Socioeconomic Inequality in Health and Health Care:
Measurement and Explanation

Sociaal-economische ongelijkheid in gezondheid en
gezondheidszorggebruik: meting en verklaring

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 12 mei 2006 om 13.30 uur

door

Alexander Hendrikus Engelbertus Koolman
geboren te Eindhoven

Promotiecommissie

Promotoren:
Prof.dr. E.K.A. van Doorslaer
Prof.dr. A.M. Jones

Overige leden:
Prof.dr. J.P. Mackenbach
Prof.dr. F.F.H. Rutten
Prof.dr. M. Sutton

ISBN 90-8559-174-0
Design of cover and interior by Pier

PUBLICATIONS

Chapters 2 to 7 are based on the following articles:

CHAPTER 2

with E.K.A. van Doorslaer: On the interpretation of a concentration index of inequality. Health Economics, 13:629-647 (2004).

CHAPTER 3

with A. Kunst: Estimating relative risks for categorical outcomes An application to health inequality measurement. Submitted for publication.

CHAPTER 4

with E.K.A. van Doorslaer: Explaining the differences in income-related health inequalities across European countries. Health Economics, 13: 609-628 (2004).

CHAPTER 5

with A.M. Jones and N. Rice: On the interpretation of a concentration index of inequality. Journal of Royal Statistical Society, Series A, 169:1-27 (2006).

CHAPTER 6

with E.K.A. van Doorslaer and A.M. Jones: Explaining income-related inequalities in doctor utilisation in Europe. Health Economics, 13: 649-656 (2004).

CHAPTER 7

with E.K.A. van Doorslaer and A.M. Jones: The impact of supplementary private health insurance on the use of specialists in selected European countries. Forthcoming in Annals Of Economics and Statistics.

Reprinted with permission.

CHAPTER 1

Introduction

1.1 INTRODUCTION

Empirical researchers in France, Germany, and the United Kingdom set out to measure inequality in health ever since the mid-nineteenth century (Farr 1839; Ackerknecht 1953; Coleman 1982). Although many studies went before it, the Black report (Townsend and Davidson 1982) played a crucial role in getting the issue of socioeconomic differences in health on the political agenda in a number of countries, in the United Kingdom especially. For that reason, it is sometimes regarded as the beginning of the current and continuing research into the measurement of health inequalities. In fact, the influential British Medical Journal regarded it the most important medical report since the Second World War.

Why do researchers study variation in health? Insight in variation in health may help to raise health levels, one might argue. That does not explain, however, why health variation is studied in relation to socioeconomic status - a complex social construct - rather than to specific (socioeconomic) factors that are thought to cause disease. Researchers seem to favour measures for socioeconomic status and other summary measures that relate to societal notions of an inequitable distribution of health (Wagstaff, Paci et al. 1991; Gakidou, Murray et al. 2000; Mackenbach, Bakker et al. 2002). So, variables that describe the (causal) relations as observed in the data best are infrequently used. It seems that inequalities in health are of most interest because of their moral concern. In fact, according to a World Health organisation (WHO) study across 51 countries (Gakidou, Murray et al. 2003), people consider the equitable distribution of health almost as important as the average level of health when evaluating the performance of a health care system.

In the next section we will discuss which part of the distribution of health this thesis is concerned with.

1.2 HEALTH INEQUALITY

1.2.1 Terminology

The normative nature of health equity studies is often obscured by popular terminology, such as inequality and disparity. Some argue that the term inequality is not inherently normative (Braveman and Gruskin 2003). Indeed, equality can be seen as a mathematical property of a distribution. Absolute health equality is however of little interest as arguing for total equality seems just as debatable as arguing for other ethical viewpoints. Health variation may result from choice or could compensate for unjust variation in other domains of life. Moreover, one can only test for total equality, because the mathematical definition of equality gives no clue as to how to measure the extent of inequality.

