Development of Income-Related Health Inequality in Europe

The Impact of Different Value Judgements

Mieke Järvinen

Emilie Toresson Grip

Supervisor: Professor Ulf-G Gerdtham

NEKP01 Master Thesis II

Department of Economics

Lund University

August, 2015

Abstract

This essay estimates the trends in Income-Related Health Inequalities (IRHI) in 11 European countries between 2006 to 2013 and investigates the sensitivity of these trends with respect to different inequality measures. The motivation for this study is both empirically and theoretically driven. Empirically, there have been many studies aiming to quantify and explain IRHI but relatively few have investigated the trends in IRHI. Of those who have, an increasing, to the least, stable trend in IRHI seem to be suggested. However, recent theoretical developments and debates have highlighted technical complications and lately, normative challenges that arises when researchers aim to estimate IRHI using bounded health variables, variables that are commonly used in the literature. Thus, by summarising the most recent theoretical developments in the field of IRHI measurements and applying these different inequality indices in an empirical analysis, this essay contributes to the current literature by, firstly, the trend analysis over time and secondly, and most importantly, by the thorough and transparent empirical design which allow us to perform an extensive sensitivity analysis not previously performed.

The results yield IRHI trends that are both increasing (Germany, the Netherlands and Spain) and decreasing (Austria, Sweden, Switzerland, Belgium and Czech Republic) between 2006 to 2013. The sensitivity analysis indicate that most of these findings are robust across different measurements but contain a few examples of different and even opposing changes over time when using different indices. Thus, the results confirm the recent theoretical debate and underline the need for using several measurements when estimating IRHI. The analysis of the consistent trends across measures are explained by the characteristics of the health variable and points to the close link between the results of different measurements and the prevalence of health.

Keywords: income-related health inequality, inequality measurements, concentration index, bounded variables, Europe

ACKNOWLEDGEMENTS

As this essay represent the closure of our university studies in Economics, there are many people to whom we would like to extend our warmest appreciation and gratitude for their support during these years, as well as during this essay. With regards to this thesis, we first and foremost would like to thank our supervisor Professor Ulf-G Gerdtham for invaluable inspiration, advice and support throughout the process. Mieke would like to thank her parents, her friends and Daniel, for all the love and encouragement. Emilie would also like to thank her family for all the support (and proof reading!) and especially Linnea, Cecilia and Henric for your consistent love and encouragement.

Mieke Järvinen & Emilie Toresson Grip

16 August 2015, Lund

This paper uses data from SHARE Wave 5 release 1.0.0, as of March 31st 2015 (DOI: 10.6103/SHARE.w5.100) or SHARE Wave 4 release 1.1.1, as of March 28th 2013 (DOI: 10.6103/SHARE.w4.111) or SHARE Waves 1 and 2 release 2.6.0, as of November 29th 2013 (DOI: 10.6103/SHARE.w1.260 and 10.6103/SHARE.w2.260) or SHARELIFE release 1.0.0, as of November 24th 2010 (DOI: 10.6103/SHARE.w3.100). The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

TABLE OF CONTENT

Chapter one

1 INTRODUCTION

Inequalities in health, both within and between countries, are a persistent pattern in Europe. These inequalities in health can in many cases be linked to socioeconomic inequality, since poor health tends to be concentrated among poorer individuals in most societies (O'Donnell et al., 2008 p. 1). Income-Related Health Inequalities (IRHI) is therefore a topical issue frequently discussed among both researchers and policy makers, and many studies have been performed with the aim to quantify and explain such inequalities. This has led to a great degree of variability in the choice of method and health data.

Despite a vast field of research on IRHI, relatively few studies have examined the development of IRHI over time. Furthermore, many studies are often motivated by the seemingly general perception of a rising trend of inequalities (e.g. Islam et al. 2010; Burstöm et al. 2005; Kunst et al 2004, 2005). Although the results seem to vary across studies, several studies also seem to suggest an increasing trend in IRHI (Burström et al., 2005; Hernández-Quevedo et al., 2006; van Ourti et al., 2009). Given the very recent theoretical developments on how to measure inequality with bounded variables (e.g. Wagstaff, 2005; Erreygers 2009a; Kjellsson and Gerdtham 2013) , one might however question how sensitive these results are to, for example, the choice of inequality measure. Furthermore, since different IRHI indicators also can be shown to entail implicit value judgements, this further underlines the motivation for a sensitivity analysis of the IRHI trends. Against this background, the focus of this essay will be to investigate how IRHI has developed over time and to evaluate the sensitivity of this trend with respect to different indices.

1.1 Aim, general method and disposition

Monitoring the progress of health inequality over time remains a challenge for this research field, which so far has relied on cross-sectional studies (O'Donnell et al., 2008 p. ix). A separate theoretical challenge that have arisen very recently in the IRHI literature, regards the measurement of inequality in bounded variables, variables defined as having both an upper and a lower bound.[1] While all measures of inequality is value laden, based on principles of either

[1] Examples of bounded health variables are (fraction of) mortality, being a smoker or self-assessed health where one equal 'excellent health' and zero 'poor health'. All bounded health variables can be expressed as a

absolute or relative inequality, complexities have been shown to arise with regards to these two value judgements for bounded variables. Since this type of variable has been frequently used in previous studies, where these complexities have not been considered, motivation for further studies arises. Thus, the aim of our paper is to contribute to the existing literature by investigating how health inequality has developed over time and if and how these trends are sensitive to a) *the inequality measure* used and b*) the inequality aversion parameter*, which can be specifically altered in the so called extended Concentration Index. Specifically, the research questions examined are:

- *Has income related health inequality increased over time in Europe?*
- *How sensitive are these trends to the choice of inequality measure?*

Key is that very few previous studies have investigated whether the trends in IRHI in different countries are affected by different measures and no one to our knowledge has investigated if and how the trends change depending on the choice of the mentioned inequality aversion parameter. Thus the contribution of this essay is twofold.

Based on the motivation and aim of this thesis, it should be noted that this essay is of descriptive nature and does not aim to provide any causal explanations to inequality (as is found in the decomposition literature[2]). Given the both normative and technical challenges entailed in the IRHI literature, it rather aims to contribute to a rightful and transparent *description* of health inequality in Europe. Our essay will thus combine the theoretical discussion on different inequality measures and their implicit value judgements with an empirical application to investigate whether health inequality has increased over time in Europe.

Theoretically, the essay will describe and discuss different rank dependent IRHI measures, which in its most generalized form could be expressed as:

$$I(h) = f(\cdot)\sum_{i=1}^{n} w(\cdot)H_i \quad (1)$$

where $I(h)$ is the inequality index, H_i represent the health status of each individual, $f(\cdot)$ represent a normalizing function and $w(\cdot)$ represent a weighting function (Erreygers and van Ourti, 2011; Erreygers et al, 2012). This type of index will be employed in order to estimate trends in IRHI, furthermore, this formal representation may also give the structure, and rational, for the sensitivity analysis: by altering $f(\cdot)$ and $w(\cdot)$ we will evaluate the robustness of these

mirror image in an ill-health variable, thus the ill-health variables of the mentioned variables are (fraction of) survival, being a non-smoker or self-assessed health with an opposite scale. These 'mirror images' of the same variable will also be referred to as health measured in attainments and health measured in shortfalls.

[2] The decomposition literature aims to decompose the inequality measure and its magnitude into different contributing factors and thus to explain inequality (see for example Wildman 2003).

estimated trends. Specifically, by altering $f(\cdot)$ we obtain different inequality indices implying different value judgements, in terms of absolute or relative inequality (or dimensions in between). By altering $w(\cdot)$, we may weight the health of the poor differently compared to the health of rich, which imply a different kind of value judgement and have been referred to as altering societies' aversions towards inequality. Empirically, this paper uses repeated cross sectional data covering 11 European countries to calculate trends in different inequality indices between 2006 and 2013. As will be further described in the theoretical section, the properties of these indices are closely related to the choice of health variable. We will therefore carefully motivate and construct a suitable health index, specifically developed for our data. The use of this index is limited in the previous literature and therefore further adds to the contributions of this thesis. Although the essay is descriptive in its nature, as the IRHI literature is to a large extent, an ordered probit model will be employed to construct the health index while ordinary least square regression models will be employed to estimate the various health indices.

Finally, the paper is laid out as follow: This first chapter presents the aim and background of the paper together with a summary of previous research. Chapter two outlines the theoretical framework by providing a description of different types of health variables and inequality indices. The third chapter begins with a description of the data and is followed by a discussion of the empirical strategy. A selection of descriptive statistics and the results are presented in chapter four, whereas the final discussion of the results, limitations of the study and conclusions are found in chapter five. Further descriptive statistics and results are found in the appendices.

1.2 Background

1.2.1 Why care about health inequality?

Economics is a normative discipline and has since the work of Adam Smith and Vilfred Pareto to a large extent been engaged in questions and concepts of economic justice, fairness and equality (Kolm, 2007). With regards to normative standpoints in the field of health care, reference is often made to the Nobel Prize winner Amartya Sen (2002) who has argued that inequalities in health are of particular concern: "In any discussion of social equity and justice, illness and health must figure as a major concern." (p. 659). Coupled with his theory of capabilities for social justice, the rationale is that health is a necessity for other human capabilities and thus for real opportunity for freedom.

To be noted here is that this essay, in line with the literature, makes a distinction between inequality in health and health inequality conditional on the socioeconomic ranking. Since poorer individuals tend to suffer from higher mortality as well as higher morbidity compared to richer individuals (O'Donnell et al., 2008 p. 1) and since these health differences across socioeconomic groups are not necessarily due to differences in preferences but rather constraints in terms of income or human capital (Kjellsson, 2014), these inequalities have been considered to be inequities (Wagstaff and van Doorslaer, 2000). This essay focuses on this particular kind of inequity, namely Income-Related Health Inequality (IRHI).

1.2.2 Previous literature

This essay will evaluate health inequality trends in Europe and this overview of previous research therefore reflect studies conducted in the same geographical and economic context. In addition, studies using data collected earlier than 1980 have been excluded from the overview. In general, the vast majority of previous research on IRHI has been done using cross-sectional data. Some of these studies have calculated and compared different IRHI indices for different countries in order to examine if the ranking of countries can change depending on the measure used. The results suggest that they may differ (Gerdtham and Kjellsson 2013, 2014; Erreygers 2009a, 2009b; Doorslaer and van Ourti, 2011). Gerdtham and Kjellsson (2013) show that the ranking of countries with two different measures (Wagstaff's (W) and Erreygers' (E) inequality indices) when applied on nine different bounded variables do differ, although to a varying extent.[3] These different indices represent different expressions of the normalising function, depicted in equation (1) above. The findings in these studies suggest, among others, that the rankings may differ depending on the bounded variable being expressed as a health variable or as an ill-health variable, and that the mean value of a health variable has a large impact on the rankings on these different indices. Wagstaff (2002) developed the extended Concentration index (extended C), an index which implies altering the weighting function as above depicted in eq (1). In the same article he applied it on 44 developing countries and found that using different weights for aversion against inequality also did change the ranking of different countries, since the impact of using different inequality aversion weights differed between countries. This was tested by Hernández-Quevedo et al. (2006) in a European context. They found that the ranking remained rather stable, with the exception of a few countries.

[3] The different rank dependent inequality measures mentioned in this section such as Wagstaff's inequality index (W), Erreygers' inequality index (E) and the Concentration index (C) will be further introduced and described in the theoretical section, chapter 2. Further description of different kinds of health variables will also be found in chapter 2.

The much more scarce previous research on longitudinal data and some of its findings have been summarized in table A1, Appendix A.[4] As can be seen in column 5, most of the studies indicate an increasing trend (Burström et al., 2005; Hong et al., 2011; Gravelle and Sutton, 2003; Leu and Schellhorn, 2006) although the results overall vary somewhat and furthermore entail a wide array of measures and variables employed. Both the rather limited amount of longitudinal studies and their very different research designs, as can be seen in table A1 in Appendix A, are noteworthy. The study by Hernández-Quevedo and Masseria (2013) on European countries is particularly interesting since they use and compare different measures *over time* in several European countries. The trends differ somewhat depending on country, time period, measure and variables used but for most countries there seem to be an increasing, to the least stable, trend in IRHI. As far as we have been able to find, the study by Hernández-Quevedo and Masseria (2013) is the only study that has compared the standard C with E over time. They show that the trends for E and C differ for two countries using a binary variable of self-assessed health and more substantially for several countries when using health-limitation (i.e. ill-health).

Not only are the previous longitudinal studies on trends in IRHI scarce, but many of them have bypassed technical and normative challenges which in recent theoretical IRHI literature have been intensively debated (and empirically evaluated in the previous mentioned cross sectional studies). In terms of value judgements, it can for example be shown that C, which is sensitive to changes in relative inequality, ranks countries differently depending on the bounded variable being expressed as a health or as ill-health, although it is the same variable. Therefore, inequality analysis requires careful consideration of which variable and which index to use, as each index entails implicit value judgements (Kjellsson and Gerdtham, 2013). Based on this, we sum up and conclude that there seem to be much research needed in order to gain more insight in the trends on health inequality in Europe. Given the theoretical debate, to be presented further in the next chapter, a comprehensive sensitivity analysis of the estimated trends seems further to be motivated.

[4] Naturally, there is a limited possibility for an accurate summary of studies in a table format, despite our intentions of making a just summary. The table should therefore be read cautiously and the interested reader is referred to the respective study for a more detailed and accurate description of the results. Note also that due to the limited amount of previous longitudinal IRHI studies, studies using education instead of income as the socioeconomic variable is included in this table.

Chapter two

2 THEORETICAL FRAMEWORK

IRHI measures aims to answer the question "To what extend are there inequalities in health that are systematically related to socioeconomic status?" (Wagstaff et al., 1991 p. 545). The aim of this thesis is to investigate how IRHI has developed over time and how sensitive these trends are to the choice of inequality measure. The motivation is both empirically and theoretically driven. Empirically, most of the few previous studies investigating health related inequality trends have suggested an increasing, to the least stable, trend over time as seen from table A1 in Appendix A. By altering the choice of measures and data used these results can be evaluated. Performing a sensitivity analysis can furthermore be motivated theoretically. In the dissertation by Kjellsson (2014) he states: "Whereas the ethical aspects of health constitute a reason for concern for health inequality, the technical aspects of health variables constitute a reason for giving specific attention to health-inequality measurement."(p. 2). This refers to the fact that many inequality measures have been developed for an income variable, which is different from many health variables, as we will see. Although these differences cause technical challenges for the (health) inequality measures as Kjellsson phrases it, in many cases these challenges also imply implicit value judgements on the definition of health inequality. A sensitivity analysis is therefore needed to evaluate how sensitive these (and indirectly previous) results are to the different methodological choices that implicitly entail different value judgements. This chapter focuses on the latter part of the motivation to this thesis: it will describe the theory behind different health variables and health inequality measures.

2.1 Measuring income related health inequality

This section covers the theoretical framework used when measuring income-related health inequality. First, the properties of different health variables are discussed since these properties are closely related to the properties of IRHI indices. This is followed by a discussion of absolute and relative inequality, and how distinguishing between these two implies making a value judgement of what kind of inequality one is interested in measuring. The preceding sections deal with the properties of the inequality indices, and present the indices used in this theses: the standard concentration index (C), Wagstaff's normalised index (W), Erreygers' corrected concentration index (E), and the extended concentration index (extended C).[5] Since this essay focuses on IRHI, we are interested in how much more healthier the richer individuals are compared to how healthy poorer individuals are. This focus implies the need of two variables: an income variable and a health variable, and thus these inequality measures to be described have been referred to as bivariate measures (Kjellsson, 2014 p. 4).[6]

2.1.1 Different properties of health variables

Erreygers and van Ourti (2011) categorize health variables by two dimensions: *measurement scale* and *boundedness*. With regards to measurement scale, Kjellsson and Gerdtham (2014) describe five different scales on which a health variable can be measured, which are summarised in table 1 below. Health variables can also differ in terms of boundedness (the second dimension): is the upper and lower bound for the health variable bounded (that is, finite) or unbounded? Income is an example of an unbounded variable and this is why complication arises when one tries to measure inequality in health based on principles from the income literature. If a health variable (H_i) is bounded (with upper limit H_{max}) it means that a corresponding *ill-health* (S_i) variable can be constructed by the health variable:

$$S_i \equiv H_{max} - H_i \quad (2)$$

In this perspective, the health variable captures *attainments* in health whereas the ill-health variable captures *shortfalls* in health. Health variables may also be *binary*, which in the light of the above categorization is an ordered variable. However, a binary variable can also be interpreted as a ratio variable where the binary variable captures an average prevalence of a health variable at different deciles/percentiles (Kjellsson and Gerdtham, 2014).

---

[5] In comparison to other bivariate inequality measures available, one of the most cited advantages of these rank dependent measures is that these measures are sensitive to inequality in the middle of the distribution; other measures (e.g. the range measure) are more sensitive to the tails of the distribution (i.e. the extreme values of the distribution) (Wagstaff et al. 1991). The relative index of inequality (RII) and the slope index of inequality (SII) are other measures, with these advantages; these have however been mostly used in research by social epidemiologists (Hernández-Quevedo et al., 2006).

[6] The Gini coefficient (G) is a well-known, univariate inequality measure which measures how much richer the richer individuals are (if used with an income variable). However, this measure may also use a single health variable to assess the question of how much more the healthy individuals are compared to the unhealthy individuals. The key point is that this inequality measure, in contrast to IRHI measures where health is conditioned on socio-economic ranking, only needs one variable, thus called a univariate measure (Kjellsson, 2014).

Table 1. Different measurement scales for health variables.

Nominal	Classification of individuals into different categories but without ranking (e.g. type of disease).
Ordinal	Ordering of individuals but without possibility to measure the difference between them (e.g. a scale of self-assessed health, where the difference between 'very poor' and 'poor' cannot be compared to the difference between two other categories).
Cardinal	Having an arbitrary 'zero point' which does not entail an intuitive interpretation of total absence of health; differences can thus be calculated but ratios cannot (e.g. body temperature).
Ratio	With the 'zero point' corresponding to zero, thus total absence of e.g. health, implying that ratios can be calculated but the variable can be expressed with different scales (e.g. health care expenditures in dollars).[7]
Unique	Similar to 'ratio' but without the possibility to rescale the variable (e.g. number of GP-visits).

Source: Kjellsson and Gerdtham (2014)

2.1.2 Changes in health and how we should measure them: absolute or relative inequality or something in between?

For unbounded, ratio scaled variables health can typically change in relative terms (health can improve by 50 percent) implying a *proportional change* or in absolute terms (improvements of a certain amount) implying a *uniform change*. Since these changes can be measured in relative or in absolute terms, how should we measure them? To answer this question the researcher is bound to make a normative choice; the researcher needs to apply an inequality equivalence criterion (IEC). To make a distinction between the normative arguments and implications behind an absolute and relative measure of inequality, the IEC defines *what change of health preserves inequality* (a proportional change preserves the inequality for a relative IEC and a uniform change preserves the inequality for an absolute IEC)(Kjellsson, 2014).

However, the normative choice of absolute or relative inequality becomes more complicated in the case of bounded variables. One debated finding is that a change in the same (bounded) variable, though in one case measured as attainments and in another case measured as shortfalls, can by a relative measure indicate different (increasing and constant) or even opposite (increasing

[7] For the sake of clarity, if you have a ratio scaled variable you can express two values of this variable as a ratio. For example, if individual A has 1000 US dollar in health expenditure a year whereas another individual B has expenditures of 500 US dollar a year, A has twice as much expenditures a year compared to B (1000/500 = 2). However you cannot work with ratios when using cardinal variables since the zero point, 0 degrees Celsius for example, does not mean absence of heat or no heat. Therefore, a temperature of 40 degrees Celsius is not twice as much as 20 degrees of Celsius.

Table 2. Illustrative example of the two relative value judgements for bounded variables.

Suppose that one society consists of two groups: group A has an average health of 20 and group B has an average health of 30 (where health is a bounded variable and maximum health equals 100, one can for example think of 100 as meaning maximum life expectancy). Average health in the whole society (A plus B) is 25 ((20+30)/2), but imagine that overall average health increases to 50, thus an increase of 25 years. To preserve the relative inequality between A and B, this increase would be distributed so that health in group A increased from 20 to 40 and health in group B increased from 30 to 60 (A still has two thirds of that of B), while overall health would have increased to 50. If one instead would look at this situation using the ill-health variable (where 100 now equals maximum ill-health, for example that 100 is the maximum shortfall in life expectancy), the change would be different. As expressed in ill-health, A has an average ill-health of 80 while B has an ill-health of 70. First, one may note that the relative inequality to start with is now different, despite looking at the same variable (the relative difference between 70 and 80 is less than the relative difference between 20 and 30, but the absolute difference is the same). Second, if health is increasing it means that ill-health is decreasing. Thus the decrease in ill-health (by 25 years) will now be distributed as to preserve relative inequality in ill-health – this would yield an ill-health level of $53^{1/3}$ and $46^{1/3}$ for A and B respectively. The two key findings are, first, the inequality-preserving changes of these two changes do not correspond, despite that both preserve relative inequality in the same variable, though in one case expressed as a health variable in the other case expressed as an ill-health variable. Second, if this increase in 50 years would have been distributed to preserve absolute inequality instead, both health and ill-health would have uniformly increased and decreased by 25 years in A and B, thus these changes do correspond. Therefore, it has been argued that there is one absolute IEC and two, different relative IEC for bounded variables (Kjellsson and Gerdtham, 2015).

and decreasing) trends in these two cases. An absolute measure on the other hand, always shows the same (change) in inequality no matter if the variable is measured in attainments or shortfalls. Since this finding is key for the understanding of this essay an illustrative example could be useful, which can be found in table 2. One strand of literature, related to the income literature, has interpreted this as a technical inconsistency. In the income inequality literature, this result would be nonsensical, since the main variable of interest, i.e. income, is a ratio scaled variable (Kjellsson and Gerdtham, 2015).[8] However, is this nonsensical in the light of health inequality when most variables are not unbounded (ratio-scaled), but bounded? Kjellsson and Gerdtham (ibid) argue, in contrast to the above, that the correct interpretation is that with bounded variables the normative choices for an IEC is not of two (relative or absolute) value judgements but a choice of one absolute and *two* different kinds of relative value judgements. The argument is based on the fact that if measured in attainments, a proportional inequality-preserving change is different to a proportional inequality-preserving change in shortfalls. As noted this has been interpreted as nonsensical but recall that this difference on what change of health that preserves inequality is also the <u>only</u> difference between an absolute and a relative inequality measure (for example in the income literature using ratio scaled variables). Thus, if these different inequality preserving changes are not an issue with an unbounded variable (which is typically the case in the income inequality literature), and only a matter of value judgements, then accordingly, these differences may not be a problem with a bounded variable (in the health inequality literature), *only*

[8] This discussion could however apply to the income literature as well in the cases where one, for example, measures poverty by some bounded variable.

a matter of <u>other</u> value judgements. Kjellsson and Gerdtham (2015) distinguish between the two different IEC:s represented by relative changes in shortfall and attainments respectively and refer to these as s(hortfall)- and h(ealth)-relative changes.[9] As will be described further below, this discussion is closely related to a discussion on desirable properties of bivariate inequality measures.

As stated indirectly in the introduction (section 1.1.), the aim of this thesis is *not* to discuss and conclude on the most suitable or superior inequality measure from a normative standpoint (this is a topic for an essay on its own). Rather, the aim is to highlight and problematize value judgements that have often been made implicitly in previous research and to empirically evaluate how these affect the calculations of IRHI trends. This is especially motivated given the above; the relative-absolute IEC's become complicated depending on what kind of health variable is chosen by the researcher.