As concluded before, the measurement of the extent of inequality is of interest because it is of moral concern. Assuming there is variation in health that is or is not of moral concern, then health inequality studies aim to measure only the variation of the first sort. This is captured by the definition of socioeconomic inequalities in health used by Mackenbach, Bakker et al. (2002). It says that inequalities are equal to the "systematic differences in morbidity and mortality rates between individual people of higher and lower socioeconomic status to the extent that these are perceived to be unfair." While the above shows the normative use of the term "health inequality", for this thesis we have preferred it above the unequivocal term "health inequity" in line with the common practice in the literature to "avoid judgmental or moral connotations that may be associated with the use of health inequities" (Braveman and Gruskin 2003).

1.2.2 From theoretical ethics to empirical ethics

In order to measure the extent of health variation for normative purposes, we need an ethical stance. A host of ethical viewpoints is readily available. Arguing for one or the other would not do injustice to societal preferences, as few people - let alone societies - qualify as pure supporters of one ethical principle. Yet, the proximity to societal ethical preferences or, if one prefers, social justice, largely determines the value of empirical research in ethics. Techniques to unravel such preferences or notions about justice, such as stated preference analysis or "veil of ignorance" experiments, are based on strong assumptions (Dworkin 1977; Carneiro, Hansen et al. 2001; Heckman 2001; Bryan and Dolan 2004). Nonetheless, these techniques may ultimately be more defendable than the sometimes implicit but always debatable choices of researchers.

While such empirically grounded ethical stances are being gathered, many researchers have tried to develop equity principles. We do not aim to review these here, but we will a summary of the efforts by Whitehead (1985). She has been - and still is - very influential in SES-related differences in health and health care research.

Whitehead distinguishes seven causes of variation:

1. Natural, biological variation.
2. Health-damaging behaviour if freely chosen.
3. The transient health advantage of one group over another when that group is first to adopt a health-promoting behaviour.
4. Health-damaging behaviour where the degree of choice of lifestyles is severely restricted.
5. Exposure to unhealthy, stressful living and working conditions.
6. Inadequate access to essential health and other public services.
7. Natural selection or health-related social mobility involving the tendency for sick people to move down the social scale.

She argues that the first three causes listed are generally considered as just causes of variation, while the latter four are unjustified causes of variation. Whitehead acknowledges the role of individual responsibility, and thus is distinctly non-egalitarian. She also points out that variation in health caused by society and beyond the control of the individual is unjust, but variation out of control of both individual and society is just. While most researchers feel that society can be held more responsible for ill-health it causes, many feel it should still aim to remove inequality due to factors outside both individual and society.

Whitehead proposes the following definition: "Equity in health implies that ideally everyone should have a fair opportunity to attain their full health potential and, more pragmatically, that no one should be disadvantaged from achieving this potential, if it can be avoided" (Whitehead 1985). This definition shows that while she acknowledges that unavoidable causes for variation may exist, society should not be acquitted from remedying the resulting ill health. Yet, the definition does not require a society to pour unlimited resources into undoing variation it cannot reasonably avoid, in order to be equitable.

In line with the above, freely chosen health-damaging behaviour should also be removed from total health variation in order to obtain that part of the health variation that is viewed as inequitable. The question remains why people would choose to damage their health. It seems fair to assume that most people will only (knowingly) risk their health if some benefit is expected elsewhere, for example by receiving a higher wage when taking up a risky job. Costs and benefits of choices may well vary systematically between groups. Other causes for variation, not mentioned by Whitehead, could be bounded rationality or lack of information. Both may cause people to make health-damaging decisions.

All of the above would cause as much variation in health as it would in opinion about their classification of free or restricted choice. In the absence of consensus over societal preferences or principles of social justice, researchers cannot avoid making normative choices when measuring inequality. These commonly made normative choices may not hold. For example, society may give more weight to a "shortage" of health compared to a set standard of health than to "excess" health compared to that standard. In other words, excess health of one individual within a group may not compensate a shortage of health of another individual on a one to one base. Consequently people may prefer to be in the group with the least variation in health even though the groups have equal average health. Similarly, society may care differently about the loss of one extra health increment as the health of the individual deteriorates. Society may also care more about one individual than another, for example a young mother versus an 87-year old person. By not acknowledging such preferences when choosing between summary measures,

researchers are not avoiding normative choices; they are merely superimposing norms that may or may not command widespread support in society.