2.1.3 Income related health inequality indices

2.1.3.1 Desirable properties of IRHI indices

Kjellsson and Gerdtham (2014) summarise properties of rank-dependent inequality indices that have been argued to be essential in the previous literature. These include the *transfer property* and *scale invariance*, which are important for all health variables, as well as the *mirror property* which however is only applicable for bounded health variables. The transfer property implies that a transfer of health from a poorer individual to a richer individual will result in a more pro-rich inequality index. Scale invariance indicates that the scale of the health variable does not influence the inequality index. This property ensures that different cardinal scales (such as Celsius and Fahrenheit) will result in identical inequality indices. Finally, the mirror property ensures that the level of inequality remains the same for both attainments and shortfalls; the inequality index should be equal in both cases, but with opposite signs. However, although the desirability of this property has been underlined by Erreygers (2009a) it is debated. The argument against the mirror property stems from the discussion on the IEC and the different value judgements involved when the health variable is bounded (described in section 2.1.2. above). As previously stated,

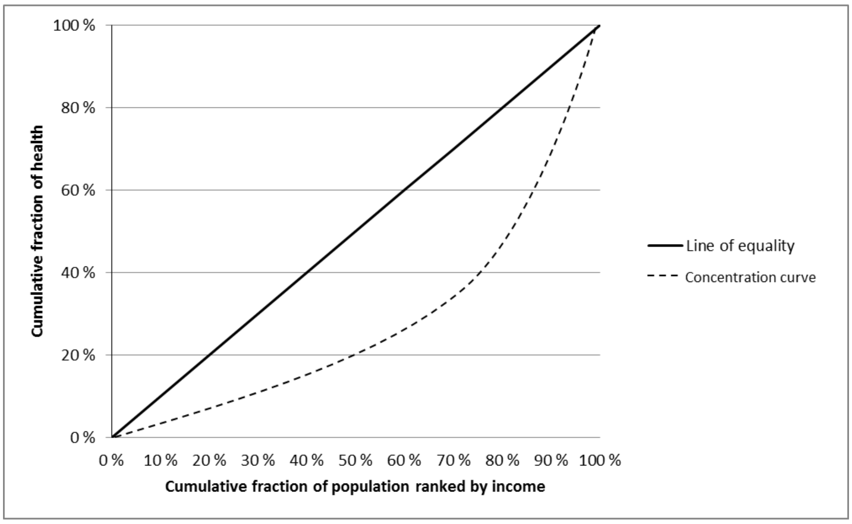
[9] For example, s-relative IEC may correspond to a value judgement where the ill are treated in proportion to their level of illness. An h-relative IEC is in principle closer to the standard relative IEC: a relative inequality preserving change implies that increasing life expectancy is distributed in proportion to how healthy individuals are from the beginning (Kjellsson and Gerdtham, 2013).

when dealing with a bounded variable, one can either make an absolute value judgement, and obtain identical results regardless of using attainments or shortfalls, or use a relative measure of inequality and deal with the resulting health-relative or shortfall-relative value judgements that leads to non-identical inequality values. Since this is a general result, that is, a relative index can never yield the same level of inequality in attainments and shortfall, this implies that to fulfil the mirror property, one needs an absolute measure of inequality (Kjellsson, 2014). Therefore, when imposing the mirror property one also imposes an implicit value judgement of regarding measures based on inequalities in ill-health and inequalities in health to be the same (Kjellsson & Gerdtham 2014).

2.1.3.2 The standard concentration index (C)

The Concentration Index (C) is the most commonly used measure of income related health inequality. This is partly due to its simple interpretation and the possibility to graphically illustrate it using the Concentration Curve (CC). C can be illustrated by the concentration curve that plots the cumulative fraction of health against the cumulative fraction of the population ranked by socioeconomic status such as income (Kjellsson and Gerdtham, 2014). The figure also includes the 45 degree line of equality which represents the case of no inequality, or equally, perfect equality.

Figure 1. The Concentration Curve



Source: Based on Kjellsson and Gerdtham (2014, p. 241)

Looking at the CC in Figure 1 one can see that for example the poorest 40 percent of the population holds less than 20% of the total health, while the line of equality implies that the poorest 40 percent holds 40 percent of the accumulated health (Kjellsson and Gerdtham, 2014). Therefore, if the CC lies above the line of equality there is a pro-poor distribution of health, and if the CC lies below the line of equality the distribution of health favours the rich. Hence, a pro-rich distribution of health causes C to take a positive value, while a negative value indicates a pro-poor distribution of health. However, if the health variable describes ill-health rather than good health, the situation is reversed (O'Donnell et al. 2008: 83-84). Moreover, for unbounded variables, C takes values between -1 and 1. However, Erreygers (2009a) based on Wagstaff, (2005) has shown that with a bounded variable the bounds of C are dependent on the mean of the health variable. This is one of the shortcomings of the standard C and will be discussed more in detail in subsequent sections.

Researchers are often interested in comparing the level of inequality between time periods and/or countries. In this case, using only concentration curves or simply graphs of the distribution of health and income in the analysis is not always sufficient.[10] Therefore it is convenient to use the concentration index which gives a numerical value of the degree of inequality, enabling comparisons between different time periods or countries (O'Donnell et al. 2008: 84). For a discrete living standard variable, C can be written as:

$$C = \frac{2}{n \cdot \mu} \sum_{i=1}^{n} H_i R_i - 1 - \frac{1}{n} \tag{3}$$

where H_i is the health variable, *n* is the number of individuals, $\mu = (\frac{1}{n}) \sum_{i=1}^{n} H_i$, i.e. the mean of the health variable, and R_i is the fractional rank of individual *i* (which is equal to λ / n, where λ equal the rank ranging from 1 (the poorest) to *n* (the richest)).[11] C therefore takes into account both the distribution of health, each individual's relative position measured by ones socioeconomic rank and the level of health of all individuals in a population (Kakwani et al. 1997; Erreygers 2009a). Since this thesis adds to the current literature by employing several measures over a time period, it would be convenient to find an expression of the respective measures that highlights both their statistical similarities and differences. This is exactly what Erreygers (2009a) does by showing that C, W and E belong to the same family of rank dependent indices of health inequality and this is a somewhat less general expression of the equation presented in the introduction (see section 1.1). He expresses the general form of these indices as the sum of socioeconomic weighted health levels (the last term in the equation below), being normalized by a function of four parameters (the first term in the equation):

[10] It is possible to test whether one CC dominates another; if one CC lies above the other at all points it dominates the other curve. If this is the case, it is possible to conclude that one country (or time period) is more equal than another. However, if the two curves cross at some point, one cannot state that one curve dominates another; hence in this case it is not possible to compare the level of inequality between the two countries using only concentration curves.

[11] For large *N*, the last term approaches zero in eq. (3), why it is typically omitted (O'Donnell et al 2008).

$$I(h) = f(\mu, n, a_H, b_H) \sum_{i=1}^{n}(Z_i H_i) \tag{4}$$

where μ equals the mean of the health variable, n equals the number of individuals in the population, a_H and b_H represent the lower and upper bound of the health variable. In all these indices the weighted sum of the health variable is normalized by these four parameters, however how this is specified in the function *f(·)* is what constitute the difference between the indices.[12] Expressed in this general form the C can then be written as eq. 5 below. In the next sections W and E will also be expressed in these general forms.

$$C = \frac{2}{n^2 \cdot \mu} \sum_{i=1}^{n}(Z_i H_i) \tag{5}$$

Despite having many desirable properties of a health inequality index (such as depending on the socioeconomic rank and taking into account the health of all members of the society), C has been found to be problematic when used with certain kinds of health variables (recall section 2.1.1 describing different categories). In short, the main problem relates to bounded variables where C for example is scale invariant only for ratio-scaled variables. Thus, C does not have all the desirable properties discussed in section 2.1.3.1 if C is used with nominal, ordinal, or cardinal variables. Regarding the more debated mirror property, C does neither satisfy this since C is a relative measure (thus it will show different results depending on the bounded variable being expressed as health or ill-health).[13]As a result, modifications correcting for the shortcomings of these indices have been suggested. Erreygers and Van Ourti (2011) have suggested a modification of C for cardinal variables, both bounded and unbounded variables, expressed as:

$$C = \frac{2}{n^2 \cdot (\mu - H_{min})} \sum_{i=1}^{n}(Z_i H_i) \tag{6}$$

By this modification[14], modified C now satisfies scale invariance for cardinal variables (Erreygers and van Ourti, 2011; Kjellström and Gerdtham, 2014). However it does not satisfy the mirror condition for bounded variables (different results appear if the health variable is measured in attainments or in shortfalls) and therefore other measures have been developed especially to be

[12] Given the above eq. (3), the last term of eq. (4) (the weighted sum of health) can be shown to be equal to $Z=(n+1)/2 - \lambda$. This definition is the same for C, W and E, but will be altered in the extended versions of these indices (when evaluating the second kind of value judgments within IRHI measures, as it was described in the introduction (recall section 1.1)).

[13] While C is a measure of relative inequality, there also exists a related measure for absolute inequalities. In order to measure *absolute* inequalities the generalized C (GC) was initially developed by Wagstaff et al. (1991).). However, we do not use GC in this essay as it requires an unbounded health variable.

[14] To be noted for later purposes, this modification corresponds to the computation of C on a normalized health variable m_i, where the minimum value is restricted to be set to 0, i.e. $m_i = H_i - H_{min}$ (Kjellsson and Gerdtham, 2014).

used for bounded variables, which will presented next.

2.1.3.3 Wagstaff's normalisation of Concentration Index (W)

For binary variables, Wagstaff (2005) shows that differences in the mean of the distribution of the chosen health variable can lead to different maximum and minimum values of C. The bounds of C will then be $1-\mu$ and $\mu-1$, implying that a higher mean causes the range for the values of C to becomes narrower. This 'bounds issue' is not resolved by the modified C and the implication is that one cannot compare the C:s measured for populations with different means of the health variable (Wagstaff, 2009). The author therefore proposes that the standard C should be normalised. Erreygers' generalises this normalisation to all bounded variables with a lower bound a_h and an upper, finite bound b_h, and following his generalized notation for the Wagstaff's correction it can be written as:

$$W = \frac{\mu(H_{max}-H_{min})}{(H_{max}-\mu)(\mu-H_{min})}C \tag{7}$$

Expressed in the general form where all indices share one term (as explained in the previous section), it can be written as:

$$W = \frac{2}{n^2}\frac{(H_{max}-H_{min})}{(H_{max}-\mu)(\mu-H_{min})}\sum_{i=1}^{n}(Z_i H_i) \tag{8}$$

This normalization of the standard C satisfies both the mirror property and is scale invariant for cardinal variables. Recalling section 2.1.3.1, satisfying the mirror property has been shown to be incompatible with a relative inequality measure (a general result). However, it can be shown that W neither is an absolute inequality index. For an increase in the (standardized) health variable in absolute terms, all else equal, W will increase if the mean of the health variable is larger than 0.5 and decrease if less than 0.5 (Kjellsson and Gerdtham, 2014). This may be and have been interpreted as counterintuitive, since the inequality equivalence criteria (IEC) (the concept that was presented in section 2.1.2) made by the researcher when using this measure can neither be a pure relative nor a pure absolute IEC (a proportional change will not always leave W unchanged and neither will an absolute change).

2.1.3.4 Erreygers' corrected Concentration Index (E)

Erreygers (2009a) has shown that the value of C used with a non-ratio scaled variable depends on the scale of the health variable (e.g. body temperature measured in Celsius or Fahrenheit), thus it is not scale invariant.[15] Furthermore, given his criticism of W (see Erreygers 2009a; 2009b; Erreygers and Van Ourti, 2011), Erreygers derived a different corrected C, which will be referred to as E. Like W, E also satisfies the transfer property, scale invariance and the mirror condition for cardinal variables and is also only applicable for bounded variables (Kjellsson and Gerdtham, 2014). The index can be expressed as (Kjellsson and Gerdtham, 2014) :

$$E = \frac{4}{(H_{max} - H_{min})} \mu C \tag{9}$$

Expressed in the general form where all indices share one term (as explained in the previous section), it can be written as:

$$E = \frac{8}{n^2 (H_{max} - H_{min})} \sum_{i=1}^{n} (Z_i H_i) \tag{10}$$

Erreygers (2009a) argues that this index is superior to W. The different corrections and modifications of C when used for bounded variables sparked a debate between the two authors, where Erreygers (2009a) criticises W by stating that it generates counterintuitive results. Wagstaff (2009) replies by highlighting that E is an index of absolute inequality, and that W never was intended to measure absolute inequalities.

2.1.3.5 A comment on the Wagstaff- Erreygers debate

In order to provide a background and to guide our empirical strategy, the above sections have tried to highlight the estimation challenges a researcher faces when using bounded variables, and the corrections that have been suggested to deal with them. It should be noted that this debate is still ongoing and there is yet no agreed upon best practice in empirical research. Kjellsson and Gerdtham (2013, 2014) have however provided a recent discussion of the debate between Wagstaff (2005) and Erreygers (2009a), and put forward some new arguments for the use of W. They specifically point to the fact that it is important to take into account both the technical and normative features of the index when measuring inequality. As for the normative features they highlight that these indices all differ in their definition of the most unequal society. C defines the most unequal state as a situation where the richest single individual has all health available in a society, E defines it as where the top 50 percent of the population has all health, while W defines it as a society where the richest proportions of individuals have all health, where this proportion corresponds to the mean health of the bounded variable. This implies that W depends on the

[15] However, C is also scale invariant for unique measurement scales, e.g. number of GP visits, which only have one scale.

prevalence of health in the society.[16] The respective indices then answer how far away the distribution at hand is from the respective definitions of the most unequal society.

Combining these definitions with the previous discussion in section 2.1.2, regarding how to measure a change in health when dealing with bounded variables, Kjellsson and Gerdtham (2013) furthermore discuss the inconsistencies of C and W. They stress that what may be and have been interpreted as counterintuitive and undesirable results when using (modified) C for bounded variables and W, may instead be interpreted as the result of value judgements, rather than technical inconsistencies, specifically for bounded variables. If one acknowledges that the relative IEC may be two folded as described in section 2.1.2, one may interpret that using C for a health variable (measuring health attainments) suggest that a certain level of absolute inequality is more severe when prevalence of health is low, while using C for an ill-health variable (measuring shortfall in health) suggests the opposite (i.e. when the level of ill-health is low). That is, a certain change in health may increase relative inequality in a health variable while at the same time decrease relative inequality in an ill-health. What can be seen is that following the above definition of the most unequal society, W actually reflects both relative inequality and relative inequality in ill-health, and is therefore referred to as a 'mirror relative' index by Kjellsson and Gerdtham (2013). From this perspective the formal definition of the value judgement of W suggests that the same level of absolute inequality is more severe for both high and low values of health prevalence (since high levels of health imply low levels of ill-health) in a society. The crucial point is that unless there can be made an explicit motivation to why one would prefer to measure relative inequalities in health rather than ill-health or vice versa, neither C(h) or C(s) is desirable. Since W merges both these value judgements simultaneously, causing the counterintuitive results, this may rather form the justification for using W to measure relative changes when using bounded variables.

Lastly, comparing the definitions of W and E (eq. (7) and (9)), we can see that the indices coincide with each other when mean health μ equal 0.5 (and the bounds for the bounded variable is one and zero). More specifically, it can be shown that the greater the difference in average health level in the various distributions, the greater the respective indices will diverge from each other. Further, by noting that the numerator of C, W and E (eq. (6), (10) and (8)) is invariant to

[16] If mean health is 0.3, the most unequal state is where the richest 30 percent of the population has all health. Furthermore, the level dependence of W (which is also the case for C) can be detected in the previous formal definitions of the respective indices, noting that mean health enter the denominators in eq. (3) and eq. (7), which is not the case for E.

an equal increase of health across a population, the change in these indices (of an equal increment) is only dependent on the denominator. This may formally summarise the previous discussion on the implicit value judgement with these indices. Noting that the impact of an absolute increment can be evaluated by the derivative of these expressions with respect to the mean of health μ, the derivative is constant (0) for E, strictly negative for C while for W, equal increments decreases the inequality measure if the mean health is below 0.5 while it increases inequality for W if average health is above 0.5 (Kjellsson and Gerdtham (2013).

2.1.3.6 The Extended Concentration Index

The indices discussed in previous sections, the standard C, and Wagstaff's and Erreygers' modifications W and E, all embody certain value judgements as discussed in previous sections. With regards to the generalised form of rank-dependent measures presented in the introduction (eq. (1)), these measures all implied modifications (and different value judgements) of the so called *normalising function*. Nevertheless, there is yet another modification to the standard C, where the inequality index is constructed to explicitly take into account *aversion to inequality*. This modification uncovers the implicit value judgements made in the *weighting function* in eq. (1) of rank dependent inequality measures. Wagstaff (2002) develops such an index, called the Extended Concentration Index, where the standard C is modified to include a so called inequality aversion parameter, *v*. The extended C can be written as the following, where *v* is the parameter measuring attitudes towards inequality[17]:

$$C(v) = 1 - \frac{v}{n \cdot \mu} \sum_{i=1}^{n} H_i (1 - R_i)^{v-1} \quad = \quad 1 - \sum_{i=1}^{n} \frac{H_i}{n \cdot \mu} w_i(R_i, v) \tag{11}$$

From the above, $w_i(R_i, v) = v(1 - R_i)^{v-1}$ is the weight of individual i's share of health, i.e. $(\frac{H_i}{n \cdot \mu})$. This means that when all individuals' health is of the same weight we have $w_i = 1$ and $v = 1$, which implies indifference to inequality and hence the value of the extended C will be zero (even though inequalities may be present). Furthermore, if $v = 2$, the extended C is equal to the standard C. Increasing the inequality aversion parameter above 1 causes the weight attached to the health of the poorest individuals in the society to increase, while the weight of the richest individuals in the distribution will decrease. In order to grasp the concept of the inequality aversion parameter, consider a *v* of 8. Such a high aversion to inequality causes the weight attached to the riches *half* of the distribution to decrease to practically zero.

[17] To see the similarity between the standard C and the extended C somewhat easier, one can rewrite eq. (3) (in section 2.1.3.2) as follows: $C = 1 - \frac{2}{n \cdot \mu} \sum_{i=1}^{n} H_i (1 - R_i)$

What the extended index shows is that all inequality indices are dependent on a value judgement of the society's preference of equality, and how averse the society is towards inequality (Wagstaff, 2002). A more inequality-averse society will have higher inequality indices than a less inequality-averse society as long as the distribution of health over socioeconomic status is monotonically increasing (i.e. better health for richer individuals). However, if it is not, extreme values of v may result in lower indices, or even negative ones. For example, if the poorest quintile has better health than the richer quintiles, raising the value of v will give a larger weight to healthier people and thus the index may instead indicate a pro-poor distribution of health (Erreygers et al. 2012).

Lastly, it should be underlined that the attitudes towards inequality of society are not given, indicating that the choice of the inequality aversion parameter remains largely arbitrary; one can expect a value of v above 1, but finding the value that realistically portrays a society's preferences is more difficult. Therefore, one challenge that still remains in the IRHI literature is to uncover these preferences in order to develop a range of values of v that are based on actual preferences rather than randomly assigning a weighting scheme to use in policy analysis (Erreygers et al. 2012).

2.2 Summary: properties and preferences

In the debate between King et al. (2010) and Asada (2010) on unbounded variables they came to conclude that a researcher aiming to investigate and describe inequality ought to use both relative and absolute measures of inequality. Kjellsson and Gerdtham (2013a) and Allanson and Petrie (2013) have in very recent contributions extended this reasoning to research using bounded variables, with the addition that both health relative and shortfall relative indices should be presented by the applied researcher. Given that the aim of this essay is to describe and evaluate IRHI trends in Europe with respect to different inequality measures, these recommendations guide our empirical design. Against the previous theoretical discussion on different available measures to use for this purpose, a summary of "appropriate" measures to be used by the researcher, is provided in table 3. Since this empirical analysis not only is (naturally) theoretically guided but furthermore theoretically motivated, the choice of health measures employed in this study will be motivated by the different inequality measures we aim to use in order to evaluate the sensitivity of IRHI trends in Europe.

Table 3. Properties of indices for bounded variables.

Inequality index	Value Judgement	Properties				
		Transfer[18]	Mirror	Scale invariance		
				Cardinal	*Ratio*	*Unique*
C	Relative	✓			✓	✓
Modified C	Relative	✓		✓	✓	✓
W	Mirror relative	✓	✓	✓	✓	✓
E[19]	Absolute[20]	✓	✓	✓	✓	✓

Source: Based on tables in O'Donnell et al. (2008) and Kjellsson and Gerdtham (2014).

By using a bounded variable, we may use all measures mentioned above and thus test whether these measures have an influence on the direction of trends in IRHI in Europe. Following the recommendations by King et al. 2010, Asada 2010, Kjellsson and Gerdtham (2013a) and Allanson and Petrie (2013), we will make use of both absolute and relative measures, where the latter will be used for a bounded variable scaled both as a health variable and as an ill-health variable. Given this, a bounded variable will be used with W and E and this bounded variable will be scaled both as an health and an ill-health variable when using the C index (since C, in contrast to W and E, does not satisfy the mirror property as was described earlier). These different measures will be employed in order to investigate IRHI in overall health in Europe. In the next chapter the methodological considerations for the empirical part of this analysis will be further explained and motivated.

[18] The transfer property is satisfied for all indices above when the health variable is nonnegative. However, with a health variable that may take negative values, only the modified C, W, and E satisfy the transfer property.

[19] In addition, E satisfies a fourth property called level independence, which implies that a health increase that is equal in (absolute) value for all people will keep the measured inequality constant (Erreygers, 2009a). Nevertheless, Wagstaff (2009) argues that the property of level independence stems from the absolute nature of E. Hence, we do not include level independence as a property of inequality indices as it implicitly entails an absolute value judgement.

[20] Formally, the bounds of a bounded variable may act as constraints to an equal increase in health across a population, thus the concept of an absolute value judgement is not entirely the same as for unbounded variable. However, while we acknowledge this, we will continue to refer to this value judgment as an absolute, following Kjellsson and Gerdtham (2013).

Chapter three

3 METHOD

This section lays out the estimation strategy and methodological considerations taken in order to investigate and provide an accurate picture of the IRHI trends in Europe, using different inequality measures.

3.1 Data

This essay uses the Survey of Health Ageing and Retirement in Europe (SHARE) which is a large longitudinal, individual-level dataset covering several European countries where data have been collected at five points in time between 2004 and 2013 on individuals aged 50 or older (Börsch-Supan et al. 2013).[21] As our purpose is to evaluate the trends in IRHI, we have aimed to study the longest time period possible, with the available data. Unfortunately, we had to omit the first wave of the survey since it was discovered during the analysis that the income measures in the first wave are not compatible with the income variables in later waves.[22] Wave three, on the other hand, differs from the other waves in the sense that it contains retrospective information of respondent's early life conditions (ibid). As a result, we are left with wave two, wave four and wave five, sampled in 2006, 2011 and 2013, respectively.[23]

Even though the aim of this essay is to measure trends in IRHI, we will not utilise a traditional panel dataset since the inequality indices will be calculated separately for the three years and any potential trend will be detected by visual evaluation of these three point estimates (for similar strategies see e.g. Hernández-Quevedo and Masseria, 2013; Ljungvall and Gerdtham, 2010). Thus, we use three separate cross-sectional datasets. However it should be noted that a part of the cross sectional data contains the same individuals, thus the data is neither pure repeated cross

[21] The countries included in the latest wave of SHARE (wave 5) are Austria, Belgium, Switzerland, Czech Republic, Germany, Denmark, Estonia, Spain, France, Israel, Italy, Luxembourg, Netherlands, Sweden and Slovenia. Furthermore, SHARE is primarily concerned with individuals aged 50 or older, however, as respondents' partners and other household members are also interviewed, younger individuals are included in the sample. We choose to drop all individuals under the age of 50, as they are not the primary focus of the survey, in order to enhance our estimates for the population aged 50 or older. The loss of individuals is limited: 2794 individuals are dropped across all countries for all years (SHARE Release Guide 5, 2015).

[22] Income in wave one is measured as gross income whereas the income for later years are measured as net of taxes (SHARE Release Guide 2.6.0, 2013). Given that disposable income is the appropriate income measure to use when studying the relationship between individual health and income, we were bound to omit this wave from our data.

[23] For some countries, data for these waves were collected over a period between 2006 to 2007 and between 2010 to 2011. However in the remaining part of the paper we will make references to these three waves by using these three years respectively.

sectional data nor pure panel data. An unbalanced panel is justified in our case as our analysis does not require a panel dimension; rather, we aim to calculate population representative indices on a yearly basis and this does not need the same individual to be answering the survey each year.[24] The final sample amounts to approximately 71 000 individuals contributing to around 100 000 observations in total, observed between 2006 and 2013 in 11 countries.

3.2 Variables

Given the aim of this thesis, to investigate trends in IRHI over time, our main variables of interest are health and income. In this section the choice and construction of these variables will be further described and motivated.

3.2.1 The income variable

The sample of individuals aged 50 and over will naturally entail both working and non-working individuals. Thus, using an income variable only considering salary or wage income would not accurately capture the economic situation of elderly. Fortunately, the SHARE survey contains rather detailed data on the sources of income for the respondents as well as a question on the total household income. This enables the researcher to take into account different sources of income, such as pensions and assets.

One variable that has been used in previous research is total income by all household members in the last month (Jürges 2010; Kjellsson and Gerdtham 2013). For this variable, one can use the respondents' answer to the single question of the households' total income. However, using this variable causes a rather large loss of observations since income is typically a survey question with many refused answer and cases where the respondent is unable to answer the question (the question may be difficult or require proper access to documentation, for example latest tax information at hand).[25] Therefore SHARE provides the possibility of using an imputed income variable, based on several different income sources (for which the individual has stated income from). These sources can be seen in table 4. This imputed data has been previously used in the literature (see for example Kjellsson and Gerdtham 2014, Pfarr et al. 2012 and for a different but

[24] Recall that the aim of this thesis is not to explain IRHI but rather to make an informed description of IRHI trends over time. Thus, a balanced panel data following the same individual each year would actually imply a less representative sample, since this sample would for example contain older and older individuals each year.
[25] This is supported by our data: using this variable we lose around 45 % of our observations due to missing observations or answers coded as "refusal" and "don't know".

similar imputed data set in Schneider et al. 2012), thus this variable will be our main income variable.

Table 4. Income sources included in the imputed income variable available in SHARE.

o Annual net income from employment	o Annual net income from self-employment
o Various annual public old age pension, pre-retirement pension, unemployment benefit, survivor pension, war pension, disability insurance pension	o Annual public long-term insurance payments and private long-term care insurance payments.
	o Income from rent
o Annual life insurance payments, private personal pension, private health insurance, regular payments from charities.	o Household income from rent on bank accounts, interest income from bonds, stocks and interest and dividends from mutual funds.
o Other household member annual net income.	

Source: SHARE Release Guide 2.6.0, 2013.

Due to the manipulation of the income variable one should however naturally be careful when using imputed values. Nonetheless, it may be highlighted that the income variable based on one single question may not be unproblematic either. Not only may the variable contain many missing values due to e.g. respondents' refusal or lack of knowledge of their income, they may also unintentionally make incorrect statements of their income, causing further measurement errors in the data.[26] Nevertheless, as a robustness check of our results we will calculate and briefly comment on the results using the non-imputed income variable.

Lastly, in line with the literature (Jürges, 2010; Kjellsson and Gerdtham, 2013), both income variables will be transformed to equivalent household income, in order to enable accurate ranking of individuals in households with different sizes. This will be executed by dividing total household income by the square root of the number of household members.

3.2.2 The health variable

In order to capture the health status of individuals in different European countries, the main health variable to be used in this study is based on one survey question asking the respondent to rank their own health by choosing one of five categories, ranging from 'poor health' to 'excellent health'.[27] This variable is generally referred to as "self-assessed health" (SAH). The choice of this health variable for this study can be motivated by the following: First, this variable has been

[26] For the purpose of transparency, it can be noted that we have for the non-imputed income variable also chosen to omit a few observations that we believe were intended to be coded as refusal or unknown but mistakenly have been coded as numerical values.

[27] The categories are: "Poor", "Fair", "Good", "Very good" and "Excellent".

shown to be a good predictor of both mortality and use of medical care (van Doorslaer and Jones, 2003). Since we are interested in a health measure that captures overall or general health of individuals in Europe, it is a suitable choice for this analysis. Second, SAH has been widely used in the inequality literature (*ibid*) and seems also to have been used frequently in the previous but limited trend analysis. This enable comparisons. Third, SAH is an ordinal variable (the variable contains ordered response categories) which by transformation may yield a cardinal, bounded variable. This gives motivation to use the W and E indices, as was described in the theoretical chapter (in addition to use the standard C, which in principle could be use for both bounded and unbounded variables).[28]

However, given the ordinal nature of SAH variable, which allows ordering of individuals but without having the possibility to measure the difference between them, we need to impose cardinality on the variable before the mentioned inequality indices can be employed.[29] This will be explained in the next section.

3.2.2.1 Constructing a health index

Given the extensive amount of information in SHARE, not only on income, but also on different health aspects of individuals aged 50 and over, we are able to impose cardinality on the ordinal variable using other information on health. In short, SHARE allows the researcher to use "objective" health measures (such as having asthma or experienced a stroke during the last year) together with answers on self-assessed health in order to calculate disability weights which can be used to construct a cardinal health index. This method has been proposed by Cutler and Richardson (1997) and Groot (2000) and has previously specifically been employed on SHARE data by Jürges 2005, 2007; Schneider et al. 2012; Pfarr et al. 2012 and Kjellsson and Gerdtham 2013). However, given that few datasets contains the same amount of health information as in SHARE, this method has not been applied extensively elsewhere and this further adds to the contribution of this thesis to the current literature.

[28] In the SHARE data set there are two different versions of SAH: one US version and one European version. The European version differs from the US version in terms of labelling of categories (ranges from very poor to very good). In this study the US version will be used since this is the only SAH variable available in all waves. Furthermore, Jürges (2007) highlights that the scale of the US version seems to work better in terms of providing more variance and a more symmetric health distribution across countries. When cross-checking his results of health inequality with the other health variable, he also receives similar results.

[29] This is actually not entirely correct; the C index requires a ratio scaled variable to the least. However, a cardinal variable may be used if one employs the modified index, as was explained in section 2.1.3.2. This will be done in this analysis implicitly, as will be further commented on in section C2 in Appendix C.

It can be noted that in the health economics literature disability weights of certain health states commonly are derived by using methods such as time trade-off (TTO) or visual analogue scale (VAS). Compared to these methods, the advantage of the probit model estimation is that the cognitive burden on respondents is lower (Groot, 2000). In addition, Fryback et al. (1993) have shown that the scores of this method highly correlate with the scores on disability weights derived from for example TTO.

The strategy implies constructing the cardinal health index as the linear prediction of an ordered probit model where the available SAH variable (with five categories) is regressed on "objective" health measures such as having been diagnosed different health conditions (by a doctor, in our case). The health index is the linear prediction of this regression and the prediction is normalised to obtain an index of health ranging from 0 (near death) to 1 (almost perfect health). Formally, the ordered probit model assumes that there is a varying range of health status that underlies the health categories. This underlying continuous range of health status (utility) is sometimes referred to as the latent health index. Given that we have no natural unit of measurement of health status the latent health index may be expressed as (Greene, 2009):

$$-\infty < H_i^* < \pm\infty \tag{12}$$

and the short notation of the translation of the latent variable into the five response categories of SAH (where μ_j equal the threshold value between two different response categories):

$$SAH_i = j \ if \ \mu_j - 1 < H_i^* \leq \mu_j, \ j = 1, \dots, m \ and \ m = 5 \tag{13}$$

Thus the ordered probit model assumes that the observed SAH responses censors the true underlying health index. In our case, the latent (unobservable) health variable *Hi** is furthermore assumed to be affected by observable variables through β and thus this latent regression determines the observed SAH response (as described in eq. (13) above):

$$H_i^* = x_i'\beta + \varepsilon_j, \ \varepsilon_j \sim N(0,1) \tag{14}$$

, where x contains objective health conditions diagnosed by a doctor and the notation of the error term reflects the normality assumption underlying the probit model (Greene, 2009).[30]

[30] A critical reader may question whether this assumption could have been motivated by for example the distribution of sample proportions across the response categories in the SAH variable (which could be graphically evaluated by a histogram or a table). However, recalling the modelling assumptions of these non-linear models, which aim to predict an unobserved, underlying variable, the sample proportions do not need to provide a picture of the proportion of the underlying variable. Nevertheless, the choice of normal distribution may be motivated by the central limit theorem and the fact that we are modelling human

Substituting eq. (14) into (13) may be rewritten as:

$$Prob(SAH_i = j|x_i) = Prob(\mu_{j-1} < x_i'\beta + \varepsilon_j < \mu_j)$$
$$= Prob(\varepsilon_j < \mu_j - \beta' x_i) - Prob(\mu_{j-1} - x_i'\beta),\ \mathrm{j} = 1{,}2 \ldots 5 \qquad (15)$$

Thus, the probability that the response category SAH_i is chosen is the probability that the latent health variable *Hi** is between two cut points or thresholds $\mu_{j-1}$and μ_j (Greene, 2009; Veerbek, 2012, p. 221). By assuming a standard normal distribution of ε as in the probit model, the implied probabilities may be written as below, where *F(·)* denote the cumulative density function (CDF) (where ϕ refers to the standard normal distribution):

$$Prob(SAH_i = j|x_i) = F(\mu_j - \beta' x_i) - F(\mu_{j-1} - x_i'\beta) = \phi(\mu_j - \beta' x_i) - \phi(\mu_{j-1} - x_i'\beta) \quad (16)$$

Or more generally:

$$Prob(SAH_i \leq j|x_i) = F(\mu_j - \beta' x_i) \qquad (17)$$

Following Jürges (2007), the vector x_i in eq. (17) will include several objective health (dummy) variables as explanatory variables, listed below in table 5 (descriptive statistics of all health variable used are found in table B2 in appendix B).

Table 5. Objective health variables available in SHARE (defined as "Doctor have told you have…").

o High blood pressure	o Parkinson
o High blood cholesterol	o Cataracts
o Stroke	o Hip fracture
o Diabetes	o Other conditions
o Chronic lung disease	o Alzheimer/ dementia
o Arthritis	o Depression
o Cancer	o Low grip
o Stomach/ duodenal ulcer	o No grip

In addition to the health conditions the respondent claims a doctor has diagnosed them with, we follow Jürges (2007) and include grip strength (coded into low grip strength and no grip strength) as a physical health measure, since grip strength has been shown to be able to predict functional limitations and disability. We also follow his specification by including a variable for depression

behaviour which is influenced by infinite numbers of factors (Greene, 2009). Also, since the probit and the logistic distributions are very similar except for a scaling term in the variance, the results are often very similar in empirical work (Verbeek, 2012, p. 201).

(measured on a four point self-assessed scale).[31] Lastly, Jürges (2007) also includes the variable BMI in his estimation, however, this would imply a loss of around 25000 observations in our analysis and we therefore choose to omit this variable.[32]

With regards to our final specification of our probit model, we need to decide whether the disability weights of the above objective health conditions should be allowed to differ across time and across countries. By estimating an ordered probit model with our panel data and including country and time dummies we may test which model is statistically preferred.[33] In his baseline regression, Jürges (2007) restricted the disability weights of each health state to be the same across countries. Recall however, that this analysis does not aim at estimating a cross-country model but rather the aim is to investigate within-country IRHI trends, thus we aim to estimate the most appropriate health index for within country comparisons mainly. For this purpose we may employ a homogeneity test. The test will be performed separately for testing the time and country dimensions.[34] The homogeneity test aims to test the hypothesis that the same parameter vector is valid for all *G* groups the pooled sample may be divided into, which in our case may be either countries or years. By estimating the model *G+1* times (for all groups separately and once for the pooled sample) and obtaining the log likelihood functions one may use the likelihood ratio test statistic to test the hypothesis (Greene, 2009).

$$LR = -2(\ln L_{pooled} - \sum_{g=1}^{G} \ln L_g) \tag{18}$$

The results of these tests will be commented on in the results section 4.1. Finally, by estimating this ordered probit model which have been described and motivated in this section (which will be based on maximum likelihood, where the probabilities enter the likelihood function (Veerbek, 2012 p. 222)), the cardinal health index will be generated by the linear prediction of the estimated coefficients for the above listed ill-health variables. Since the variables are coded to 1 in case the respondent do suffer from any of the health conditions, the prediction yield an index of ill-health

[31] Jürges (2007) included a variable that indicates whether the respondent has been treated for depression. Including this variable in our case would result in a large loss of observations. Instead we use the variable where respondents are asked to self-assess on a 4 point scale if and how depressed they are at the moment. By including this variable we increase the sample size by 10 000 observations.

[32] While we on the one hand acknowledge the possible importance of BMI as a determinant of self-assessed health, BMI is in older cohorts closely related to diabetes, cholesterol, arthritis and heart problems and is furthermore not per se a disease (Pfarr et al 2012).

[33] We do acknowledge that to evaluate whether disability weights do differ between countries and across time should also be accompanied by theoretical arguments as well. An argument for the former could be that country differences in health care affect the disability weights (Jürges, 2007). An argument for varying weights across time, may be that respondents with certain health conditions adapt and get used to them over time which affect the disability weight they indirectly assign to them (Pfarr et al, 2012).

[34] One might have tested this by including time and country dummies and the respective interactions between each dummy and health variable. However employing a random effects ordered probit model with all these interactions with our data set of around 100 000 observations requires very large computer capacity.

(where 1 equal near death). This bounded variable can of course be recoded to an index of health (by taking 1 minus the estimated ill-health index). Given the previously described properties of different inequality indices for health and ill-health variables, we will make use of both the health and ill-health version of the generated health index.

3.2.2.1.1 Imposing cardinality: some further motivation and a robustness check

In the literature there are different methods that can be employed to impose cardinality on ordinal responses, in addition to the ordered probit model, which may be appropriate to comment on. Van Doorsaler and Jones (2003) have written a well-cited review of the most common approaches, e.g. to dichotomize the ordinal variable or to employ strategies based on some kind of scaling or scoring algorithm, probit or interval regression methods. To dichotomize (i.e. to transform the variable into a binary variable) implies a great loss of information and has, more importantly, been shown to make comparisons of inequality over time unreliable (Wagstaff and van Doorslaer, 1994). Another common method is to take advantage of another generic scale (another continuous health variable) that has been constructed and validated in another context and to map this scale to the categories of the SAH variable used by the researcher. By employing this strategy, a stable mapping from the generic index to the latent variable that determines SAH is assumed and this relationship is furthermore assumed to be valid for all individuals. In practice, this implies that the rank by the generic index correspond to the rank by SAH. This strategy can be employed in different ways, but a simple method is by directly attribute the mean value of each SAH category assigned from one (continuous) index in a data set that contains both SAH and a utility index, for example the Canadian HUI[35], to the corresponding health category of the self-assessed health variable the researcher has access to. As long as the health categories are the same (e.g. poor, fair, good, very good and excellent) the mean HUI value can be directly applied to the corresponding category in the SAH variable and this strategy will be employed as a robustness check to the above results by probit model. However, this does not take into account for example within group differences based on gender and age groups (as with a dichotomizing approach this yield limited individual level variability in the health variable). Perhaps the most restrictive assumption is that this strategy assumes that the mean health in each category is the same in the sample used by the researcher (for example in all countries in SHARE) as in the sample providing the HUI scores (for example Canada in the NPHS survey, see footnote 35).

[35] The McMaster Health Utility Index Mark III (HUI) is a health status index that has been used to derive numerous different health states based on eight different attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain). The index takes a value between 0 and 1, where 1 implies perfect health (Feeny et al. 2002). In the Canadian National Population Health Survey (NPHS), individuals provided answers both on SAH and on each attribute of the HUI classification system so that each individual also was assigned an HUI score in addition to their answers on SAH.

That is, one assumes a similar health distribution (O'Donnell et al. 2008, p. 59; van Doorslaer and Jones, 2003).

A third strategy, using the interval regression approach, also requires the use of external data, however in this case the upper and lower limits of each SAH category interval needs to be known or assumed to be known.[36] Aiming to overcome the ordinal nature of the variable, all these methods use secondary data from a different population, which is however not required by the probit model approach. Nonetheless, in their review of the mentioned methods, van Doorslaer and Jones (2003) did conclude that the interval regression method seemed to outperform the probit model. However a few arguments may support the choice of the previously described probit estimation strategy: First, when evaluating the probit model, van Doorslaer and Jones (2003) refer to Cutler and Richardson (1997) and Groot (2000), however when validating this method by comparison to the interval regression, they do not use the same or a similar specification as these authors, despite the fact that the health index predicted from the probit model most likely is very dependent on the independent variables included (O'Donnell et al., 2008 p. 59).[37] Thus, we would argue that the validation of the probit model in van Doorslaer and Jones (2003) is very dependent on the particular specification used by them and we do not find these results convincing enough to discard the probit strategy all together. Secondly, employing the interval regression method by using the data from the NPHS study would imply using very old data (almost 20 years for the last wave in our sample) and we are not aware of any other available other data at this point in time. Third, as already stated, the probit model approach does have the advantage of avoiding the potential restrictive assumption that the relationship between the generic health index and SAH is the same across all countries. We therefore argue that the chosen method is motivated, although at the same time acknowledging that further validation and comparison between methods, in line with van Doorslaer and Jones (2003), seems to be needed.

[36] In short, this strategy implies to obtain the upper and lower category limits from a generic health index by computing the cumulative frequency of observations for each category of the SAH variable that is found in the same dataset as the generic health index. By matching these frequencies with the quintiles of the empirical distributions function (EDF) the upper and lower thresholds may be found. For a more detailed description, see van Doorsaler and Jones (2003).

[37] The only control variables included in van Doorslaer and Jones (2003) are socioeconomic variables such as education, marital status and age whereas in the specification used by Cutler and Richardson (1997) and Groot (2000) is similar to the one in this study (which is also the case in Jürges (2007), Schneider et al (2012) and Pfarr et al. (2012)).

3.3 Calculation of income related health inequality indices

3.3.1 Standardisation, sample weights and imputations

When calculating all our measures we will standardise for age groups and sex, as is standard in the literature (O'Donnell et al. 2008: 60). The motivation is that the measures otherwise may be misleading if there are other factors affecting health which are correlated with income. For example, the oldest group of individuals in a population is likely to report worse health but they are also likely to report a lower income. Thus, not controlling for these factors (i.e. standardising) may imply that the relation between income and health may be biased (Gravelle and Sutton, 2003). More specifically, in addition to controlling for sex by dummies, we standardise by age by including four age groups (50-54, 55-59, 60-69 and 70 and over).

Given the choice of using an unbalanced panel, the possibility for drawing inferences about the target population is contingent on the sample being representative and appropriately collected. However, different observations in SHARE have different probabilities of selection and therefore the researcher needs to weigh each observation in the analysis to generate unbiased estimates of parameters of interest and to mitigate the effect of non-response (O'Donnell et al, 2008; SHARE, 2013).[38] Given the descriptive nature of this thesis, in contrast to traditional economic literature aiming to investigate causal relationships, this provides motivation to employ sample weights in the estimation (for a recent, informative guide on when to use sample weights, see Solon et al (2015)). As a baseline design for our calculation we will therefore employ calibrated sample weights, provided by SHARE to be used to reduce selectivity bias, in the calculation of the covariance, the mean of the health variable and the fractional rank as suggested by O'Donnell et al (2008). However, unweighted indices will also be calculated as a robustness test.

As noted briefly above we will make use of imputed data on the income variable. SHARE provides five complete but slightly different datasets since each imputed observation is a unique draw from a probability distribution, conditional on the observed data. As previously mentioned, the imputed data is based on several survey questions on different income sources and for those individuals where a response has been given to all observations, no imputed observations are created, thus the observations for these individuals are the same across all the five data sets. An average estimate is then computed for the estimated inequality indices and their standard errors

[38] It can be noted that different countries have different types of sample design in SHARE. The sample weights available in SHARE takes into account the respective design. For more detailed information on the SHARE sampling procedure and weights design, see Alcser et al (2005).

using the imputation strategy by Christelis (2011). A more detailed description of this method can be found in section C1 in Appendix C.

3.3.2 Estimating (extended) C, W and E

In section 2.1.3 the formal concepts and equations of the indices were described. For the purpose of estimation these equations can however be rewritten and it is convenient to note that C can be expressed in terms of the covariance between the health variable *h* and the income rank variable *r*. We will therefore employ a method that has been referred to as the convenient regression model to estimate these indices. For the standard C, W and E, the convenient regression models are outlined in Appendix C2. However, in order to modify the inequality aversion parameter in these indices, which are outlined explicitly in the Extended Concentration Index, the basic convenient regression models as depicted in Appendix C2 can be generalised as to estimate the extended version of C as the following (O'Donnell et al., 2008):

$$-v \cdot var\{(1-\mathrm{R})^{(v-1)}\}\left\{\frac{\boldsymbol{H_i}}{\boldsymbol{\mu_H}}\right\} = \alpha + \beta_c \cdot (1-R)^{(v-1)} + \varepsilon_i \tag{19}$$

where v represents the inequality aversion parameter as seen in eq. (10) in section 2.1.3.6, H_i is the health variable, μ_H is the mean of the health variable and R is the income rank variable. The Extended Concentration Index is equal to the slope coefficient β. Recall that the Extended Index makes different weighting schemes in the ranking of individuals' health explicit, which are "hidden" in the weighting function, as presented in the introduction (see eq. (1)). It can furthermore be seen by comparing eq. (6), (8) and (10) that the standard weighting function is identical in the indices C, W and E. Therefore we may make use of the generalised convenient regression also for W and E, in order to modify the implicit inequality aversion parameter also made in these measures. By noting the similarities between C and W (compare eq. (5) and (8), where the difference is found in the denominator of the normalising function), the above generalised convenient regression model may be rewritten to estimate the extended version of W using the following equation:

$$-v \cdot var\{(1-\mathrm{R})^{(v-1)}\}\left\{\frac{\boldsymbol{H_i}}{(\boldsymbol{1-\mu_H})\times\boldsymbol{\mu_H}}\right\} = \alpha + \beta_c \times (1-R)^{(v-1)} + \varepsilon_i \tag{20}$$

Following Erreygers et al (2012), extended E can furthermore be estimated using the following equation:

$$-v \cdot \frac{v^{\overline{(v-1)}^{\,v}}}{(v-1)}(var\{(1-\mathrm{R})^{(v-1)}\}\{\boldsymbol{H_i}\} = \alpha + \beta_c \times (1-R)^{(v-1)} + \varepsilon_i \tag{21}$$

These indices will be calculated for v equal to 1.5, 4, 6 and 8. The choice of these weights will imply a stepwise increase from the standard value of 2 weighting the health of the relatively poorer individuals to the richer individuals up till the point of where the health of the top half of the population goes to zero, as was described in section 2.1.3.6. By using the value of 1.5 we also impose the value judgement that inequality is less of a concern than implied by the standard measure (using the standard value of v equals 2, recall that a value of 1 would imply that we do not care about inequality and all the resulting indices would be zero).

Lastly, as we have weighted data, the standard procedure of calculating standard errors of the concentration index cannot be used. Two alternative methods are proposed by O'Donnell et al (2008). First, one can simply use the standard error of the estimated rank variable to represent the standard error of the inequality index. This makes sense since the coefficient of the rank variable is the respective measure, and hence the resulting standard error can be seen as the standard error of the index as well. However, the limitation of this is that the standard error of the estimated rank coefficient does not take into account the sampling variability of the estimate of the health variable mean. The authors suggest the delta method for the standard error, where the sampling variability of the mean is taken into account. Nevertheless, these two methods seem to produce almost identical results, and the authors conclude that standard errors from the convenient regression method can be used and should provide accurate results. Thus, when using the non-imputed income variable, we present standard errors and significance for the respective index from the slope coefficient in the regression. When using imputed income data, we use the same standard errors but take into account the variance in the imputation process by the method described in appendix C, section C1.