1.2.3 Choosing between measures of inequality

When measuring the extent of variation in health, further choices have to be made. These choices concern the health concept, the unit of time, and univariate versus bivariate or multivariate analysis. The concept of health in question depends on the reason why health variation is a cause for moral concern (Asada 2005). If health is an intrinsic part of utility or well-being, then we need to measure health-related quality of life or health utility. If health is seen as instrumental to other intrinsic components of well-being (Sen 1992), then health-related functional limitations should be focussed on. Variation, disorders in particular, are merely instrumental to both concepts of inequity, and thus important in explaining health variation, yet are of little intrinsic interest.

If health is an intrinsic part of wellbeing, then we probably want to evaluate health-related wellbeing over an individual's life span. If health is seen as a contributing factor to other intrinsic components of wellbeing, then health-related functions may serve a different purpose in different stages of life and may thus be valued differently in different life stages. For example, the loss of reproductive health may only worry people of a certain age. Weights for each health-related function at each stage in life is not yet empirically substantiated, but early work in this field is getting published (Anand and Dolan 2005; Brouwer, Van Exel et al. 2005).

Another question is whether health variation per se is of moral concern or only health variation that is related to other characteristics, such as SES. The first approach is sometimes referred to as univariate, (Braveman, Krieger et al. 2000) or individual (Murray, Gakidou et al. 1999), and the latter is referred to as bivariate or group based.[a] Both approaches provide answers to different questions (Braveman and Gruskin 2003).

Finally, a choice between relative and absolute measures of inequality has to be made. Both Wagstaff, Van Doorslaer et al. (1991) and Mackenbach and Kunst (1997) discuss measures of absolute differences that indicate the difference in expected health between the two extreme SES groups or individuals, and relative differences that indicate the absolute difference relative to either the expected health of one group or individual or to the average level. Both groups of authors and many other researchers in both social epidemiology and economics have focussed mostly on relative inequalities, because relative inequalities are insensitive to a change in the mean (Wagstaff, Paci et al. 1991) and are more suitable to evaluate egalitarian policies (Mackenbach, Cavelaars et al. 1997). In this thesis, we apply relative measures of inequality.

1.2.4 Inequalities in health or ill-health?

When applying relative measures of variation, the question of whether to focus on health or ill-health becomes important. Both the extent of the variation and the relative ranking of groups or countries depend upon this choice (Clarke, Gerdtham et al. 2002; chapter three). This can easily be illustrated. Let A and B be two countries, let health be defined as (quality adjusted) life expectancy that may vary between zero and one hundred years and ill-health be defined as a one hundred minus health. Table 1.1 depicts the average health levels in groups 1 and 2.

Clearly, the ratios are much more unequal when using ill-health rather than health. Perhaps more importantly, the ranking of countries depends on the choice of health outcome measure, with country B having lower health inequality than A when health is chosen, and vice versa when ill-health is chosen. The choice to measure inequality in ill-health rather than health probably explained the relatively high inequality in ill-health (mortality) of Sweden (Mackenbach, Kunst et al. 1997); a country which performed well in both absolute inequality in mortality and in life expectancy (Mackenbach, Kunst et al. 1997; Mackenbach, Bakker et al. 2002).

Equally, the choice between health and ill-health also considerably affects the time trends in health inequality. This becomes clear when writing the health concentration index as a ratio:

$$C^h = 2\sigma_r^2 \frac{SII}{\bar{h}} \qquad (1.1)$$

and the ill-health concentration index as:

$$C^{ih} = 2\sigma_r^2 \frac{SII}{(\tilde{h} - \bar{h})} \qquad (1.2)$$

(see chapter 2) where $\bar{h}$ indicates the mean level of health, $\tilde{h}$ the (arbitrary) standard level of health, $\tilde{h} - \bar{h}$ the mean level of ill-health and SII the slope index of inequality (the estimated difference between the health of the best off and the worst off person). The probability limit for $2\sigma_r^2$ equals the constant 1/6 and may be ignored for now.