3.4 Summary

In order to be able to provide an accurate description of health inequality trends, not only does the researcher need to consider implicit value judgements for the inequality indices as was described in chapter 2, but this chapter has also described the various methodological considerations needed in the empirical estimation procedure. By generating a health index, specifically developed for our data, and by motivating a careful research design including sample weights and imputed data to deal with sample selection and measurement errors, we aim to make a thorough evaluation of IRHI trends in Europe. To sum up and to provide an overview of the results to be presented, the table below (table 6) summarizes the indices that will be calculated and commented on in the next chapter, aiming to investigate trends in IRHI in Europe.

Table 6. IRHI indices employed in this paper.

Inequality aversion weights/ Indices	C (modified)	W	E
V=1.5	✓	✓	✓
V=2	✓	✓	✓
V=4	✓	✓	✓
V=6	✓	✓	✓
V=8	✓	✓	✓

Chapter four

4 RESULTS

4.1 The health index

In section 3.2.2.1 we described the estimation strategy in order to conclude on a suitable specification of the ordered probit model employed to impose cardinality on SAH. As part of our strategy, we employ a homogeneity test in order to check whether the health index should be estimated separately for every country and time period. First, the hypothesis that the same model applies to all countries is tested. Second, we test homogeneity across the three points in time. Both these hypotheses are rejected.[39] These results support a specification that yield country and time specific health indices. The implication is that we, in contrast to Jürges (2007), allow both thresholds and coefficients to differ between countries. The final country specific equation employed in this study can be expressed as the above eq. 17 (see section 3.2.2.1), calculated for 33 separate subsamples. The results from the ordered probit model regressing the categorical SAH variable on various, objective health variables are presented in appendix D. In table D1 the country specific regressions for wave 2 (2006) are found and in table D2 and D3 the corresponding regressions for wave 4 (2011) and wave 5 (2013) are presented.

In these tables, each coefficient corresponds to the estimated contributing "weight"[40] of certain health conditions on self-assessed health, in the respective country, at three points in time.[41] All coefficients have the expected, positive signs (the SAH variable is coded between 1 and 5, where 5 equals 'very poor' health) where the coefficients are significant. Following the formal language

---

[39]Testing the hypothesis that the same model applies to all countries yields a LR statistic of around 7582 (the log likelihood function for the pooled model is -146960.5 and the sum of the likelihood functions from all country specific estimations is -143169.39). Given the critical value of 213 taken from a chi statistic table (with 180 degrees of freedom and alpha 0.05) the hypothesis of the pooled model is rejected. The LR statistic testing homogeneity across the years is -9290 and given the critical value of 36 this hypothesis is also rejected.

[40] After normalising each coefficient (thus restricting each coefficient to be between 1 and 0), Jürges (2007) refers to these coefficients as disability weights. However, given that sign and relative magnitude between the coefficients can be evaluated directly by the regression output and since the construction of the health index is of only secondary importance for this analysis, we do not present these normalised coefficients.

[41] Given the aim of our study, we will not put particular emphasis on interpreting the exact impact of the coefficients on the respective SAH categories. Note however as with all non-linear models, the coefficients of the regression cannot be interpreted directly as marginal effects, only the sign holds valid information. Furthermore, with the ordered probit models, if one conceptualise the model in terms of probabilities, the positive effect of the coefficients can only be directly interpreted for the highest category of SAH (the lowest health state in our case) and the opposite effect for the lowest category (the highest health state in our case), whereas not even the sign of the coefficient could be directly interpreted for the intermediate categories (Veerbek 2012, p. 210-211, 221-222). Nevertheless, following Jürges (2007) we will not pursue a more detailed interpretation of the respective coefficients, except for what will follow in the text.

of section 3.2.2.1, this implies that the latent variable of ill-health H^* (interpreted as ill-health due to the coding of SAH) increases if the individual suffers from any of the included health conditions (since these variables are binary). In general, the magnitudes of the impacts of the respective health conditions are similar to the coefficients in Jürges (2007), although the coefficients differ somewhat between the countries, indicating that the restriction of homogenous disability weights across countries in Jürges is restrictive (which again, has been supported by the likelihood ratio test). More specifically, it seems that relatively more grave diseases such as stroke, cancer and Alzheimer's disease have very different impact on self-assessed health in different countries. These differences may partly be explained by differences in health and elderly care. Other conditions, such as high blood pressure and cholesterol seem to have more even impacts on general health across countries and these are also, as expected, in the lower end of relative magnitude when comparing the impact of the different health conditions. What also seems to be plausible is that having had a heart attack, having chronic lung disease, and Parkinson's disease yield a relatively larger coefficient compared to for example having high cholesterol.

As a linear prediction of these disability weights, a health index has been generated for each respondent. In the below table 7 the mean values per country and per wave are presented where the Netherlands has the highest value in wave two and Austria has the lowest. In wave five Italy has the highest value and Germany has the lowest. Despite differing coefficients for some health conditions, the mean values are very similar, over time and between countries. This will have an impact on our results, as we will see later on.

Table 7. Mean values of the estimated health index

	Wave 2		**Wave 4**		**Wave 5**	
	N	**Mean H**	**N**	**Mean H**	**N**	**Mean H**
Austria	1187	0.791	5087	0.833	3970	0.815
Germany	2526	0.807	1548	0.786	5632	0.811
Sweden	2673	0.810	1919	0.822	4435	0.843
Netherlands	2609	0.858	2697	0.813	4084	0.856
Spain	2162	0.823	3341	0.808	6115	0.831
Italy	2950	0.831	3505	0.852	4578	0.871
France	2819	0.845	5626	0.845	4391	0.854
Denmark	2581	0.842	2237	0.801	4085	0.863
Switzerland	1447	0.853	3722	0.850	2991	0.854
Belgium	3128	0.851	5204	0.853	5532	0.842
Czech Rep.	2780	0.812	5946	0.849	5578	0.841

As a robustness test of our generated variable we also generate a health variable by the direct mapping method described in section 3.2.2.1.1 (summary statistics of the resulting health variable are presented in table D4 in Appendix D) and compare it with mean health values of some countries predicted by the health index (see column 7 and 8 in table 8 below). We found that the latter method generates slightly lower mean values; however, the overall results are similar. However, as mentioned in section 3.2.2.1.1, the validity of this application depends on how similar the sample distribution in the used data set (in our case SHARE) is to the sample distribution in the sample that provides a cardinal health variable (in our case NPHS). As can be seen from the below table, the distribution in the NHPS is skewed, whereas for example Sweden and Italy as well as the SHARE average have a more normal distribution.[42] This weakens the validity of using this method and given the previous presented motivation for the probit model we opt to use the predicted health index for the inequality indices. The two last columns show the means of the predicted health index per SAH category for Sweden and Italy. We can see that for Sweden the values are very similar to the mean HUI values, while the values are slightly higher for Italy. Because the values are fairly similar to the values found in the NPHS, it further validates the use of the predicted health index over the simple mapping method.

Table 8. Distribution of sample proportions of SAH and HUI score in the SHARE survey and NPHS survey.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SAH categories (same in NPHS and SHARE)	**Sample proportion (%) SHARE (11 European countries, 2006)**	**Sample proportion (%) SHARE (Italy, 2006)**	**Sample proportion (%) SHARE (Sweden, 2006)**	**Sample proportion (%) NPHS (Canada, 1994)**	**Mean HUI per SAH category (from NPHS)**	**Mean health index per SAH category (Italy, 2006)**	**Mean health index per SAH category (Sweden, 2006)**
Poor	9.23	12.93	6.36	2.4	0.557	0.660	0.584
Fair	24.88	31.07	22.04	8.6	0.758	0.780	0.722
Good	37.75	35.84	30.75	27.0	0.876	0.883	0.801
Very good	18.47	12.90	23.23	37.2	0.923	0.923	0.883
Excellent	9.67	7.26	17.62	24.8	0.947	0.930	0.918

Source for the mean values from the NPHS survey: Doorslaer and Jones (2003).

[42] We believe that this may be caused by the fact that the SHARE survey only includes individuals aged 55 and above, whereas the NHPS survey include all ages.

4.2 Income related health inequality trends

The research question of this essay reads:

- *Has income related health inequality increased over time in Europe?*
- *How sensitive are these trends to the choice of inequality measure?*

Given the first question, to investigate the trends in IRHI, we will naturally focus on commenting the results in terms of direction of the indices. Further comments will be made on the sign and relative strength (i.e. magnitude) of the indices. As to answer the second research question, to evaluate the sensitivity of these trends, two dimensions (recall table 6 above) will be in focus of the description of results: a) any possible differences in trends retrieved by different indices (holding the inequality parameter *v* constant) (section 4.2.1) and b) any possible differences in trends retrieved by different values of the inequality parameter *v* (focusing on the same index across all values of *v*) (section 4.2.2). The first dimension refers to a horizontal comparison of the different indices, given a value of *v,* as presented in table 6, and the second dimension refers to a vertical comparison between the different values of *v,* given a certain health index. Again, in addition to contribute to the rather scarce literature on trends in IRHI, this thorough evaluation of IRHI trends has not been done elsewhere and constitutes the main contribution of this essay.

Specifically, the inequality indices are presented per country, per wave, i.e. per year (recall that wave 2, 4 and 5 correspond to year 2006, 2011 and 2013). The baseline results for the Concentration Index (C), Wagstaff's normalised Concentration Index (W) and Erreygers' corrected Concentration Index (E) are found in table 9 below, with the standard assumption of inequality parameter (*v*) set to 2. Since the C does not satisfy the mirror property for bounded variables, C is presented for both a health and an ill-health variable denoted C(h) and C(s). These baseline results are calculated using weighted data and imputed income, as previously has been described. For all indices, the sign of the index shows whether the distribution of health is pro-rich or pro-poor, and the magnitude of the index indicates how the strength of the pro-rich (or pro-poor) the distribution is relative to other countries or other points in time. As will be explained later, the magnitude of an inequality index can only be compared between countries for the same index.

By looking at table 9 one can see that good health is concentrated amongst the rich in Europe, as the indices are positive for C(h), W and E. The index C(s) measures IRHI in ill-health and the negative sign implies that ill-health is concentrated among low income earners. This means that income-related health inequalities are present in all European countries included in this study.

Overall we find a strictly *decreasing* trend of income-related health inequality for Sweden, Switzerland and Belgium, for all indices and all inequality aversion parameters (*v)* (except for a stable trend in Belgium with the equality parameter set to 8, see further comments in section 4.2.2.). That is, all indices generate the same pattern of decreasing inequality from wave two to wave five. Furthermore, we find an overall decreasing trend for Austria and the Czech Republic, where the indices decrease from wave two to wave five and from wave two to wave four, but the indices in wave four are lower than the indices in wave five. Thus, based on the calculated indices for these three data points, it seems to be difficult to establish the magnitude of the inequality in some countries around 2011-2013, nonetheless, given the relatively larger difference between 2006 and 2011-2013, it seems that the trend in IRHI is decreasing in these 5 countries, no matter the inequality index or value of inequality aversion parameter.

Consistently *increasing* trends are found in Germany and the Netherlands for all indices and inequality parameters (focusing on the difference between 2006 to 2011-2013). There also seems to be an overall increasing trend in inequality in Spain, though much more limited in scope and with a stagnating trend between 2011 and 2013, so the results should be interpreted very cautiously. More importantly, the indices, and thus the trend, are not significant when using the inequality parameter of 4 and above, which will be further commented on in section 4.2.2.

4.2.1 Evaluating the inequality trends with different inequality indices

The indices C, W and E deal differently on how to measure inequality when the health variable is bounded, which also implies that they entail different normative value judgements (i.e. different IECs as described in section 2.1.1). Due to this, the absolute magnitude of the respective indices cannot be compared (since they measure different things), while the trend direction can. Since different IECs imply that different kinds of changes of health preserve the given inequality, a certain change of health over time, may be 'interpreted' differently by for example the absolute index E and the mirror relative index W. However, although a change in health may imply opposite effects on the indices, it does not exclude that both indices may decrease (increase) by a certain change. Thus, from a theoretical point of view, what impact has the choice of these different indices on the estimated IRHI trends? Or equivalent, looked upon from the empiricists perspective, will the trend differ depending on the measure chosen?

Given the data and variables employed in this study, the impact of the different indices on the estimated direction of the trend seems to be limited. However, the magnitude of change can be seen to be different, as will be commented on further below. When examining the indices over time for the different indices, the results, seen in table 9, suggest that most measures indicate the same direction of change in inequality over time in the respective countries. In other words, when the distribution of health across the socioeconomic ranking of individuals have changed in these countries over this particular period of time, the direction in absolute (E), relative (C(h) and C(s)) and mirror relative (W) inequality seem to be the same same and both absolute and relative inequality have increased or decreased in the respective countries with increasing or decreasing trends. All estimates are also significant. There is however at least one exception to the consistent trend found by the indices. In Denmark the change between 2006 and 2011 varies between indices: using C(h) and E the change in distribution implies an increase in inequality while the opposite is depicted by C(s) and W. Since E is independent of mean health (see eq. (10)), the increase in E suggests that the association (covariance) between the fractional rank and the health variable has increased. However, between 2006 and 2011, mean health in Denmark fell from 0.84 to 0.8, which implies that mean ill-health increased. Given that C is level dependent (mean health enter the equation (5) in the denominator), this change may contribute to the fact that we see C(h) increase, while C(s) decrease (mean health in the denominator has decreased but mean ill-health has increased, thus possibly "dominating" the effect of the increase in covariance between the health variable and the rank variable). A pattern where the indices E and C(h) yield a similar direction of change and slightly different to the changes indicated by W and C(s), during the same time period, is also seen in the Netherlands and Spain. In these countries the overall trend (i.e. the change from wave 2 to wave 4 and the change between wave 2 to wave 5) indicates that the level of inequality in 2011-2013 has increased compared to 2006. However, the change between 2011 to 2013 yields a stable or on the margin decreasing level of inequality for the former two indices while an increasing level of inequality is depicted by the latter two indices for this period. This pattern has been found in previous studies investigating the impact of the indices on rankings, when the average health level is high in the sample countries (Kjellsson and Gerdtham 2014). Recalling the theoretical section, this may be explained by the fact that W tend to reflect C(h) or C(s) with the lower prevalence. Since mean of health in this sample is around 0.8 (see table 7 above), the ill-health has the lower mean (0.2), thus we see this grouping trend between the indices (Kjellsson and Gerdtham 2014; van Doorslaer and van Ourti 2011).

In terms of magnitude of inequality and trend, recalling the theoretical section and the respective equations for each index, it was highlighted that W and E coincide when the average health level of the bounded variable is 0.5 and the larger variability of the mean of the health variable, the greater the divergence by the respective measures. This can partly explain the results of few differences between the trends depicted by W and E using this data. First of all, given that mean health in all countries are above 0.5, we see in table 9 below, that the magnitudes of the respective indices are different. Second, since mean health within the countries is very stable across the period studied, this implies that the difference between the W and E will be "constant". It can be understood by recalling the respective equations (eq. (8) and (10)) for the indices; For each year it is only the normalising function (the first term of the respective definitions of the indices) that differ between the indices. Given that the mean health is close to constant across all years, the difference between these measures is constant for each year (would the mean health variable instead vary substantially for each year, this would affect W but not E, as seen from these equations). Thus the "ranking" of the distributions across time by W and E will be similar and thus will the direction of trend be similar.[43] Kjellsson and Gerdtham (2013) show that low variation in the health variable causes the inequality rankings of the countries to be fairly stable over indices and this is also the case with our results, if the results in table 9, instead is assessed based on the trend in ranking, see table D5 (in appendix D). Rankings within countries over the three points in time do change when comparing within each index, due to the changes in distributions over time. However, when looking at how the rankings change over time between different indices, we can see that the different indices give fairly similar rankings over time. Nevertheless, even though all indices in table 9 for example indicate an increasing trend for Spain, this ranking table highlight what different messages different indices still may give: according to C(h), Spain rises in ranking by four positions (the higher ranking, the higher is the inequality relative to the other countries), whereas with C(s), Spain rises only by one step. This picture is clearly different depending on what measure one uses, despite the fact that the depicted trend by the indices indicate the same direction. There is furthermore, as shown in previous research, examples of different ranking by different indices for separate years (see for example the different positions of Switzerland and Austria using W and E respectively). Again, the pattern where the indices divide in two groups (C(h) with E and C(s) with W, respectively) can be seen.

[43] More precisely, recalling the formal definition for W, see eq. (8) in section 2.1.3.3, mean health enter the first term in the definition (the normalising function), in contrast to E. This is why a different level of prevalence across the years would have an impact on W, while not on E and thus possibly cause different trends (E has the property of level independence).

Similarly to W and E, C(s) and C(h) will, of course, coincide when mean health is 0.5 (*ibid*). However, the fact that they do not yield similar levels of inequality for our sample, in which mean health is around 0.8, is linked to the debated mirror property (recall that if they would have satisfied this property, the indices would have had the same value but with opposite signs). As have been described previously, the reason is because of the relative value judgment within C: when the health prevalence is high (as in our sample) the same absolute difference will be in relative terms lower compared to the same absolute differences in a sample where health prevalence is low (Clarke et al., 2002). Therefore, the magnitude of inequality estimated by C(h) is lower than C(s) since C is level dependent (as is all pure relative measures). What can be noted from the theoretical section is though that for C(h) and C(s) to indicate the same magnitude in change (i.e the same relative change), the mean value for the two points in time does not have to be equal to 0.5, what matters is that the mean is the same across the two points in time compared (Erreygers, 2009a). This is crucial in order to understand our results of similar trends between C(h) and C(s) as well. If the mean would have been substantially different between the years, the relative change over time would have been different, depending on the use of C(h) or C(s), which could have implied different trends.

Unless there would be a normative argument for while levels of inequality is more important to measure in health or in ill-health, W and E have been suggested to be the most appropriate indices for bounded variables. Nevertheless the results from this empirical application provides an example of when the differences in trends (not magnitude) still may be similar depending on the level of health prevalence. Thus, although finding a few exceptions, we find few notably contradicting results when using different inequality indices.

4.2.2 Evaluating the inequality trends with different inequality aversion parameters

In order to further investigate the sensitivity of the trends found in self-assessed health in Europe, we investigate the trends depicted when altering the inequality parameter v. Tables D7, D8, D9 and D10 (found in appendix D) show the results of all indices calculated for four different inequality aversion parameters, where v equal 1.5, 4, 6 and 8. As stated earlier, v equal to 2 gives the standard C, W and E and these results were presented above and in focus when we compared the trends depicted between the different indices. Increasing the value of v above 2 will give a higher weight to the health of individuals in the bottom of the distribution compared to the standard indices, thus the interpretation of increasing aversion towards inequality.

Table 9. IRHI trends in Europe, inequality aversion parameter = 2 (standard C, W and E).

Countries	Overall trend	C(h)			C(s)			W			E		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5
AUSTRIA	↘	**0.017*** **	**0.007*** **	**0.012*** **	**-0.063*** **	**-0.033*** **	**-0.055*** **	**0.080*** **	**0.040*** **	**0.068*** **	**0.053*** **	**0.022*** **	**0.040*** **
		(0.004)	(0.002)	(0.002)	(0.016)	(0.009)	(0.011)	(0.021)	(0.011)	(0.013)	(0.014)	(0.015)	(0.016)
GERMANY	↗	**0.009*** **	**0.019*** **	**0.020*** **	**-0.035*** **	**-0.070*** **	**-0.081*** **	**0.045*** **	**0.089*** **	**0.100*** **	**0.030*** **	**0.061*** **	**0.063*** **
		(0.003)	(0.004)	(0.002)	(0.013)	(0.015)	(0.007)	(0.016)	(0.019)	(0.009)	(0.011)	(0.012)	(0.013)
SWEDEN	↘	**0.020*** **	**0.014*** **	**0.010*** **	**-0.084*** **	**-0.073*** **	**-0.054*** **	**0.103*** **	**0.087*** **	**0.063*** **	**0.064*** **	**0.048*** **	**0.032*** **
		(0.004)	(0.003)	(0.002)	(0.016)	(0.016)	(0.010)	(0.020)	(0.019)	(0.012)	(0.012)	(0.013)	(0.014)
NETHERLANDS	↗	**0.006** **	**0.017*** **	**0.017*** **	**-0.033** **	**-0.076*** **	**-0.099*** **	**0.039** **	**0.093*** **	**0.116*** **	**0.020** **	**0.056*** **	**0.057*** **
		(0.003)	(0.003)	(0.002)	(0.016)	(0.013)	(0.012)	(0.019)	(0.016)	(0.014)	(0.015)	(0.016)	(0.017)
SPAIN	↗	**0.005* **	**0.008*** **	**0.008*** **	**-0.027* **	**-0.035*** **	**-0.038*** **	**0.032* **	**0.044*** **	**0.046*** **	**0.018* **	**0.027*** **	**0.026*** **
		(0.003)	(0.003)	(0.002)	(0.013)	(0.011)	(0.011)	(0.015)	(0.013)	(0.013)	(0.012)	(0.013)	(0.014)
ITALY	**stable**	**0.006*** **	**0.007** **	**0.005*** **	**-0.031*** **	**-0.040** **	**-0.033*** **	**0.037*** **	**0.047** **	**0.037*** **	**0.021** **	**0.023** **	**0.016*** **
		(0.002)	(0.002)	(0.002)	(0.012)	(0.014)	(0.012)	(0.015)	(0.016)	(0.014)	(0.013)	(0.014)	(0.015)
FRANCE	**stable**	**0.009*** **	**0.013*** **	**0.008*** **	**-0.049*** **	**-0.069*** **	**-0.052*** **	**0.059*** **	**0.082*** **	**0.060*** **	**0.031*** **	**0.023*** **	**0.029*** **
		(0.003)	(0.002)	(0.002)	(0.015)	(0.010)	(0.010)	(0.018)	(0.011)	(0.011)	(0.016)	(0.017)	(0.018)
DENMARK	–	**0.020*** **	**0.023*** **	**0.015*** **	**-0.102*** **	**-0.090*** **	**-0.093*** **	**0.122*** **	**0.113*** **	**0.108*** **	**0.067*** **	**0.073*** **	**0.052*** **
		(0.003)	(0.003)	(0.002)	(0.017)	(0.013)	(0.011)	(0.020)	(0.017)	(0.013)	(0.013)	(0.014)	(0.015)
SWITZERLAND	↘	**0.014*** **	**0.010*** **	**0.008*** **	**-0.078*** **	**-0.057*** **	**-0.047*** **	**0.092*** **	**0.067*** **	**0.055*** **	**0.047*** **	**0.034*** **	**0.027*** **
		(0.003)	(0.002)	(0.002)	(0.015)	(0.011)	(0.012)	(0.018)	(0.013)	(0.014)	(0.014)	(0.015)	(0.016)
BELGIUM	↘	**0.011*** **	**0.010*** **	**0.009*** **	**-0.064*** **	**-0.057*** **	**-0.050*** **	**0.075*** **	**0.067*** **	**0.059*** **	**0.039*** **	**0.034*** **	**0.031*** **
		(0.002)	(0.002)	(0.002)	(0.012)	(0.010)	(0.010)	(0.014)	(0.012)	(0.012)	(0.017)	(0.018)	(0.019)
CZECH REP	↘	**0.016*** **	**0.007*** **	**0.007*** **	**-0.065*** **	**-0.041*** **	**-0.042*** **	**0.082*** **	**0.048*** **	**0.049*** **	**0.052*** **	**0.024*** **	**0.025*** **
		(0.004)	(0.001)	(0.002)	(0.017)	(0.008)	(0.009)	(0.021)	(0.010)	(0.011)	(0.014)	(0.015)	(0.016)

Standard error in parentheses, *p<0.1 **p<0.05 ***p<0.01

Decreasing the value of v implies less weight on the health of the poorest, however as soon as the weight is larger than 1, the weight of the population ranked by income above the 55th percentile decreases.

Although the use of different inequality measures instead of graphical analysis by histogram or Lorentz/concentration curves, was motivated earlier in order to facilitate comparisons between countries or years, we will complement this section by some graphical analysis in order to better understand what drives the results.

Overall the results presented in tables D7, D8, D9 and D10 suggest that when altering v, the direction of trend over time does not change for most countries, no matter the index. One exception from this general conclusion is however possible to discern, furthermore, the impact on the inequality levels when altering v, varies. The Czech Republic shows overall the same trend, where the level of inequality is higher compared to 2011-2013, for all indices and weights. Notable is that the decreasing trend in Czech Republic seems to become more steep when increasing aversion towards inequality, mostly driven by the fact that the inequality level for 2006 increases relatively more than inequality in 2011-2013. In fact, the level of inequality in 2011-2013 just slightly decreases for v above 4. That is, given a larger weight on the health of the very poorest (and thus less on the upper part of the distribution) seems to not generate any larger inequality in 2011 to 2013 for this country. Similar patterns arises for Austria which as Czech Republic has an overall decreasing trend for all years and indices (i.e. 2011 and 2013 are both lower than 2006) however for v equals 6 and above, the inequality level in 2011 starts to decrease for all indices. This uneven impact of raising v on the inequality levels for different countries over time is also seen for Belgium. Belgium is an interesting case since this effect implies that the trend seen for v equals 2 tends to be on the margin reversed when v equals 8 (to the least stable, that is, at least less negative than what was suggested with *v e*quals 2*)*. With v equals 2 the overall trend was negative in Belgium (more negative for W and C(s) than for C(h) and E). For v equal 4, all indices increase (except for E which will be commented on below) but for v equals 6 the trend seems to stabilize (the differences between the waves get smaller). When v is raised to 8, it is especially wave 5 for W and C(s) that increases which could suggest a slight increase over time in IRHI in Belgium. For France, the same trend as with v equals 2 appears, although by raising v the trend tends to go from stable (or without the possibility to determine the direction of the trend) to a trend that may suggest a slightly increasing inequality from 2006 to 2011-2013.

We also find an additional example of indices suggesting opposite direction of changes when altering v, similar to what was found for Denmark with the standard weight 2. The trend in

Sweden is overall decreasing across *v,* however for *v* equals 8 the change between 2011 to 2013 implies less inequality for C(h) and E, but an increase in C(s) and W. This change in direction is though due to very small changes, thus, it is more an indication of the possibility for these indices to yield different results.

Denmark is ranked as having among the highest inequality in terms of magnitude of the indices (recall that we cannot compare magnitude between indices only within indices, but this ranking is very similar across indices, see table D5 for ranking in 2006). The magnitude of the indices more than double, across all years for all indices, in Denmark when raising *v* to 8. This large increase, in contrast to the above described different impact of raising *v* on the inequality levels, implies that there is a relatively larger difference between the poorest group in this country compared to the rest of the sample, and compared to other country populations for which the index of C does not increase as much.

This varying impact of changing *v* on the inequality indices is directly linked to the distribution of health over income. The original extended index (C) developed by Wagstaff (2002) which was described in section 2.1.3.6, is expected to increase with raising *v.* However, this particular result is also dependent on the distribution of health and income, and more specifically if the distribution of health is monotonic in income. If one observes a distribution where for example the richest group is not the healthiest, or where the poorest group actually is healthier than the second and third poorest decile, then increasing *v* to more extreme values, such as 8, may not yield as large impact on the level of inequality as for monotonic distribution. In extreme cases it may even cause the inequality index to change sign (Wagstaff 2002; Erreygers et al. 2012). Thus, these results are driven by our sample and its distribution.

By a graphical analysis we might get an indication of what is driving the results for the slightly different result for e.g. Belgium. What can be seen in figure D1 and figure D2 (appendix D) is that both the poorest and the richest deciles in Belgium seems to have worse health in 2013 compared to 2006. With the standard index, relative inequality however seem to decrease slightly according to table 9 above. By raising *v,* further weight is however put on the poorest getting worse health and by raising *v* high enough, less consideration is put on the worse health of the upper part of the distribution, which in the case of Belgium seem to imply that relative inequality instead starts to increase. However, the four richest deciles still have better health than the remaining six, which causes the indices to imply that the distribution of health still is pro-rich. In another graph (see figure D4 in appendix D) we may further see that the health index is not monotonic in income in the Czech Republic for 2011, a year the indices did not increase as much as for the year 2006 when increasing *v*. In figure D4 in appendix D, the distribution for the Czech Republic is shown which indicate that the poorest decile has slightly less ill-health than the second poorest group.[44] Furthermore, with regards to the distribution in Denmark, the relatively high rank as well as large increase in level of inequality when increasing *v* may be explained by looking at figure D3, where the distribution of health is monotonic in income and the poorest decile has relatively high ill-health (i.e. less health) compared to for example the distribution of the Czech Republic in figure D4. From these graphs it may also be noted that the overall variability of the health distribution across the income deciles are rather limited in our sample.

Thus, the levels of inequality rise as expected for most countries for the relative and mirror relative measures. However, one of the more notable findings, yet not commented on, is that in contrast to these measures (C and W), raising *v* implies that the absolute inequality is reduced according to the index of E. For example by raising *v* to 4 most values for E decreases for most countries, for example in German and Sweden. However an increase in absolute inequality is found for single years for Denmark, Switzerland and France, which implies that changing *v* may change the ranking of countries for certain years. These results are similar to the findings of Erreygers et al. (2012) when ranking developing countries using the extended E and under-five mortality levels. The level of absolute inequality also decreased for most of these countries, although the impact varied. However, given our focus on the trends of IRHI over time we can conclude that no trends do change when using E and altering *v*. Nonetheless, given that one would be interested in the magnitude of inequality, it is clear that altering the choice of *v* yield very different findings depending on the measure one chooses. Furthermore, although any change of direction of trend was hard to depict with our data, the re-ranking commented on occasionally in this text suggest that a trend change also could be possible.

Lastly, to be noted is that increasing *v* implies that some significance of the estimates is lost. This applies to Spain and Italy, for *v* equals 4 and above. Furthermore, when increasing the weight to 6 and 8, all indices in wave 2 for Germany and the Netherlands are insignificant. For Germany, this means that we cannot detect any consistent (increasing) trends any longer which can be seen with *v* equal to 2 and 4; different indices yield different trends between 2011 and 2013 for *v* equal 6 and 8. For Austria and the Czech Republic, insignificant indices are found for wave 4 for *v* equal 6 and 8, however as with the Netherlands the same trends (decreasing) can still be detected

[44] It should be acknowledged that this pattern of distribution is also prevalent for the other two waves, however, the difference between the level of ill-health for the poorest group and the upper part of the distribution in 2011 is significantly less compared to the other waves.

from 2006 to 2013. Lastly we note that neither Wagstaff (2002) nor Erreygers et al. (2012) present any information on significance of their extended indices, thus on this matter our results are hard to compare.

4.2.3 Robustness tests

In the beginning of our analysis we investigated if and how the use of unweighted data and an non-imputed income variable changed our results for our base case using the standard value of *v*. For the sake of brevity these results are not presented, however a few comments on these will be made. Using unweighted data did change some of the inequality indices for some countries whereas other countries did seem to be unaffected. More importantly, the overall trends previously described remain. Given our previous motivation, weighted inequality indices were however our favourable specification and these results gave no particular reason to why we should re-consider this. On the other hand, using the non-imputed income variable instead of the imputed one showed that the trend in Austria, Belgium and Germany actually was opposite to the imputed based trend. In the other countries trends were either consistent with the above results or lost significance due to the large loss of observations. This further underlines the need for cautious conclusions of the above presented results in the empirical trends in IRHI in some countries in Europe.

Chapter five

5 DISCUSSION

The aim of this thesis was to investigate whether a) IRHI has increased in Europe over time and b) whether these trends may be sensitive to the choice of inequality index. With regards to the first research question we find rather consistent increasing trends in Germany and the Netherlands, while decreasing trends in Sweden, Switzerland and Belgium between 2006 to 2013. The trend in Austria and the Czech Republic also suggests decreasing inequality after 2006, however the magnitude of the decreasing change is difficult to determine depending on the different inequality levels estimated for 2011 and for 2013. Since the direction of the trend changed for Germany, Belgium and Austria when we used non-imputed data (thus a much smaller sample) the results for these countries should be interpreted cautiously.

Interestingly, these results are only partly consistent with several of the previous studies finding an increasing trend in IRHI, as was described in section 1.2.2 on previous literature (for example Burström et al. (2005); Islam et al. (2010); Hernández-Quevedo et al (2006); van Ourti et al (2009)). However, although a few previous studies have investigated inequality using self-assessed health and C, we were unable to find any studies including these components during a period as recent as our period of study. The possibility to compare our results with these is therefore somewhat limited. One of the most recent studies is the study by Hernández-Quevedo and Masseria (2013), in which inequality was investigated by a binary variable based on self-assessed health and a health limitations variable over time 1994-2001 and 2005-2007, respectively. The trends calculated for the SAH variable are only partly similar to the trends found in this study. For example, the authors find an increasing trend in Denmark and France, while stable trends in Germany, the Netherlands and Austria. Similar trends as in this study are found for Spain and Italy. However, one should be aware of that imposing cardinality on the SAH variable by dichotomising the variable, as is done by Hernández-Quevedo and Masseria (2013), implies a great loss of information on the health of each individual in the sample. In fact, Wagstaff and van Doorslaer (1994) have shown this method to make comparisons of inequality over time to be unreliable. Thus, comparing our analysis with these results should be done cautiously. Studying health limitations for a more recent time period, they found a similar decreasing trend in inequality for Sweden, but opposite trends in Austria and Belgium. They also found some unclear directions of trends, although a tendency for increasing inequality in most countries. Given the enhanced estimation strategy with regards to the health variable, compared to Hernández-Quevedo and Masseria (2013), and given that our study period is even more recent, our results may provide indications of changing trends in IRHI in some countries since 2006. What can be noted is that, although not part of the main rationale for this thesis, this analysis provides an insight of how IRHI has developed over the financial crisis, which hit Europe in 2008. This might in fact be part of the explanation to why somewhat rather different IRHI levels were found in 2011 (wave 4) compared to 2013 (wave 5), which prevented us from giving hard answers on the level on the inequality: the year of 2011 may to a large extent pick up effects of the crisis that temporarily have affected income distributions across Europe. Except for the longer time period between wave 2 and wave 4, compared to the period between wave 4 and wave 5, the effect of this crisis may therefore be the reason to why one can see substantially larger changes in IRHI between 2006 and 2011, compared to 2011 to 2013, in for example Austria, Germany, the Netherlands and Switzerland. Furthermore, it is possible that the indications of decreasing trends in some countries, not as commonly depicted in previous research, could also partly be explained by the financial crisis. On this note, a very recent and (and preliminary) paper, by Coveney, Garcia-Gomez, van Doorslaer and van Ourti, seems interesting enough to suggest a decreasing trend in IRHI in almost all countries, when investigating the effect of the financial crisis on IRHI between 2008-2011 (International Health Economics Association, 2015). Similar to this study, a decreasing trend was found in, inter alia, Belgium and the authors argues that the decline in IRHI in this country is driven by a large income loss of high income earners. While the graphs of ours on Belgium, presented in the end of appendix D, cannot directly depict changes in income, we noted previously that the decline in IRHI in our results could seem to be driven by an increase of ill-health of the richer, rather than improvements in the health of the poor. It is therefore interesting to note these related findings that might suggest deterioration of both health and income of the richer individuals, which possibly could be linked to the crisis. Given the indications of possibly changing trends in some European countries, the recent impact of substantial financial distress in Europe in combination with the difficulties we experienced in finding recent trend studies, this provides motivation for further studies on this topic.

Regarding how the change in the above discussed distribution of income and health in Belgium may translate into different values (and partly trends) of IRHI when altering v, as previously was discussed in the results, it is interesting to note that this particular feature of the extended index, exemplified by the change in Belgium, recently has been criticized by Erreygers et al. (2012). What could be seen from graph D2, is that since the decreasing trend in relative inequality could be linked to higher ill-health in the relatively richer part of the population, rather than health improvements for the poor, it was possible to detect a stable or almost increasing trend in IRHI, when raising v enough, since this weighting scheme only took into account the slight deterioration of the health of the poor. This is due to the asymmetric feature of this measure which implies that it is only the health of the poor that is weighted higher with an higher v, even in a situation where health would be distributed pro-poor. That is, by using these measures, we do only care for relatively poorer health among the poor and not, as much, for relatively poorer health among the rich. Thus, these modifications (of raising v) is sensitive to the health of the poor, and not sensitive to inequality per se. As was discussed previously, this feature is likely the reason to the slightly changed trend in Belgium, when raising v. Due to this feature, Erreygers et al (2012) asks, normatively, whether we should care about poor health among the richer and as to provide the possibility of concluding on a positive answer on this question, they developed a symmetric extended Concentration Index. The point raised by Erreygers et al. (2012) is interesting and this measure could be useful for further studies. By for example employing both the asymmetric and the symmetric measure, the drivers behind the estimated inequality levels could be further understood. Nonetheless, again, the choice between the indices do come down to a question of value judgments: should poor health among rich be weighted as much as poor health among poor?

With regards to the second research question, on the sensitivity of these results with respect to the measures chosen, we do not find strong indications that these results are driven by different indices, for the chosen health variable and period of time studied in this essay. That is, the reason why we find somewhat opposite trends to previous studies does not seem to be a matter of inequality measure chosen, when comparing trend directions. Looking at the results, we can see that for Austria, Sweden, the Netherlands, Germany, and Switzerland we find consistent trends over all indices and inequality aversion parameters. This indicates that the trend may be the same (increasing or decreasing) regardless of the value judgements applied, which of course is not implausible, since absolute and relative inequality may rise (or decrease) simultaneously. However, we did find at least two examples (in Denmark and in Sweden) where the change from one year to another, did differ depending on the choice of index. While it did not affect the conclusions on the overall trend for the indices, these results point towards the need for caution when measuring IRHI over time using only one measure, unless the value judgement made implicitly in the chosen measure is motivated and thus the results. These findings are similar to the conclusions by Hernández-Quevedo and Masseria (2013), which provides the only article we have been able to find that previously have investigate the impact of different measures on IRHI trends. By acknowledging the challenges of the bounds issue and the mirror property of the standard C, they compare the trends between E and C and find that in some cases the trends differ between indices (for example Italy and the UK), however, overall the indices yield similar results, similar to this study.[45] However, the simple fact that even just a few opposite trends or changes are found by us and by Hernández-Quevedo and Masseria (2013), indicate that the choice of indices does matter. Therefore, and given that the choice between these measures are a matter of value judgements, rather than a matter of technical features, one may question how adequate it is to motivate the choice of using E, instead of W, by the fact that the latter may provide "larger inequalities than the Erreygers index when prevalence is low…"(p. 68), as is done by for example Costa-Font et al. (2014). Given that these possibly larger inequalities are a consequence of a certain value judgement (of a combination of s-relative and h-relative inequality, more specifically, as is the case for W), there might be reason for a more transparent approach by researchers in studies of IRHI.

With regards to the second part of the sensitivity analysis, focusing on altering the inequality aversion parameter (v) rather than on the difference between the indices, we also found a limited impact on the direction of trends (but not on level, as expected) when altering the v. However, indications of changing trends were found in for example Belgium and furthermore, in Sweden we noted that altering v also could imply different directions of change depending on the measure used. In addition, it was interesting to note that the extended version of E (as defined by Erreygers et al. (2012) but based on the extended C defined by Wagstaff (2002)), notably decreased for most (but not all) countries when raising v. Except for the application by Erreygers et al. (2012) when defining this index, we have not found any other article applying this measure before, thus, the decrease they found when raising v, using this index on developing countries, is confirmed for a different health variable and in a different economic-geographical context by this study. Again, the sensitivity of the results in trends with respect to v could however not be judged as to be very sensitive, and again, the characteristics of the health variable could be part of the explanation of the results.

On the one hand, our essay summarizes very recent theoretical developments in the measurement of IRHI, in which the different implicit value judgments within these measures have gained increasing attention, thus pointing to the need to employ different indices from a

[45] Hernández-Quevedo and Masseria (2013) state that for a few countries their trend changes "substantially" (p. 469) and point to Italy as an example. However since C indicates a decrease from 0.14 to 0.11 while E instead indicates a slight increase from 0.110 to 0.111, we are not convinced that this change is to be referred to as substantial. Overall rather limited changes seem to be found in their study, as is found in ours.

normative transparent standpoint. On the other hand, we have empirically assessed these differences for a certain sample but found limited indications of large differences between these indices. What the above results and discussion seems to underline is that the sensitivity of the trends in IRHI not merely is an isolated question regarding measurement, but is closely related to the characteristics of the health variable, in terms of mean, variance and the distribution in the sample, such as the aspect of monotonicity as earlier discussed.

Given the close link between the health variable and the inequality measures, a few more comments on the health variable used in this study can be made. First of all, previous alternative methods on how to impose cardinality on an ordinal variable have already been discussed in the method (see section 3.2.2.1.1). Although acknowledging the statement by van Doorslaer and Jones (2003) that imposing cardinality by the ordered probit model may need further validation, we note that the model seems to yield plausible results in terms of signs and relative magnitude of the impacts of the health conditions. The mean value of the health index per SAH category and the mean of the health index per country also seem to be rather close to values estimated by other methods (see for example TTO estimates in Burström et al. (2014) and values from the interval regression methods in van Doorslaer and Jones (2003)). Secondly, we noted briefly in the results section that there seems to be a limited variability in the distribution of health over income in our sample, i.e. that there are relatively low differences in mean health value between the poorest and the richest deciles (for example by looking at the graph for the Czech Republic). However, given that the probit model seems to yield plausible results, this finding rather seems to suggest that most of the elderly in Europe in all income deciles report relatively good health. Extending the analysis by employing other but similar measurements would have been interesting. Nonetheless, since the method of imposing cardinality is a transformation implying further adjustments of the data and that the distribution of the health variable have been shown to be important in order to explain the pattern of inequality that arises, the validity of the study is of course contingent on this method. Third, given the finding that many individuals tend to report rather good health one might face both the technical and the normative question whether self-assessed health is a good variable to use when investigating overall health. For example, in Leu and Schellhorn (2006), they note that the proportion of individuals who report good or very good health remained stable around 85 percent between 1982 to 2002, a period of 20 years. Hernández-Quevedo and Masseria (2013) argue further that limitations in daily activities may be considered to be a more objective measure, which thus would capture the level of health "more accurately" (p. 464) while van Doorslaer and Jones (2003), on the matter of reporting bias in the

SAH variable, points to the normative issue in trusting the reports of individuals' own health. As the basis to suggest improvements for future research, some additional technical limitations of this study will be discussed in the next section, before ending with conclusions.

5.1 Limitations of the study

The study at hand and its results should be viewed in the light of its limitations, linked to the chosen method, available data and limited scope of time for this paper. Thus, some important limitations may be worth to be highlighted as they may provide considerations for future and enhanced studies on related topics.

First, when comparing our study with the very few previous, similar studies, it is hard to omit from one of the more important limitations of this study, namely the limited amount of data points over time used. The inclusion of the fourth wave in the SHARE survey, which also would have expanded the time period by two years, would have strengthen the empirical application in order to provide a more general trend. The change of income survey question in SHARE between years, which preventing us from using the first wave, is thus truly unfortunate. Nevertheless, the period as such, given the financial crisis, do constitute an interesting period to study changes in IRHI, for which future studies may be able to confirm or reject as temporary changes.

Secondly, our dependent variable may suffer from systematic reporting bias. If the same level of (true) health is assessed differently across sub-groups of individuals, heterogenous reporting behaviour may cause bias in the measured IRHI (Lindeboom and Doorslaer, 2004). In general, the assumption of homogenous reporting behaviour may be restrictive.[46] In this study we have controlled for systematic reporting differences between gender and age groups by the standardisation procedure (see section 3.3.1) as suggested by Lindeboom and Doorslaer (2004). However this captures only one of two types of reporting heterogeneity, where the first one can be referred to as index shift and the second as cut point shift. The former refers to the situation where the reporting behaviour affects all response categories similarly (for example if elderly tend to assess all health levels more positively than younger individuals would). This kind of

[46] Firstly, differences in national health care may have an impact on how the same level of health is rated by different individuals (Jürges, 2007). Secondly, there might be different subgroups within populations that use different reference points when assessing their health, linked to for example age and gender (Lindeboom and Doorslaer, 2004). Thirdly, differences in how response categories have been translated in different languages, may affect how some individuals in some countries choose to rate their health (see Jürges 2007).

reporting bias is controlled for by the standardisation procedure. The latter reporting bias implies that the thresholds are affected differently by the response behaviour (for example if Swedish respondents are less likely to pick one specific SAH category, due to the translation of this specific response category)(Jürges, 2007). Recalling our model as depicted in eq. (13) in section 3.2.2.1 one sees that this expression captures several different equations (in total 5 equations in our case) used for calculating the cumulative probabilities, where each equation models the probability that the respondent chooses that specific SAH category (in total 5 different categories). What this equation also shows is that the slope coefficients captured in the vector β is assumed to be the same in all equations. This illustrates the *parallel regression assumption* made in the standard ordered probit model, which also implies that no cut point shifts are modelled. To evaluate whether the SHARE survey indicates cut point shifts and to take this into account, an alternative estimation strategy could have been the generalised ordered probit model which relaxes the parallel regression assumption. However since we employed country specific models, country specific cut points were allowed indirectly and since Pfarr et al. (2012) did not find evidence for cut off point shifts for gender, also using SHARE data, we do however have reason to believe that this should not constitute large source of bias in our results. Given these aspects, it is worth underlining for future research that the methods for imposing cardinality on ordinal variables by probit or interval regression methods are closely related to the issue of heterogeneous reporting behaviour. It seems that the literature would benefit from additional, extensive work to investigate the sources of misspecification and potential bias in the (self-assessed) health variable.

Thirdly, as we have concluded in the discussion, IRHI is not only dependent on the normative aspects of measuring inequality, but also on the characteristics of the health variable. However, what we have not yet highlighted is that the measured degree of inequality may also be dependent on the choice of socioeconomic variable. Current income is not necessarily the best variable to capture socioeconomic status. For example Islam et al. (2010) examine IRHI over time using both current income and lifetime income and find that when individuals are ranked according to current income, IRHI is decreasing, while a stable trend is detected when lifetime income is used. Another variable that can be used as a proxy for socioeconomic status is education (used by for example Shkolnikov et al. 2011). Therefore, it would also be interesting to examine these trends in IRHI taking into account not only measurement issues and the properties of the health variable, but also the choice of socioeconomic variable.

5.2 Conclusions

Investigating IRHI in the context of relatively rich European market economies, we find both rising and declining IRHI trends in different countries. When investigating the sensitivity of these results with respect to different IRHI index, which rather recently have been proposed in the literature, the estimated trends remain rather consistent between measures.