The concentration index for health will not change if both nominator and denominator increase or decrease by the same percentage. An increase in health that is accompanied by a smaller increase in the slope index of inequality leads to a decrease in the

Table 1.1 Ratios of health versus ill-health

	A		B	
	Health	Ill-health	Health	Ill-health
1	80	20	97	3
2	90	10	99	1
Ratio	0.89	2	0.98	3

concentration index. If, however, one measures inequality in ill-health, a reduction in ill-health will have to be matched by an equal percentage reduction in the slope index of inequality in order to hold the concentration index constant. Therefore, when inequality in ill-health is more likely to rise over time than inequality in health, this may explain the consistent increase in inequality found by Wagstaff (2002), but not the consistent increase found by Gravelle and Sutton (2003).

Both health and ill-health ratios compare the obtained level of health to a certain standard. In the example of country A and B, it is zero in the case of health and a hundred in the case of ill-health. Ratio measures, however, require a ratio scale and a ratio scale requires a natural zero point, one that may not depend on the data (see chapter 8). While the standard for health is natural, the standard for the ill-health is open for debate. It may vary from country to country and within countries over time, effectively ruling out ill-health based measures.

A choice for a health based measure in line with the measurement of inequities in health care that are traditionally measured in terms of health care contacts or resources used, rather than in terms of lack of contacts or lack of resources used)Wagstaff, Van Doorslaer et al, 1991). Like in other areas in which egalitarian policies are evaluated, such as income inequality, it is natural to focus on resources rather than the lack of resources compared to some arbitrary standard.

The above reasoning equally applies when the measure of health is not quality-adjusted life expectancy but some proxy for it, such as health-related utility. However, when the focus of research shifts from an ethical perspective to disease aetiology or other modelling of causation, the above considerations need not apply as long as one avoids ethical interpretations of the extent of the inequality.

We argue for a relative measure of inequality based on health rather than ill-health, and an application may be found in chapter 4.

1.2.5 Choice of index

Several articles compare the properties of the available indices. Frequently referred to articles have been written by economists (Wagstaff, Paci et al. 1991) and social epidemiologists (Mackenbach and Kunst 1997). Both arrive at complex measures that are argued to capture SES related differences in health best but have the drawback of easily leading to misinterpretation (Mackenbach and Kunst 1997). While the earlier relative index of inequality is mathematically similar to the concentration index, the relative index of inequality as defined by Mackenbach and Kunst (1997) is not (chapter 2). The latter may be easier to interpret but has different properties. To help interpret the extent of inequality as measured by a concentration index, we developed a redistribution interpretation for the concentration index (chapter 2). Both

the concentration index and the relative index of inequality require a health measure that is either measured on a dichotomous or a ratio scale (chapter 8).

1.2.6 Health status measurement

From the above it follows that a health measure should measure health rather than ill-health. Lacking information about quality adjusted life expectancy, we have chosen a cross sectional measure of health measured by self-reported health. Variation in this subjective measure may not accurately capture variation in true health, but on the other hand variation in true health may affect health-related utility differently for different individuals. The subjectivity may well capture those health limitations that matter to the individual (Simon 2002; Simon, De Boer et al. 2005). As such, true health is sensitive to different circumstances and coping capabilities among individuals, which may well be systematically related to socio-economic class. From an equity perspective, it is unclear to what extent health measures should be sensitive to these factors.

1.2.7 Applying ratio scale properties to an ordinal health measure

The measure of health above is measured on an ordinal scale. In order to use inequality measures such as the concentration index or the relative index of inequality, one needs to dichotomize the health measure or a ratio scale variable. Dichotomising multiple categories into healthy and non-healthy always leads to a loss of efficiency, and the choice of a cut-off point may well affect the measured inequality (see chapter 3). Alternatively one could convert the ordinal scale into a ratio scale. However, converting a lower quality variable into a higher quality variable is usually viewed as impossible (Sen 2001) as it requires adding information to the variable that was usually lacking in the first place.

Wagstaff and Van Doorslaer (1994) first upgraded an ordinal scale health variable by assuming a latent ill-health variable that had a log-normal distribution from which the ordered responses were drawn. In international comparative research (Van Doorslaer, Wagstaff et al. 1997) they assumed a separate log-normal distribution to underlie the observed ordered health for each of the countries, implying that while responses to the self assessed health questions differed between countries, the underlying mean level of health and the distribution of health were identical across countries.