Despite the somewhat limited conclusions one can draw from the results of this empirical study due to the limited data available, we believe that this essay has made an important contribution to the current research in terms of research design by discussing and applying several inequality measures, while acknowledging the different value judgements they imply, as have been highlighted only very recently in the literature. Although it seems that the indices often yield similar results, a change in IRHI over time may be depicted differently by different measures. In addition, although we believe that the different trends found when altering the inequality parameter *v* may be of less importance (since the different trends were found for very high values of *v,* which we believe have less legitimacy) this point to the same findings: a researcher that claims to be describing IRHI as objectively and transparent as possibly should first of all underline any value judgements implicitly or explicitly made in the indices, and should second of all possibly consider to take into account several different indices when investigating IRHI. The importance of this is likely to be increasing in the length of period covered by the analysis and in the degree of variance in the mean of health in the countries or regions in focus for the study. While this thesis have not taken a normative stance on the most appropriate index to use when dealing with bounded variables, we would like to conclude by underlining the important normative value of research meeting principles of transparency and as far as possible, principles of objectivity. Lastly, although the answer to the first research question provided by this essay should be interpreted somewhat cautiously, given the limited time period and other estimations strategies available as previously discussed, the decreasing trends depicted, in contrast to much previous research, is an interesting finding. Future studies should evaluate whether these trends are specific for this limited period, possibly influenced by economic retrenchment, or indications of somewhat changing patterns in IRHI.

By combining some of the possible technical consideration discussed above together with more extensive data there seem to be ample possibility for further empirical studies in this field. Given the limited research on how IRHI has developed over time, the different results the research have yielded and the increasing focus on matters of inequality amongst policy makers, further studies seem to be well motivated. Whether IRHI has increased or decreased over time is a rather simple question to ask, however, as have been discussed in this essay, it is a rather difficult question to answer.

6 References

Alcser, K. H., Benson, G., Börsch-Supan, A., Brugiavini, A., Christelis, D., Croda, E., ... & Weerman, B. (2005). The Survey of Health, Aging, and Retirement in Europe–Methodology. Mannheim Mannheim Research Institute for the Economics of Aging (MEA).

Allanson, P., & Petrie, D. (2013). On The Choice Of Health Inequality Measure For The Longitudinal Analysis Of Income‐Related Health Inequalities. Health economics, 22(3), 353-365.

Asada, Y. 2010. On the Choice of Absolute or Relative Inequality Measures. The Milbank Quarterly 88(4):616–22.

Burström, K., Johannesson, M., & Diderichsen, F. (2005). Increasing socio‐economic inequalities in life expectancy and QALYs in Sweden 1980–1997. Health economics, 14(8), 831-850.

Burström, K., Sun, S., Gerdtham, U. G., Henriksson, M., Johannesson, M., Levin, L. Å., & Zethraeus, N. (2014). Swedish experience-based value sets for EQ-5D health states. Quality of Life Research, 23(2), 431-442.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S. and Zuber, S. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology DOI: 10.1093/ije/dyt088.

Christelis, D. (2011). Imputation of Missing Data in Waves 1 and 2 of SHARE. Available at SSRN 1788248.

Clarke, P. M., Gerdtham, U. G., Johannesson, M., Bingefors, K., & Smith, L. (2002). On the measurement of relative and absolute income-related health inequality. Social Science & Medicine, 55(11), 1923-1928.

Costa-Font, J., Hernández-Quevedo, C., & Jiménez-Rubio, D. (2014). Income inequalities in unhealthy life styles in England and Spain. *Economics & Human Biology*, *13*, 66-75.

Cutler, D. M., Richardson, E., Keeler, T. E., & Staiger, D. (1997). Measuring the health of the US population. Brookings papers on economic activity. Microeconomics, 217-282.

Dalstra, J., Kunst, A. E., Geurts, J. J. M., Frenken, F. J. M., & Mackenbach, J. P. (2002). Trends in socioeconomic health inequalities in the Netherlands, 1981–1999. Journal of Epidemiology and Community Health, 56(12), 927-934.

Erreygers, G. (2009a). Correcting the concentration index. Journal of health economics, 28(2), 504-515.

Erreygers, G. (2009b). Correcting the concentration index: a reply to Wagstaff. Journal of Health Economics, 28(2), 521-524.

Erreygers, G., Clarke, P., & Van Ourti, T. (2012). "Mirror, mirror, on the wall, who in this land is fairest of all?"—Distributional sensitivity in the measurement of socioeconomic inequality of health. Journal of health economics, 31(1), 257-270.

Erreygers, G., & Van Ourti, T. (2011). Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. Journal of health economics, 30(4), 685-694.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., ... & Boyle, M. (2002). Multiattribute and single‐attribute utility functions for the health utilities index mark 3 system. Medical care, 40(2), 113-128.

Fryback, D. G., Dasbach, E. J., Klein, R., Klein, B. E., Dorn, N., Peterson, K., & Martin, P. A. (1993). The beaver dam health outcomes study initial catalog of health-state quality factors. Medical Decision Making, 13(2), 89-102.

Gravelle, H., & Sutton, M. (2003). Income related inequalities in self assessed health in Britain: 1979–1995. Journal of Epidemiology and Community Health,57(2), 125-129.

Greene, W. H., & Hensher, D. A. (2009). Modeling ordered choices. Available at http://pages.stern.nyu.edu/~wgreene/DiscreteChoice/Readings/OrderedChoiceSurvey.pdf (accessed 12.08.2015)

Groot, W. (2000). Adaptation and scale of reference bias in self-assessments of quality of life. Journal of health economics, 19(3), 403-420.

Harper, S., & Lynch, J. (2007). Trends in socioeconomic inequalities in adult health behaviors among US states, 1990–2004. Public health reports, 122(2), 177.

Hernández-Quevedo, C., Jones, A. M., López-Nicolás, A., & Rice, N. (2006). Socioeconomic inequalities in health: A comparative longitudinal analysis using the European Community Household Panel. Social Science & Medicine, 63(5), 1246-1261.

Hernández-Quevedo, C., & Masseria, C. (2013). Measuring income-related inequalities in health in multi-country analysis. Estudios de Economía Aplicada, 31(2), 455-476.

Hong, J., Knapp, M., & McGuire, A. (2011). Income‐related inequalities in the prevalence of depression and suicidal behaviour: a 10‐year trend following economic crisis. World Psychiatry, 10(1), 40-44.

International Health Economics Association, 2015. Abstract: The health gradient in Europe during the financial crisis. [Web page]. Retrieved from https://ihea2015.abstractsubmit.org/presentations/13762/.

Islam, M. K., Gerdtham, U. G., Clarke, P., & Burström, K. (2010). Does income‐related health inequality change as the population ages? Evidence from Swedish panel data. Health economics, 19(3), 334-349.

Jones, A. M., & Nicolás, A. L. (2004). Measurement and explanation of socioeconomic inequality in health with longitudinal data. Health economics, 13(10), 1015-1030.

Jürges, H. (2005). Computing a comparable health index. In Börsch-Supan, A., Brugiavini, A., Jürges, H., et al. (eds.) Health, ageing and retirement in Europe – First results from the Survey of Health, Ageing and Retirement in Europe, p. 357. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA).

Jürges, H. (2007). True health vs response styles: exploring cross‐country differences in self‐reported health. Health economics, 16(2), 163-178.

Jürges, H. (2010). Health inequalities by education, income and wealth: a comparison of 11 European countries and the US. Applied Economics Letters, 17(1), 87-91.

Kakwani, N., Wagstaff, A., & Van Doorslaer, E. (1997). Socioeconomic inequalities in health: Measurement, computation, and statistical inference. Journal of Econometrics, 77(1), 87-103.

King, N. B., Harper, S. A. M., Meersman, S. C., Reichman, M. E., Breen, N., & Lynch, J. (2010). We'll take the red pill: a reply to Asada. Milbank Quarterly, 88(4), 623-627

Kjellsson, G., & Gerdtham, U. G. (2013). On correcting the concentration index for binary variables. Journal of health economics, 32(3), 659-670.

Kjellsson G., and Gerdtham U.-G. (2014). Measuring Health Inequalities Using the Concentration Index Approach. In: Anthony J. Culyer (ed.), Encyclopedia of Health Economics, Vol 2. San Diego: Elsevier. pp. 240-246.

Kjellsson, G., Gerdtham, U. G., & Petrie, D. (2015). Lies, damned lies, and health inequality measurements: understanding the value judgments. *Epidemiology*, *26*(5), 673-680.

Kjellsson, G. (2014) Health, Inequality, and Smoking. Doctoral Dissertation, Lund Economic Studies Nr 175.

Kolm, S. (2007). Economic justice. Mimeo. Available from: http://sergekolm.org/wp-content/uploads/2014/02/ECONOMIC-JUSTICEFev07.pdf

Kunst, A. E., Bos, V., Andersen, O., Cardano, M., Costa, G., Harding, S., ... & Mackenbach, J. P. (2004). Monitoring of trends in socioeconomic inequalities in mortality: experiences from a European project. Demographic research, 2(9), 229-54.

Kunst, A. E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., ... & Mackenbach, J. P. (2005). Trends in socioeconomic inequalities in self-assessed health in 10 European countries. International Journal of Epidemiology, 34(2), 295-305.

Leu, R. E., & Schellhorn, M. (2006). The evolution of income-related health inequalities in Switzerland over time. CESifo Economic Studies, 52(4), 666-690.

Lindeboom, M., & Van Doorslaer, E. (2004). Cut-point shift and index shift in self-reported health. Journal of health economics, 23(6), 1083-1099.

Ljungvall, Å., & Gerdtham, U. G. (2010). More equal but heavier: A longitudinal analysis of income-related obesity inequalities in an adult Swedish cohort. Social Science & Medicine, 70(2), 221-231.

Nolte, E., & McKee, M. (2004). Changing health inequalities in east and west Germany since unification. Social Science & Medicine, 58(1), 119-136.

O'Donnell, O., van Doorslaer, E., Wagstaff, A., Lindelow, M. (2008). Analyzing health equity using household survey data: a guide to techniques and their implementation. Washington DC: The World Bank

Pfarr, C., Schmid, A., & Schneider, U. (2012). Reporting heterogeneity in self-assessed health among elderly Europeans. Health economics review, 2(1), 1-14.

Schneider, U., Pfarr, C., Schneider, B. S., & Ulrich, V. (2012). I feel good! Gender differences and reporting heterogeneity in self-assessed health. The European Journal of Health Economics, 13(3), 251-265.

Sen, A. (2002). Why health equity?. Health economics, 11(8), 659-666.

Shkolnikov, V. M., Andreev, E. M., Jdanov, D. A., Jasilionis, D., Kravdal, Ø., Vågerö, D., & Valkonen, T. (2011). Increasing absolute mortality disparities by education in Finland, Norway and Sweden, 1971–2000. Journal of epidemiology and community health, jech-2009.

SHARE (2013). Release Guide 2.6.0 Wave 1 & Wave 2. Mannheim Research Institute for the Economics of Aging.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for?. Journal of Human Resources, 50(2), 301-316.

Van Doorslaer, E., & Jones, A. M. (2003). Inequalities in self-reported health: validation of a new approach to measurement. Journal of health economics, 22(1), 61-87

Van Doorslaer, E., & Van Ourti, T. (2011). Measuring inequality and inequity in health and health care. Oxford Handbook on Health Economics. Oxford: Oxford University, 837-869.

Van Ourti, T., van Doorslaer, E., & Koolman, X. (2009). The effect of income growth and inequality on health inequality: Theory and empirical evidence from the European Panel. Journal of Health Economics, 28(3), 525-539.

Verbeek, M. (2012). A guide to modern econometrics. John Wiley & Sons.

Wagstaff, A. (2002). Inequality aversion, health inequalities and health achievement. Journal of health economics, 21(4), 627-641.

Wagstaff, A. (2005). The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. Health economics, 14(4), 429-432.

Wagstaff, A. (2009). Correcting the concentration index: a comment. Journal of Health Economics, 28(2), 516-520.

Wagstaff, A., Paci, P., & Van Doorslaer, E. (1991). On the measurement of inequalities in health. Social science & medicine, 33(5), 545-557.

Wagstaff, A., & Van Doorslaer, E. (1994). Measuring inequalities in health in the presence of multiple-category morbidity indicators. Health economics.

Wagstaff, A., & Van Doorslaer, E. (2000). Equity in health care finance and delivery. Handbook of health economics, 1, 1803-1862.

Wildman, J. (2003). Income related inequalities in mental health in Great Britain: analysing the causes of health inequality over time. Journal of health economics, 22(2), 295-312.

Zhang, Q., & Wang, Y. (2007). Using concentration index to study changes in socio-economic inequality of overweight among US adolescents between 1971 and 2002. International journal of epidemiology, 36(4), 916-925.

APPENDIX A

THEORY

TABLE A1. Previous literature on development of IRHI over time.

(1)	(2)	(3)	(4)	(5)	(6)
Author(s)	**Time period**	**Country (-ies)**	**Indicators**	**Inequality trend**	**Health variable**
Burström et al. (2005)	1980-1997	Sweden	C, other	+	QALYs, life expectancy
Dalstra et al. (2002)	1981-1999	Netherlands	Relative index of inequality, other	+	SAH
Gravelle and Sutton (2003)	1979-1995	UK	C	+	SAH
Harper and Lynch (2007)	1990-2004	US (50 states)	Relative concentration index	+ (smoking, phys inactivity) -(alcohol, obesity)	Smoking, alcohol use, other
Hernández-Quevedo et al. (2006)	1994-2001	European countries	C, Extended C, Wagstaff's achievement index	+	Health limitations
Hernández-Quevedo & Masseria (2013)	1994-2001 and 2005-2007	14-20 European countries	C, E	+ (DK, FRA, ES other) Stable (DE, NE, BE, ITA, AT, other)	SAH, health limitation.
Hong et al. (2011)	1998-2007	South Korea	C	+	Depression, suicide attempts, other
Islam et al. (2010)	1980-1997	Sweden	C	+/stable	Health state score (EQ-5D)
Jones & López Nicolás (2004)	1991-1999	UK	C	-	GHQ measure of psychological well-being
Kunst et al. (2004)	1980s-1990s	8 European countries	Relative index of inequality	+	Mortality
Kunst et al. (2005)	1980s-1990s	Western Europe	Prevalence rates , other	stable/ +	SAH
Leu & Schellhorn (2006)	1982 - 2002	Switzerland	C	+ (until 1997, then stable)	SAH
Ljungvall and Gerdtham (2010)	1980-1997	Sweden	W	-	Binary obesity variable
Nolte &McKee (2004)	1992-1997	Germany	Odds ratios	+	SAH
Shkolnikov et al. (2011)	1971-2000	FI, NO, SE	Rate differences, other	+	Mortality
Van Ourti et al. (2009)	1994-2001	European countries	C	+ (except DE, NE)	SAH
Wildman (2003)	1992-1998	UK	C	stable/ -	Mental health
Zhang & Wang (2007)	1971-2002	US	C	-	Obesity

APPENDIX B

METHOD: DESCRIPTIVE STATISTICS

Table B1. Data description.

VARIABLE	Description
income	total income received by all household members in last month
SAH	self-perceived health - us version
heart attack	doctor told you had: heart attack
high blood pressure	doctor told you had: high blood pressure or hypertension
high blood cholesterol	doctor told you had: high blood cholesterol
stroke	doctor told you had: stroke
diabetes	doctor told you had: diabetes or high blood sugar
chronic lung disease	doctor told you had: chronic lung disease
arthritis	arthritis
cancer	doctor told you had: cancer
stomach/ duodenal ulcer	doctor told you had: stomach or duodenal ulcer, peptic ulcer
Parkinson	doctor told you had: Parkinson’s disease
cataracts	doctor told you had: cataracts
hip fracture	doctor told you had: hip fracture or femoral fracture
other conditions	doctor told you had: other conditions
Alzheimer/ dementia	doctor told you had: Alzheimer's disease, dementia, senility
depression	depression (part of EURO-D)
low grip	
no grip	
imputed income	

Table B2. Descriptive statistics

VARIABLE	Mean	Std. Dev.	Min	Max	Obs
SAH	3,047	1,073	1,000	5,000	N =119244
heart attack	0,110	0,313	0,000	1,000	N =119244
high blood pressure	0,364	0,481	0,000	1,000	N =119244
high blood cholesterol	0,223	0,416	0,000	1,000	N =119244
stroke	0,036	0,185	0,000	1,000	N =119244
diabetes	0,117	0,321	0,000	1,000	N =119244
chronic lung disease	0,060	0,238	0,000	1,000	N =119244
arthritis	0,162	0,368	0,000	1,000	N =119244
cancer	0,052	0,221	0,000	1,000	N =119244
stomach/ duodenal ulcer	0,039	0,194	0,000	1,000	N =192444
parkinson	0,007	0,083	0,000	1,000	N =119244
cataracts	0,082	0,274	0,000	1,000	N =119244
hip fracture	0,020	0,140	0,000	1,000	N =119244
other conditions	0,158	0,365	0,000	1,000	N =119244
alzheimer/ dementia	0,011	0,105	0,000	1,000	N =119244
depression	0,384	0,486	0,000	1,000	N =119244
low grip	0,339	0,473	0,000	1,000	N =119244
no grip	0,064	0,245	0,000	1,000	N =119244

Table B3. Descriptive statistics for imputed income

Variable	Mean	Std. Dev.	Min	Max	Obs
imputed income 1	56776,070	110502,300	0,000	10100000,000	N =119244
imputed income 2	56922,300	112062,400	0,000	10000000,000	N =119244
imputed income 3	56664,100	109539,500	0,000	10000000,000	N =119244
imputed income 4	56616,920	109461,300	0,000	10000000,000	N =119244
imputed income 5	56675,420	109764,900	0,000	10000000,000	N =119244

Table B4. Descriptive statistics of imputed income , country specific[47]

	Obs	Mean	Std.Dev	Min	Max
Austria	10244	31273.63	25377.13	0	544533.6
Germany	9706	36141.63	31583.84	0	1035135
Sweden	9067	138104.6	195295.2	0	4511470
Netherlands	9431	42889.77	38864.5	0	1021589
Spain	11696	20194.51	21337.46	0	351950.6
Italy	11033	25855.06	34760.59	0	1056074
France	12836	38531.97	134544.8	0	1.01e+07
Denmark	8903	128668.6	177710.6	0	3000000
Switzerland	8160	96773.23	93785.28	0	2293830
Belgium	13864	50563.55	63887.09	0	1264591
Czech Rep	14304	55235.62	133933.6	0	3736456

[47] The mean, min and max values are very similar between the five different imputed income dataset provided in the SHARE survey, employed in this analysis. Thus, we only provide descriptive statistics for one of these imputed variables.

Table B5. Self-assessed health (SAH) for each country (1=excellent health, 5=poor health)

Countries	Observations	Mean SAH	Std. Dev.	Min	Max
Austria	10244	2.958	1.050	1	5
Germany	9706	3.226	1.004	1	5
Sweden	9067	2.73	1.156	1	5
Netherlands	9431	2.934	1.034	1	5
Spain	11696	3.351	.999	1	5
Italy	11033	3.250	1.075	1	5
France	12836	3.190	1.034	1	5
Denmark	8903	2.519	1.153	1	5
Switzerland	8160	2.650	.9726	1	5
Belgium	13864	2.98	.996	1	5
Czech Rep	14304	3.346	1.003	1	5

APPENDIX C

METHOD: ESTIMATIONS

C1. Imputation procedure

Following Christelis (2011) we calculate the imputed estimate as the average of the estimates (of the respective inequality indices) from each imputed data set, according to the following equation:

$$\bar{\hat{\beta}} = \frac{1}{M}\sum_{m=1}^{M}\hat{\beta}_m \tag{1}$$

Since this imputed based inequality index is an estimate of its own, one needs to adjust the standard errors accordingly. This procedure is slightly more complicated and based on a three-step equation where the variance within (WV) and between (BV) the imputations are taken into account as well as the total variance (TV), according to the following equations:

$$WV = \frac{1}{M}\sum_{m=1}^{M} V_m \tag{2}$$

$$BV = \frac{1}{M}\sum_{m=1}^{M}(\hat{\beta}_m - \bar{\hat{\beta}}) \tag{3}$$

$$TV = WV + \frac{M+1}{M}BV \tag{4}$$

By calculating the degrees of freedom (DF) according to the below equation, one can then assess the significance of the inequality index by a standard two-tailed t-test.

$$df = (M-1)\left(1 + \frac{1}{M+1}\frac{WV}{BV}\right)^2 \tag{5}$$

C2. Estimating the C, W and E by the convenient regression model

For the purpose of estimation it is convenient to note that C can be expressed in terms of the covariance between the health variable *H* and the income rank variable *R*:

$$C = \frac{2}{\mu}\ cov(H,R) \tag{6}$$

Recalling that the slope coefficient β from an *ordinary least squares* (OLS) regression of *h* on *r* is equal to:

$$\beta = \frac{cov(H,R)}{var(R)} \tag{7}$$

Given the above, one can then rewrite C as:

$$C = \frac{2}{\mu_H}\ var(R) \cdot \beta \tag{8}$$

,where μ_h equal mean of the health variable. By multiplying the left and the right hand side of an regression of health on socioeconomic rank by *A,* the inequality index is obtained as the regression coefficient β_C:

$$C = A \cdot \beta \qquad \text{where } A = \frac{2}{\mu_H} var(R) \tag{9}$$

$$H_c = \alpha + \beta_c \cdot R_i + \varepsilon_i \qquad \text{where } H_c = \frac{2}{\mu_H} var(R) \cdot H \text{ and } \beta_c = C \tag{10}$$

Given the relationship between covariance and OLS one can then retrieve C directly from the slope coefficient as in eq. (10) (Wagstaff et al., 1991). This method is typically referred to as the convenient regression method (O'Donnel et al., 2008). It should here be noted that although employing the estimation method for C we are de facto estimating the modified C, which was briefly introduced and described in the end of section 2.2.3.2, since the cardinal variables we are using (the predicted health index) is normalised and restricted to be between 0 and 1.[48] As stated in section 2.2.3.2, estimating C with a normalised health variable corresponds to estimating the modified C. Furthermore, by recalling from section 2.2.3 how C, E and W can be written in similar ways with only one different term (Erreygers 2009a), the convenient regression method can be employed in order to calculate all these inequality indices by adjusting the transformation of the health variable accordingly. To calculate W we note that W can be written as:

$$W = \frac{2}{(1-\mu_H) \cdot \mu_H \cdot cov(H,R)} \tag{11}$$

and following the above steps, the below equation can be employed to estimate W by the convenient regression method:

$$H_w = \alpha + \beta_w \cdot R_i + \varepsilon_i \quad \text{where } H_w = \frac{2}{(1-\mu_h) \cdot \mu_h \cdot var(R)} \cdot H \text{ and } \beta_w = W \tag{12}$$

For E the corresponding equation is the following:

$$E = 8 \cdot cov(H, R) \tag{13}$$

and the regression can be estimated by this specification:

$$H_E = \alpha + \beta_E \cdot R_i + \varepsilon_i \qquad \text{where } H_E = 8 \cdot var(R) \cdot H \text{ and } \beta_E = E \tag{14}$$

Based on these derivations and as extensions of above equations (10), (12) and (14), the extended forms, our final models, of the respective indices are estimated. These regressions are explained in section 3.3.2.

[48] Recall that the constructed health index we use in this analysis is not a ratio scale variable; rather the probit model impose cardinality on the ordinal SAH. However, as described in section 2.1.3.2, C is not suitable to use for a cardinal variable since it does not satisfy scale invariance, hence the need to use the modified C.

APPENDIX D

RESULTS

Table D1. Ordered probit model (2006). Predictions of these equations generate the health index.

Dependent variable: SAH	**(1) Austria**	**(2) Germany**	**(3) Sweden**	**(4) Netherlands**	**(5) Spain**	**(6) Italy**	**(7) France**	**(8) Denmark**	**(9) Switzerland**	**(10) Belgium**	**(11) Czech Slovakia**
heart attack	0.664***	0.732***	0.529***	0.802***	0.691***	0.701***	0.786***	0.545***	0.774***	0.766***	0.742***
	(0.104)	(0.0725)	(0.0606)	(0.0771)	(0.0879)	(0.0701)	(0.0653)	(0.0730)	(0.124)	(0.0621)	(0.0618)
high blood pressure	0.259***	0.397***	0.363***	0.345***	0.254***	0.228***	0.294***	0.208***	0.366***	0.230***	0.406***
	(0.0682)	(0.0475)	(0.0462)	(0.0503)	(0.0522)	(0.0431)	(0.0468)	(0.0488)	(0.0659)	(0.0421)	(0.0444)
high blood cholesterol	0.128	0.0953	0.0331	0.135*	0.188**	0.0740	0.137**	0.108	0.159	0.0660	0.0883
	(0.0813)	(0.0615)	(0.0588)	(0.0621)	(0.0582)	(0.0501)	(0.0500)	(0.0583)	(0.0870)	(0.0433)	(0.0565)
stroke	0.529**	0.808***	0.545***	0.601***	0.918***	0.881***	0.585***	0.989***	0.447*	0.466***	0.854***
	(0.187)	(0.132)	(0.112)	(0.128)	(0.182)	(0.139)	(0.133)	(0.104)	(0.198)	(0.120)	(0.112)
diabetes	0.618***	0.720***	0.671***	0.646***	0.472***	0.434***	0.676***	0.410***	0.612***	0.427***	0.468***
	(0.107)	(0.0710)	(0.0764)	(0.0777)	(0.0705)	(0.0660)	(0.0728)	(0.0849)	(0.120)	(0.0698)	(0.0628)
chronic lung disease	0.756***	0.692***	1.198***	0.808***	0.870***	0.685***	0.921***	0.884***	0.742***	0.777***	0.502***
	(0.158)	(0.0978)	(0.123)	(0.0943)	(0.105)	(0.0830)	(0.102)	(0.0910)	(0.154)	(0.0900)	(0.102)
arthritis	0.494***	0.625***	0.495***	0.503***	0.567***	0.547***	0.475***	0.605***	0.620***	0.568***	0.457***
	(0.0923)	(0.0673)	(0.0689)	(0.0722)	(0.0563)	(0.0461)	(0.0489)	(0.0490)	(0.0923)	(0.0489)	(0.0582)
cancer	1.194***	0.911***	0.308***	0.674***	1.040***	0.862***	0.712***	0.573***	0.451**	0.977***	0.689***
	(0.209)	(0.107)	(0.0847)	(0.110)	(0.175)	(0.118)	(0.101)	(0.0879)	(0.144)	(0.106)	(0.111)
stomach/ duodenal ulcer	0.597***	0.539***	0.146	0.478**	0.582***	0.173	0.575***	0.419***	0.550*	0.336***	0.351***
	(0.175)	(0.141)	(0.119)	(0.153)	(0.134)	(0.0896)	(0.123)	(0.101)	(0.223)	(0.0913)	(0.0845)
Parkinson	0.809*	2.091***	1.127**	1.018***	0.284	1.037***	1.637***	1.621***	1.984***	1.311***	0.779**
	(0.363)	(0.337)	(0.402)	(0.285)	(0.345)	(0.248)	(0.289)	(0.270)	(0.510)	(0.247)	(0.277)

Table D1. *cont.*

cataracts	0.163 (0.122)	0.164 (0.0965)	0.170* (0.0692)	0.0614 (0.0960)	0.437*** (0.101)	0.172 (0.0896)	0.247** (0.0933)	0.0727 (0.0739)	0.254* (0.124)	0.0666 (0.0873)	0.119 (0.0854)
hip fracture	0.115 (0.194)	0.565** (0.219)	0.466*** (0.127)	0.151 (0.245)	0.360* (0.184)	0.409* (0.174)	0.548** (0.174)	0.191 (0.157)	0.0283 (0.300)	0.361* (0.145)	0.363* (0.142)
other conditions	0.544*** (0.0923)	0.819*** (0.0606)	0.556*** (0.0502)	0.596*** (0.0591)	0.492*** (0.0609)	0.466*** (0.0639)	0.584*** (0.0678)	0.626*** (0.0590)	0.834*** (0.0804)	0.544*** (0.0738)	0.474*** (0.0548)
Alzheimer/ dementia	0.846** (0.277)	1.193*** (0.248)	0.980*** (0.246)	0.247 (0.237)	1.185*** (0.217)	0.556** (0.214)	0.959*** (0.196)	0.288 (0.339)	0.805 (0.437)	1.033*** (0.233)	1.044*** (0.302)
depression	0.412*** (0.0691)	0.498*** (0.0466)	0.374*** (0.0458)	0.359*** (0.0469)	0.463*** (0.0521)	0.525*** (0.0431)	0.269*** (0.0419)	0.315*** (0.0463)	0.317*** (0.0598)	0.332*** (0.0412)	0.353*** (0.0456)
low grip	0.254*** (0.0725)	0.222*** (0.0521)	0.230*** (0.0480)	0.227*** (0.0490)	0.412*** (0.0529)	0.300*** (0.0457)	0.277*** (0.0454)	0.175*** (0.0456)	0.358*** (0.0640)	0.226*** (0.0432)	0.308*** (0.0483)
no grip	0.828*** (0.103)	0.520*** (0.0896)	0.706*** (0.106)	0.822*** (0.118)	0.633*** (0.0870)	0.726*** (0.0652)	0.607*** (0.0738)	0.993*** (0.143)	0.559*** (0.164)	0.733*** (0.105)	0.882*** (0.113)
Threshold 1	-0.891*** (0.0669)	-1.023*** (0.0486)	-0.351*** (0.0381)	-0.666*** (0.0393)	-1.231*** (0.0637)	-0.788*** (0.0438)	-0.870*** (0.0455)	-0.245*** (0.0387)	-0.494*** (0.0515)	-0.855*** (0.0400)	-1.355*** (0.0535)
Threshold 2	0.103 (0.0572)	-0.0651 (0.0389)	0.470*** (0.0377)	-0.0294 (0.0358)	-0.292*** (0.0469)	-0.0708 (0.0381)	-0.126** (0.0394)	0.819*** (0.0400)	0.563*** (0.0505)	0.0387 (0.0352)	-0.210*** (0.0374)
Threshold 3	1.375*** (0.0660)	1.378*** (0.0454)	1.450*** (0.0435)	1.341*** (0.0424)	1.224*** (0.0513)	1.189*** (0.0421)	1.322*** (0.0447)	1.669*** (0.0467)	1.864*** (0.0652)	1.407*** (0.0413)	1.107*** (0.0410)
Threshold 4	2.666*** (0.0930)	2.751*** (0.0648)	2.649*** (0.0608)	2.796*** (0.0671)	2.571*** (0.0677)	2.563*** (0.0577)	2.527*** (0.0586)	2.795*** (0.0656)	2.916*** (0.0926)	2.680*** (0.0588)	2.304*** (0.0533)
N	1187	2526	2673	2609	2162	2950	2819	2581	1447	3128	2780

Table D2. Ordered probit model (year 2011). Predictions of these equations generate the health index.

Dependent variable: SAH	(1) Austria	(2) Germany	(3) Sweden	(4) Netherlands	(5) Spain	(6) Italy	(7) France	(8) Denmark	(9) Switzerland	(10) Belgium	(11) Czech Slovakia
heart attack	0.657***	0.520***	0.631***	0.629***	0.692***	0.785***	0.725***	0.794***	0.587***	0.617***	0.644***
	(0.0498)	(0.0956)	(0.0700)	(0.0701)	(0.0624)	(0.0657)	(0.0453)	(0.0888)	(0.0719)	(0.0500)	(0.0422)
high blood pressure	0.331***	0.343***	0.375***	0.354***	0.325***	0.186***	0.281***	0.250***	0.302***	0.219***	0.382***
	(0.0329)	(0.0585)	(0.0535)	(0.0482)	(0.0411)	(0.0398)	(0.0325)	(0.0518)	(0.0405)	(0.0329)	(0.0301)
high blood cholesterol	0.1000**	0.00596	0.124	0.190**	0.112**	0.115*	0.0590	0.0458	0.0335	0.0117	0.0744*
	(0.0384)	(0.0690)	(0.0682)	(0.0581)	(0.0428)	(0.0460)	(0.0341)	(0.0587)	(0.0509)	(0.0333)	(0.0355)
stroke	0.493***	0.726***	0.479***	0.669***	0.520***	1.019***	0.627***	0.687***	0.619***	0.619***	0.638***
	(0.0743)	(0.149)	(0.128)	(0.114)	(0.127)	(0.127)	(0.0916)	(0.135)	(0.121)	(0.0837)	(0.0712)
diabetes	0.576***	0.544***	0.425***	0.569***	0.511***	0.455***	0.565***	0.675***	0.617***	0.490***	0.472***
	(0.0496)	(0.0826)	(0.0827)	(0.0703)	(0.0525)	(0.0576)	(0.0480)	(0.0884)	(0.0740)	(0.0504)	(0.0398)
chronic lung disease	0.634***	0.727***	0.665***	0.743***	0.785***	0.685***	0.634***	0.742***	0.575***	0.767***	0.451***
	(0.0681)	(0.101)	(0.119)	(0.0823)	(0.0821)	(0.0776)	(0.0617)	(0.0848)	(0.0911)	(0.0642)	(0.0603)
arthritis	0.505***	0.649***	0.531***	0.528***	0.595***	0.560***	0.490***	0.640***	0.599***	0.508***	0.571***
	(0.0444)	(0.0792)	(0.0757)	(0.0721)	(0.0434)	(0.0426)	(0.0334)	(0.0538)	(0.0466)	(0.0350)	(0.0383)
cancer	0.355***	0.861***	0.711***	0.591***	0.849***	1.119***	0.698***	0.715***	0.551***	0.606***	0.759***
	(0.0653)	(0.120)	(0.129)	(0.0994)	(0.102)	(0.108)	(0.0628)	(0.135)	(0.0760)	(0.0637)	(0.0666)
stomach/ duodenal ulcer	0.336***	0.590***	0.110	0.309	0.443***	0.364***	0.425***	0.0997	0.425**	0.252***	0.206***
	(0.0697)	(0.169)	(0.211)	(0.158)	(0.101)	(0.104)	(0.0766)	(0.147)	(0.129)	(0.0579)	(0.0583)
parkinson	1.099***	1.848***	1.547***	1.418**	1.431***	1.024***	1.332***	1.052***	1.580***	0.710***	1.124***
	(0.206)	(0.434)	(0.308)	(0.517)	(0.253)	(0.246)	(0.190)	(0.262)	(0.307)	(0.163)	(0.200)
cataracts	0.0670	0.0808	0.202*	0.303***	0.180**	0.361***	0.142**	0.189*	-0.0879	0.0464	0.119*
	(0.0549)	(0.0930)	(0.0854)	(0.0884)	(0.0654)	(0.0808)	(0.0542)	(0.0904)	(0.0627)	(0.0584)	(0.0512)

Table D2. *cont.*

hip fracture	0.635***	0.635**	0.319*	-0.0477	0.595***	0.498***	0.398***	0.230	0.457**	0.262**	0.666***
	(0.0844)	(0.219)	(0.158)	(0.219)	(0.132)	(0.146)	(0.0936)	(0.233)	(0.151)	(0.0907)	(0.102)
other conditions	0.426***	0.545***	0.623***	0.662***	0.595***	0.485***	0.669***	0.690***	0.577***	0.557***	0.549***
	(0.0427)	(0.0747)	(0.0641)	(0.0558)	(0.0482)	(0.0558)	(0.0433)	(0.0638)	(0.0519)	(0.0432)	(0.0396)
alzheimer/ dementia	0.963***	0.869***	1.138***	0.572**	0.670***	1.578***	0.692***	0.449	0.757***	0.617***	0.954***
	(0.130)	(0.257)	(0.220)	(0.217)	(0.180)	(0.264)	(0.139)	(0.302)	(0.229)	(0.162)	(0.165)
depression	0.406***	0.360***	0.376***	0.418***	0.527***	0.454***	0.267***	0.287***	0.308***	0.365***	0.463***
	(0.0328)	(0.0573)	(0.0531)	(0.0456)	(0.0406)	(0.0400)	(0.0293)	(0.0501)	(0.0363)	(0.0315)	(0.0303)
low grip	0.156***	0.164**	0.208***	0.250***	0.289***	0.252***	0.220***	0.244***	0.170***	0.270***	0.240***
	(0.0340)	(0.0631)	(0.0533)	(0.0469)	(0.0406)	(0.0413)	(0.0312)	(0.0522)	(0.0389)	(0.0333)	(0.0320)
no grip	0.556***	0.610***	0.925***	0.783***	0.852***	0.513***	0.689***	0.910***	0.671***	0.780***	0.548***
	(0.0520)	(0.116)	(0.147)	(0.113)	(0.0826)	(0.0629)	(0.0575)	(0.156)	(0.116)	(0.0808)	(0.0617)
Threshold 1	-0.764***	-1.142***	-0.391***	-0.618***	-1.087***	-0.853***	-0.864***	-0.246***	-0.675***	-0.881***	-1.277***
	(0.0307)	(0.0708)	(0.0469)	(0.0398)	(0.0495)	(0.0401)	(0.0328)	(0.0406)	(0.0343)	(0.0325)	(0.0377)
Threshold 2	0.312***	-0.229***	0.516***	0.0925*	-0.0493	-0.0161	-0.0667*	0.886***	0.423***	0.106***	-0.112***
	(0.0276)	(0.0558)	(0.0461)	(0.0366)	(0.0378)	(0.0345)	(0.0285)	(0.0424)	(0.0321)	(0.0280)	(0.0271)
Threshold 3	1.430***	1.179***	1.416***	1.461***	1.314***	1.255***	1.312***	1.759***	1.734***	1.452***	1.234***
	(0.0321)	(0.0602)	(0.0522)	(0.0434)	(0.0421)	(0.0386)	(0.0320)	(0.0508)	(0.0402)	(0.0327)	(0.0301)
Threshold 4	2.742***	2.630***	2.593***	2.965***	2.671***	2.614***	2.422***	2.871***	2.931***	2.700***	2.459***
	(0.0461)	(0.0832)	(0.0700)	(0.0691)	(0.0554)	(0.0536)	(0.0404)	(0.0712)	(0.0615)	(0.0454)	(0.0388)
N	5087	1548	1919	2738	3419	3505	5626	2237	3722	5204	5946

Table D3. Ordered probit model (year 2013). Predictions of these equations generate the health index.

Dependent variable: SAH	**(1) Austria**	**(2) Germany**	**(3) Sweden**	**(4) Netherlands**	**(5) Spain**	**(6) Italy**	**(7) France**	**(8) Denmark**	**(9) Switzerland**	**(10) Belgium**	**(11) Czech Slovakia**
heart attack	0.577***	0.556***	0.693***	0.619***	0.676***	0.682***	0.748***	0.558***	0.628***	0.618***	0.569***
	(0.0586)	(0.0485)	(0.0581)	(0.0585)	(0.0494)	(0.0589)	(0.0524)	(0.0588)	(0.0826)	(0.0508)	(0.0462)
high blood pressure	0.296***	0.369***	0.351***	0.240***	0.284***	0.385***	0.311***	0.233***	0.426***	0.228***	0.333***
	(0.0364)	(0.0306)	(0.0347)	(0.0399)	(0.0304)	(0.0348)	(0.0364)	(0.0381)	(0.0456)	(0.0323)	(0.0306)
high blood cholesterol	0.0780	0.0803*	0.0830	0.187***	0.187***	0.187***	0.0432	0.0811	0.0899	0.0820*	0.212***
	(0.0431)	(0.0373)	(0.0459)	(0.0466)	(0.0322)	(0.0401)	(0.0402)	(0.0422)	(0.0580)	(0.0331)	(0.0355)
stroke	0.472***	0.607***	0.582***	0.571***	0.704***	1.041***	0.657***	0.641***	0.915***	0.649***	0.667***
	(0.0852)	(0.0717)	(0.0733)	(0.0980)	(0.105)	(0.108)	(0.0987)	(0.0905)	(0.152)	(0.0913)	(0.0664)
diabetes	0.531***	0.588***	0.583***	0.633***	0.492***	0.452***	0.506***	0.401***	0.530***	0.501***	0.347***
	(0.0542)	(0.0451)	(0.0544)	(0.0591)	(0.0413)	(0.0521)	(0.0528)	(0.0633)	(0.0798)	(0.0481)	(0.0392)
chronic lung disease	0.544***	0.593***	0.669***	0.744***	0.765***	0.835***	0.726***	0.705***	0.759***	0.816***	0.596***
	(0.0756)	(0.0559)	(0.0822)	(0.0618)	(0.0645)	(0.0746)	(0.0716)	(0.0654)	(0.104)	(0.0616)	(0.0602)
arthritis	0.583***	0.561***	0.675***	0.700***	0.621***	0.515***	0.662***	0.745***	0.529***	0.453***	0.352***
	(0.0611)	(0.0486)	(0.106)	(0.0892)	(0.0403)	(0.0555)	(0.102)	(0.101)	(0.100)	(0.0541)	(0.0434)
cancer	0.551***	0.501***	0.478***	0.581***	0.631***	0.711***	0.841***	0.388***	0.746***	0.677***	0.704***
	(0.0916)	(0.0502)	(0.0572)	(0.0735)	(0.0686)	(0.0829)	(0.0767)	(0.0705)	(0.103)	(0.0695)	(0.0671)
stomach/ duodenal ulcer	0.306***	0.504***	0.214*	0.241	0.323***	0.210*	0.415***	0.497***	0.455*	0.468***	0.335***
	(0.0908)	(0.0748)	(0.0894)	(0.129)	(0.0761)	(0.0936)	(0.104)	(0.0965)	(0.185)	(0.0636)	(0.0720)
parkinson	1.219***	1.007***	1.105***	1.473***	0.881***	1.174***	1.444***	1.833***	1.395***	1.021***	1.050***
	(0.208)	(0.180)	(0.200)	(0.283)	(0.152)	(0.237)	(0.178)	(0.261)	(0.314)	(0.162)	(0.174)
cataracts	0.204***	0.0978*	0.167***	0.0356	0.235***	0.243***	0.231***	0.117	0.0956	0.173**	0.220***
	(0.0611)	(0.0494)	(0.0497)	(0.0694)	(0.0517)	(0.0684)	(0.0662)	(0.0628)	(0.0798)	(0.0575)	(0.0486)

Table D3. *cont.*

hip fracture	0.989***	0.631***	0.371***	0.360*	0.536***	0.484***	0.723***	0.426**	0.393*	0.241*	0.392***
	(0.159)	(0.107)	(0.0858)	(0.142)	(0.107)	(0.128)	(0.148)	(0.154)	(0.161)	(0.0981)	(0.103)
other conditions	0.560***	0.614***	0.540***	0.575***	0.506***	0.542***	0.498***	0.603***	0.697***	0.535***	0.500***
	(0.0490)	(0.0390)	(0.0394)	(0.0437)	(0.0346)	(0.0492)	(0.0494)	(0.0437)	(0.0615)	(0.0400)	(0.0415)
alzheimer/ dementia	0.637***	0.800***	0.880***	0.487**	0.960***	1.098***	0.708***	0.690**	0.531	0.552***	0.536**
	(0.130)	(0.148)	(0.158)	(0.181)	(0.118)	(0.163)	(0.172)	(0.235)	(0.287)	(0.129)	(0.175)
depression	0.446***	0.337***	0.377***	0.358***	0.572***	0.554***	0.357***	0.378***	0.295***	0.397***	0.385***
	(0.0373)	(0.0294)	(0.0349)	(0.0374)	(0.0313)	(0.0343)	(0.0333)	(0.0369)	(0.0410)	(0.0305)	(0.0301)
low grip	0.255***	0.257***	0.291***	0.183***	0.271***	0.275***	0.324***	0.290***	0.228***	0.266***	0.260***
	(0.0385)	(0.0333)	(0.0353)	(0.0381)	(0.0304)	(0.0354)	(0.0343)	(0.0384)	(0.0435)	(0.0318)	(0.0319)
no grip	0.667***	0.669***	0.815***	0.828***	0.942***	0.706***	0.924***	1.069***	0.587***	0.907***	0.760***
	(0.0669)	(0.0683)	(0.105)	(0.0851)	(0.0569)	(0.0545)	(0.0801)	(0.114)	(0.101)	(0.0700)	(0.0592)
Threshold 1	-0.871***	-0.935***	-0.281***	-0.637***	-1.159***	-0.757***	-0.946***	-0.216***	-0.745***	-0.889***	-1.115***
	(0.0362)	(0.0330)	(0.0300)	(0.0317)	(0.0367)	(0.0346)	(0.0375)	(0.0294)	(0.0383)	(0.0313)	(0.0370)
Threshold 2	0.242***	-0.0999***	0.572***	-0.0227	-0.0605*	-0.0337	-0.135***	0.790***	0.330***	0.0970***	-0.193***
	(0.0315)	(0.0280)	(0.0299)	(0.0293)	(0.0278)	(0.0305)	(0.0322)	(0.0306)	(0.0351)	(0.0271)	(0.0293)
Threshold 3	1.385***	1.219***	1.629***	1.369***	1.349***	1.210***	1.258***	1.597***	1.746***	1.513***	1.120***
	(0.0364)	(0.0310)	(0.0355)	(0.0346)	(0.0315)	(0.0337)	(0.0358)	(0.0357)	(0.0447)	(0.0322)	(0.0314)
Threshold 4	2.686***	2.552***	2.807***	2.821***	2.645***	2.577***	2.402***	2.709***	2.939***	2.810***	2.311***
	(0.0511)	(0.0415)	(0.0501)	(0.0544)	(0.0416)	(0.0468)	(0.0457)	(0.0512)	(0.0695)	(0.0463)	(0.0393)
N	3970	5632	4475	4084	6115	4578	4391	4085	2991	5532	5578

Table D4. Descriptive statics of mapped SAH (using HUI) for each country in wave 2 (2006).

	Obs	Mean	Std. Dev.	Min	Max
Austria	760	0.837	0.102	0.557	0.945
Germany	1286	0.818	0.107	0.557	0.945
Sweden	1772	0.853	0.103	0.557	0.945
Netherlands	1449	0.848	0.093	0.557	0.945
Spain	908	0.799	0.116	0.557	0.945
Italy	1642	0.809	0.115	0.557	0.945
France	1611	0.828	0.105	0.557	0.945
Denmark	1552	0.864	0.101	0.557	0.945
Switzerland	852	0.874	0.080	0.557	0.945
Belgium	1747	0.843	0.098	0.557	0.945
Czech Rep.	1413	0.803	0.116	0.557	0.945

Table D5. Trends in IRHI ranking in Europe, inequality aversion parameter = 2 (standard C, W and E).

	C(h)				C(s)				W				E			
Countries	**Wave 2**	**Wave 4**	**Wave 5**	**Ranking trend**	**Wave 2**	**Wave 4**	**Wave 5**	**Ranking trend**	**Wave 2**	**Wave 4**	**Wave 5**	**Ranking trend**	**Wave 2**	**Wave 4**	**Wave 5**	**Ranking trend**
AUSTRIA	3	9	4	**stable/increasing**	6	11	4	**unclear**	5	11	4	**stable/unclear**	3	11	4	**stable/increasing**
GERMANY	7	2	1	decreasing	8	4	3	decreasing	8	2	3	decreasing	8	2	1	decreasing
SWEDEN	1	4	5	increasing	2	3	5	increasing	2	4	5	increasing	2	4	5	increasing
NETHERLANDS	9	3	2	decreasing	9	2	1	decreasing	9	3	1	decreasing	10	3	2	decreasing
SPAIN	11	8	7	**decreasing**	11	10	10	**stable/decreasing**	11	10	10	**stable/decreasing**	11	8	9	**decreasing**
ITALY	9	9	11	**stable/increasing**	10	9	11	**stable/unclear**	10	9	11	**stable/unclear**	9	10	11	**increasing**
FRANCE	7	5	7	**stable**	7	5	6	**decreasing**	7	5	6	**decreasing**	7	5	7	**stable**
DENMARK	1	1	3	stable/decreasing	1	1	2	stable/decreasing	1	1	2	stable/decreasing	1	1	3	stable/decreasing
SWITZERLAND	5	6	7	increasing	3	6	8	increasing	3	6	8	increasing	5	6	8	increasing
BELGIUM	6	6	6	**stable**	5	6	7	**increasing**	6	6	7	**stable/increasing**	6	6	6	**stable**
CZECH REP	4	9	10	increasing	4	8	9	increasing	4	8	9	increasing	4	9	10	increasing

1 equals the highest IRHI level, thus, decreasing trend in ranking implies relatively higher IRHI level, compared to the other countries. Bold trends means different indications of trends by different indices (for the same country).

Table D6. Description of 'overall trend' notations in tables D7-D10.

➘	Decreasing trend in IRHI.
➚	Increasing trend in IRHI.
stable	Stable trend.
–	Unable to determine direction of trend (the trend from wave 2 to wave 4 indicate opposite trend compared to the trend from wave 2 to wave 5).
(➚)	Due to the insignificance of one index in one wave, the trend interpretation is based on only two waves and should be interpreted cautiously.
Not significant	At least two waves are insignificant, thus no trend can be detected.

Table D7. IRHI trends in Europe, inequality aversion parameter = 1.5

Countries	Overall trend	C(h) (1) Wave 2	C(h) (2) Wave 4	C(h) (3) Wave 5	C(s) (4) Wave 2	C(s) (5) Wave 4	C(s) (6) Wave 5	W (7) Wave 2	W (8) Wave 4	W (9) Wave 5	E (10) Wave 2	E (11) Wave 4	E (12) Wave 5
AUSTRIA	↘	**0.010*** **	**0.004*** **	**0.008*** **	**0.036*** **	**0.021*** **	**0.034*** **	**0.046*** **	**0.025*** **	**0.042*** **	**0.052*** **	**0.024*** **	**0.042*** **
		(0.003)	(0.001)	(0.001)	(0.010)	(0.005)	(0.007)	(0.013)	(0.006)	(0.008)	(0.014)	(0.006)	(0.008)
GERMANY	↗	**0.006*** **	**0.012*** **	**0.012*** **	**0.023*** **	**0.041*** **	**0.048*** **	**0.030*** **	**0.053*** **	**0.060*** **	**0.033*** **	**0.061*** **	**0.064*** **
		(0.002)	(0.002)	(0.001)	(0.008)	(0.009)	(0.004)	(0.010)	(0.011)	(0.005)	(0.011)	(0.013)	(0.005)
SWEDEN	↘	**0.011*** **	**0.008*** **	**0.005*** **	**0.049*** **	**0.042*** **	**0.029*** **	**0.060*** **	**0.051*** **	**0.034*** **	**0.063*** **	**0.047*** **	**0.029*** **
		(0.002)	(0.002)	(0.001)	(0.009)	(0.010)	(0.006)	(0.011)	(0.011)	(0.007)	(0.012)	(0.011)	(0.006)
NETHERLANDS	↗	**0.003***	**0.011*** **	**0.010*** **	**0.018***	**0.046*** **	**0.058*** **	**0.021***	**0.057*** **	**0.067*** **	**0.018***	**0.058*** **	**0.056*** **
		(0.002)	(0.002)	(0.001)	(0.010)	(0.007)	(0.007)	(0.011)	(0.009)	(0.008)	(0.010)	(0.009)	(0.007)
SPAIN	↗	**0.003** **	**0.005*** **	**0.006*** **	**0.015** **	**0.021*** **	**0.027*** **	**0.019** **	**0.026*** **	**0.033*** **	**0.017** **	**0.027*** **	**0.031*** **
		(0.002)	(0.002)	(0.001)	(0.008)	(0.006)	(0.006)	(0.009)	(0.008)	(0.008)	(0.009)	(0.008)	(0.007)
ITALY	**stable**	**0.004*** **	**0.005** **	**0.003*** **	**0.021*** **	**0.027** **	**0.022*** **	**0.025*** **	**0.031** **	**0.025*** **	**0.024*** **	**0.026** **	**0.019*** **
		(0.001)	(0.001)	(0.001)	(0.007)	(0.009)	(0.007)	(0.009)	(0.010)	(0.008)	(0.008)	(0.008)	(0.006)
FRANCE	**stable**	**0.005*** **	**0.005** **	**0.005*** **	**0.028*** **	**0.027** **	**0.029*** **	**0.034*** **	**0.031** **	**0.033*** **	**0.030*** **	**0.026** **	**0.027*** **
		(0.002)	(0.001)	(0.001)	(0.009)	(0.009)	(0.006)	(0.011)	(0.010)	(0.007)	(0.010)	(0.008)	(0.005)
DENMARK	–	**0.012*** **	**0.013*** **	**0.008*** **	**0.059*** **	**0.051*** **	**0.051*** **	**0.071*** **	**0.063*** **	**0.059*** **	**0.065*** **	**0.070*** **	**0.048*** **
		(6.227)	(6.495)	(7.811)	(6.227)	(6.495)	(7.811)	(6.227)	(6.495)	(7.811)	(6.227)	(6.495)	(7.811)
SWITZERLAND	↘	**0.008*** **	**0.006*** **	**0.004*** **	**0.044*** **	**0.032*** **	**0.025*** **	**0.051*** **	**0.038*** **	**0.029*** **	**0.044*** **	**0.033*** **	**0.024*** **
		(0.002)	(0.001)	(0.001)	(0.009)	(0.007)	(0.007)	(0.011)	(0.008)	(0.009)	(0.009)	(0.007)	(0.007)
BELGIUM	↘	**0.007*** **	**0.006*** **	**0.005*** **	**0.037*** **	**0.032*** **	**0.028*** **	**0.043*** **	**0.038*** **	**0.033*** **	**0.038*** **	**0.032*** **	**0.030*** **
		(0.001)	(0.001)	(0.001)	(0.007)	(0.006)	(0.006)	(0.009)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)
CZECH REP	↘	**0.009*** **	**0.004*** **	**0.005*** **	**0.037*** **	**0.027*** **	**0.027*** **	**0.046*** **	**0.031*** **	**0.032*** **	**0.049*** **	**0.026*** **	**0.027*** **
		(0.003)	(0.001)	(0.001)	(0.010)	(0.005)	(0.006)	(0.013)	(0.006)	(0.007)	(0.014)	(0.005)	(0.006)

Standard error in parentheses.
*p<0.10 **p<0.05 ***p<0.01

Table D8. IRHI trends in Europe, inequality aversion parameter = 4.

Countries	Overall trend	C(h)			C(s)			W			E		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5
AUSTRIA	↘	**0.027*****	**0.009****	**0.020*****	**-0.103*****	**-0.046****	**-0.090*****	**0.130*****	**0.056****	**0.110*****	**0.046*****	**0.016****	**0.034*****
		(0.009)	(0.004)	(0.005)	(0.034)	(0.018)	(0.022)	(0.043)	(0.022)	(0.027)	(0.015)	(0.006)	(0.008)
GERMANY	↗	**0.012***	**0.034*****	**0.035*****	**-0.044***	**-0.124*****	**-0.142*****	**0.055***	**0.158*****	**0.177*****	**0.020***	**0.057*****	**0.059*****
		(0.007)	(0.009)	(0.004)	(0.025)	(0.031)	(0.015)	(0.032)	(0.040)	(0.018)	(0.011)	(0.014)	(0.006)
SWEDEN	↘	**0.035*****	**0.026*****	**0.020*****	**-0.147*****	**-0.130*****	**-0.113*****	**0.182*****	**0.156*****	**0.133*****	**0.060*****	**0.045*****	**0.036*****
		(0.007)	(0.006)	(0.003)	(0.031)	(0.033)	(0.018)	(0.039)	(0.039)	(0.021)	(0.013)	(0.011)	(0.006)
NETHERLANDS	↗	**0.010***	**0.029*****	**0.032*****	**-0.060***	**-0.128*****	**-0.189*****	**0.070***	**0.157*****	**0.220*****	**0.019***	**0.050*****	**0.057*****
		(0.006)	(0.006)	(0.004)	(0.035)	(0.027)	(0.024)	(0.041)	(0.033)	(0.028)	(0.011)	(0.010)	(0.007)
SPAIN	not significant	**0.007**	**0.015*****	**0.007**	**-0.038**	**-0.062*****	**-0.033**	**0.045**	**0.077*****	**0.040**	**0.013**	**0.025*****	**0.012**
		(0.005)	(0.005)	(0.005)	(0.026)	(0.021)	(0.022)	(0.031)	(0.026)	(0.027)	(0.009)	(0.009)	(0.008)
ITALY	**stable**	**0.010***	**0.008***	**0.005**	**-0.047***	**-0.045***	**-0.036**	**0.056***	**0.053***	**0.041**	**0.017***	**0.014***	**0.009**
		(0.005)	(0.004)	(0.004)	(0.025)	(0.026)	(0.026)	(0.030)	(0.030)	(0.030)	(0.009)	(0.008)	(0.007)
FRANCE	**stable**	**0.017*****	**0.025*****	**0.017*****	**-0.091*****	**-0.136*****	**-0.103*****	**0.108*****	**0.162*****	**0.120*****	**0.031*****	**0.045*****	**0.030*****
		(0.006)	(0.003)	(0.003)	(0.029)	(0.019)	(0.019)	(0.035)	(0.022)	(0.022)	(0.010)	(0.006)	(0.006)
DENMARK	–	**0.037*****	**0.043*****	**0.029*****	**-0.190*****	**-0.168*****	**-0.181*****	**0.227*****	**0.212*****	**0.211*****	**0.066*****	**0.073*****	**0.054*****
		(0.007)	(0.007)	(0.004)	(0.038)	(0.028)	(0.022)	(0.046)	(0.035)	(0.026)	(0.013)	(0.012)	(0.007)
SWITZERLAND	↘	**0.025*****	**0.019*****	**0.015*****	**-0.141*****	**-0.105*****	**-0.092*****	**0.166*****	**0.124*****	**0.107*****	**0.045*****	**0.034*****	**0.028*****
		(0.005)	(0.004)	(0.004)	(0.030)	(0.023)	(0.025)	(0.035)	(0.027)	(0.029)	(0.009)	(0.007)	(0.008)
BELGIUM	↘	**0.020*****	**0.019*****	**0.018*****	**-0.111*****	**-0.106*****	**-0.097*****	**0.131*****	**0.124*****	**0.115*****	**0.036*****	**0.033*****	**0.032*****
		(0.005)	(0.004)	(0.004)	(0.026)	(0.021)	(0.021)	(0.031)	(0.025)	(0.024)	(0.008)	(0.007)	(0.007)
CZECH REP	↘	**0.027*****	**0.009*****	**0.010*****	**-0.109*****	**-0.054*****	**-0.056*****	**0.136*****	**0.063*****	**0.065*****	**0.046*****	**0.016*****	**0.018*****
		(0.008)	(0.003)	(0.003)	(0.034)	(0.017)	(0.016)	(0.042)	(0.020)	(0.019)	(0.014)	(0.005)	(0.005)

Standard error in parentheses.
*p<0.10 **p<0.05 ***p<0.01

Table D9. IRHI trends in Europe, inequality aversion parameter = 6.

		C(h)			C(s)			W			E		
	Overall trend	(1) Wave 2	(2) Wave 4	(3) Wave 5	(4) Wave 2	(5) Wave 4	(6) Wave 5	(7) Wave 2	(8) Wave 4	(9) Wave 5	(10) Wave 2	(11) Wave 4	(12) Wave 5
AUSTRIA	↘	**0.029****	**0.008**	**0.021*****	**-0.110****	**-0.040**	**-0.097*****	**0.140****	**0.048**	**0.119*****	**0.040****	**0.011**	**0.030*****
		(0.012)	(0.005)	(0.006)	(0.046)	(0.024)	(0.029)	(0.058)	(0.029)	(0.036)	(0.017)	(0.007)	(0.009)
GERMANY	(↗)	**0.011**	**0.041*****	**0.039*****	**-0.041**	**-0.146*****	**-0.161*****	**0.053**	**0.187*****	**0.201*****	**0.015**	**0.055*****	**0.054*****
		(0.009)	(0.012)	(0.005)	(0.034)	(0.043)	(0.020)	(0.043)	(0.055)	(0.025)	(0.012)	(0.016)	(0.007)
SWEDEN	↘	**0.038*****	**0.028*****	**0.024*****	**-0.162*****	**-0.144*****	**-0.137*****	**0.200*****	**0.172*****	**0.161*****	**0.053*****	**0.040*****	**0.035*****
		(0.010)	(0.008)	(0.004)	(0.041)	(0.043)	(0.024)	(0.050)	(0.052)	(0.028)	(0.013)	(0.012)	(0.006)
NETHERLANDS	(↗)	**0.010**	**0.033*****	**0.038*****	**-0.059**	**-0.143*****	**-0.226*****	**0.069**	**0.175*****	**0.264*****	**0.015**	**0.046*****	**0.056*****
		(0.008)	(0.008)	(0.005)	(0.046)	(0.035)	(0.033)	(0.054)	(0.044)	(0.038)	(0.012)	(0.011)	(0.008)
SPAIN	not	**0.006**	**0.017****	**0.002**	**-0.032**	**-0.069****	**-0.011**	**0.038**	**0.085****	**0.013**	**0.009**	**0.023****	**0.003**
	significant	(0.007)	(0.007)	(0.006)	(0.035)	(0.027)	(0.030)	(0.041)	(0.034)	(0.036)	(0.010)	(0.009)	(0.009)
ITALY	not	**0.011**	**0.005**	**0.004**	**-0.053**	**-0.032**	**-0.025**	**0.064**	**0.038**	**0.028**	**0.016**	**0.008**	**0.005**
	significant	(0.007)	(0.006)	(0.005)	(0.033)	(0.035)	(0.034)	(0.040)	(0.040)	(0.039)	(0.010)	(0.009)	(0.007)
FRANCE	↗	**0.020*****	**0.031*****	**0.020*****	**-0.105*****	**-0.166*****	**-0.126*****	**0.125*****	**0.197*****	**0.146*****	**0.029*****	**0.045*****	**0.030*****
		(0.007)	(0.005)	(0.004)	(0.037)	(0.025)	(0.025)	(0.044)	(0.030)	(0.029)	(0.010)	(0.007)	(0.006)
DENMARK	–	**0.044*****	**0.050*****	**0.035*****	**-0.225*****	**-0.196*****	**-0.218*****	**0.269*****	**0.246*****	**0.253*****	**0.063*****	**0.069*****	**0.052*****
		(0.010)	(0.010)	(0.005)	(0.052)	(0.038)	(0.030)	(0.062)	(0.048)	(0.035)	(0.015)	(0.013)	(0.007)
SWITZERLAND	↘	**0.026*****	**0.021*****	**0.016*****	**-0.148*****	**-0.120*****	**-0.098*****	**0.174*****	**0.141*****	**0.115*****	**0.038*****	**0.031*****	**0.024*****
		(0.007)	(0.005)	(0.006)	(0.039)	(0.031)	(0.033)	(0.046)	(0.036)	(0.039)	(0.010)	(0.008)	(0.008)
BELGIUM	**stable**	**0.022*****	**0.021*****	**0.022*****	**-0.122*****	**-0.119*****	**-0.119*****	**0.144*****	**0.140*****	**0.141*****	**0.032*****	**0.030*****	**0.032*****
		(0.006)	(0.005)	(0.005)	(0.035)	(0.028)	(0.028)	(0.042)	(0.033)	(0.033)	(0.009)	(0.007)	(0.007)
CZECH REP	↘	**0.029****	**0.008****	**0.009****	**-0.117****	**-0.048****	**-0.051****	**0.146****	**0.056****	**0.060****	**0.040****	**0.012****	**0.013****
		(0.011)	(0.004)	(0.004)	(0.045)	(0.022)	(0.020)	(0.057)	(0.026)	(0.024)	(0.016)	(0.006)	(0.005)

Standard error in parentheses.
$^{*}p<0.10$ $^{**}p<0.05$ $^{***}p<0.01$

Table D10. IRHI trends in Europe, inequality aversion parameter = 8.

Countries	Overall trend	C(h)			C(s)			W			E		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5	Wave 2	Wave 4	Wave 5
AUSTRIA	(↘)	**0.030***	**0.005**	**0.021*** **	**-0.113***	**-0.026**	**-0.095*** **	**0.143***	**0.032**	**0.116*** **	**0.036*** **	**0.007**	**0.026*** **
		(0.015)	(0.006)	(0.008)	(0.056)	(0.028)	(0.035)	(0.071)	(0.034)	(0.042)	(0.018)	(0.007)	(0.010)
GERMANY	–	**0.011**	**0.045*** **	**0.041*** **	**-0.041**	**-0.161*** **	**-0.166*** **	**0.051**	**0.205*** **	**0.206*** **	**0.013**	**0.053*** **	**0.050*** **
		(0.011)	(0.015)	(0.006)	(0.040)	(0.052)	(0.024)	(0.051)	(0.067)	(0.030)	(0.013)	(0.017)	(0.007)
SWEDEN	↘	**0.038*** **	**0.028*** **	**0.026*** **	**-0.161*** **	**-0.141*** **	**-0.145*** **	**0.199*** **	**0.168*** **	**0.171*** **	**0.047*** **	**0.035*** **	**0.034*** **
		(0.011)	(0.010)	(0.005)	(0.047)	(0.050)	(0.029)	(0.059)	(0.060)	(0.034)	(0.014)	(0.013)	(0.007)
NETHERLANDS	(↗)	**0.009**	**0.033*** **	**0.041*** **	**-0.050**	**-0.146*** **	**-0.243*** **	**0.059**	**0.180*** **	**0.284*** **	**0.011**	**0.042*** **	**0.054*** **
		(0.009)	(0.010)	(0.007)	(0.054)	(0.042)	(0.040)	(0.063)	(0.052)	(0.046)	(0.012)	(0.012)	(0.009)
SPAIN	not significant	**0.004**	**0.016*** *	**-0.002**	**-0.022**	**-0.067*** *	**0.011**	**0.027**	**0.084*** *	**-0.013**	**0.006**	**0.020*** *	**-0.002**
		(0.008)	(0.008)	(0.008)	(0.041)	(0.033)	(0.036)	(0.049)	(0.041)	(0.044)	(0.010)	(0.010)	(0.010)
ITALY	not significant	**0.012**	**0.003**	**0.002**	**-0.057**	**-0.017**	**-0.011**	**0.068**	**0.019**	**0.012**	**0.015**	**0.004**	**0.001**
		(0.008)	(0.007)	(0.006)	(0.040)	(0.042)	(0.040)	(0.048)	(0.049)	(0.046)	(0.010)	(0.009)	(0.008)
FRANCE	↗	**0.020*** *	**0.034*** **	**0.023*** **	**-0.108*** *	**-0.181*** **	**-0.139*** **	**0.129*** *	**0.214*** **	**0.162*** **	**0.026*** *	**0.044*** **	**0.030*** **
		(0.008)	(0.006)	(0.005)	(0.042)	(0.030)	(0.029)	(0.050)	(0.036)	(0.034)	(0.010)	(0.007)	(0.006)
DENMARK	–	**0.047*** **	**0.054*** **	**0.039*** **	**-0.241*** **	**-0.210*** **	**-0.238*** **	**0.288*** **	**0.263*** **	**0.276*** **	**0.061*** **	**0.066*** **	**0.051*** **
		(0.012)	(0.012)	(0.006)	(0.063)	(0.046)	(0.037)	(0.075)	(0.058)	(0.042)	(0.016)	(0.014)	(0.008)
SWITZERLAND	↘	**0.024*** **	**0.022*** **	**0.015*** *	**-0.139*** **	**-0.124*** **	**-0.091*** *	**0.163*** **	**0.146*** **	**0.106*** *	**0.032*** **	**0.029*** **	**0.020*** *
		(0.008)	(0.007)	(0.007)	(0.046)	(0.036)	(0.039)	(0.054)	(0.043)	(0.046)	(0.010)	(0.009)	(0.009)
BELGIUM	-	**0.022*** **	**0.021*** **	**0.025*** **	**-0.122*** **	**-0.121*** **	**-0.133*** **	**0.144*** **	**0.142*** **	**0.158*** **	**0.028*** **	**0.028*** **	**0.032*** **
		(0.008)	(0.006)	(0.006)	(0.043)	(0.034)	(0.034)	(0.051)	(0.039)	(0.040)	(0.010)	(0.008)	(0.008)
CZECH REP	↘	**0.030*** *	**0.007**	**0.008***	**-0.121*** *	**-0.040**	**-0.043***	**0.151*** *	**0.047**	**0.050***	**0.037*** *	**0.009**	**0.010**
		(0.014)	(0.004)	(0.004)	(0.054)	(0.027)	(0.024)	(0.068)	(0.031)	(0.028)	(0.017)	(0.006)	(0.005)

Standard error in parentheses.
*p<0.10 **p<0.05 ***p<0.01

Mean ill-health (y-axis) across income deciles (x-axis) in three countries.

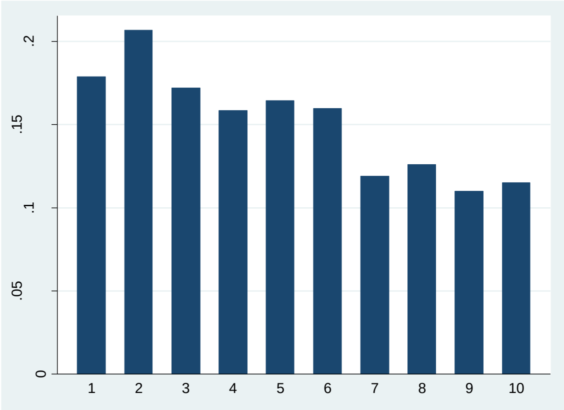


Figure D1. Belgium wave 2 (2006).

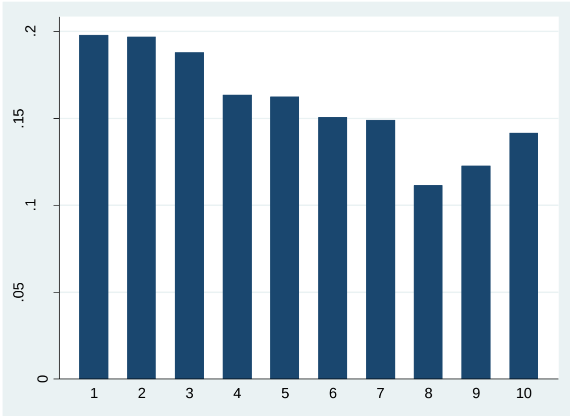


Figure D2. Belgium wave 5 (2013).

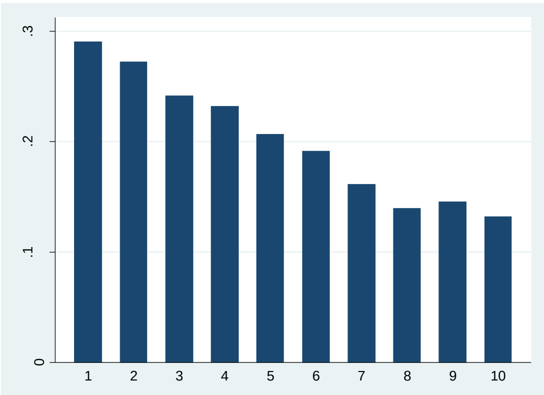


Figure D3. Denmark wave 4 (2011).

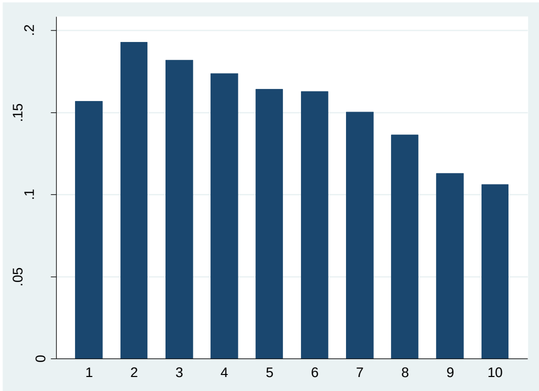


Figure D4. Czech Republic, wave 5 (2011).