To relax the assumption above, Van Doorslaer and Jones (2003) mapped the self assessed health responses on the McMaster Health Utility Index Mark III (Feeny, Furlong et al. 1995; Torrance, Feeny et al. 1996; Horsman, Furlong et al. 2003). Rather than using mid-point values as we did earlier (Van Doorslaer and Koolman 2000), they used cut-point values in combination with grouped data or interval regression. This approach had several advantages (Van Doorslaer and Jones 2003), but also one disadvantage in

that the Health Utility Index had no natural zero point. In fact, recent Canadian surveys include negative Health Utility Index scores (health states worse than death), and thus do not meet the requirements of a ratio-scale variable. These negative scores did not arise in the first wave (in 1994-1995) of the Canadian National Population Health Survey used for the study Van Doorslaer and Jones (2003) and we applied their mapping approach and treated the Health Utility Index as if it had ratio scale properties.

1.2.8 Factors that may cause unavoidable variation

Whitehead mentions biological variation as an example of an unavoidable cause of variation, and therefore as equitable. Most empirical researchers seem to follow this line of reasoning (Kakwani, Wagstaff et al. 1997; Gravelle 2003). Age- and sex-related variations in health were deemed examples of such unavoidable health, as both could not be caused by SES and thus could not mediate the effect of SES on health. This is obviously to some extent contentious, as actual observed health variation by age and sex may partly reflect differential treatment of men and women and of young and old by certain societies. Nonetheless, when comparing health inequalities between groups, researchers generally choose to standardise for differences in the age and sex distribution. As gene information gradually becomes more available one would expect that the current set of standardising factors would be extended.

1.2.9 Adjustment

Adjusting for age and sex was performed using direct and indirect standardisation techniques (Kakwani, Wagstaff et al. 1997; Gravelle 2003), a choice of terminology that does not seem grounded in either epidemiology or economics. Indirect standardisation was used to perform what some refer to as regression-based standardisation (Szklo and Nieto 2000) and others refer to as regression-based stratification (Hernan, Hernandez-Diaz et al. 2004). The introduction of regression-based adjustment methods followed the upsurge in popularity in regression-based adjustment in epidemiology that started about 15 years ago. This was mostly a result from the developments in computer technology and modern software packages that made such computations convenient, which overcame the drawback of the required burdensome computations.

Another possibly more helpful distinction is one between non-parametric adjustment, such as direct adjustment, matching or stratification versus parametric adjustment, such as regression-ased stratification. Non-parametric adjustment techniques share the attraction of not having to make regression model assumptions. On the downside these approaches are often cell-based and quickly become inefficient or even impossible as the number of factors (the dimensions) to control for increases, or when the standardising factors are measured as continuous variables (the curse of dimensionality). On the

flipside, parametric techniques require assumptions such as the strong assumption that the estimation of the effect of the standardising factors has to be the true causal effect.

1.2.10 Decomposing the concentration index

Two new decompositions were made available before the first paper of this thesis was written. Wagstaff, Van Doorslaer et al. (2003) building on work by Rao (1969) introduced a decomposition of the health concentration index into factors. This decomposition is very powerful both when the factors are causally related and merely associated with the outcome of interest. When the model is associational then the decomposition still allows one to compute the contributions of particular groups, thus allowing policy makers and scientist alike to focus their efforts on specific groups. Chapter 4 presents a first application of this decomposition to health inequalities in a comparative study of European countries, allowing a better understanding of the differences in health inequality.

1.3 HEALTH CARE INEQUITY

1.3.1 From ethics to empirical ethics

Interest in the distribution of health care is often derived from an interest in the inequitable distribution of health (Murray and Frenk 2000; Wagstaff and Van Doorslaer 2000). Health care use could both be a cause of health inequity and it may serve to reduce health inequity caused by other factors. This is, perhaps contrary to expectation, not how inequity in health care is typically defined. Policy documents typically state a preference for equality of access, i.e. access should depend on need and not on ability to pay (Wagstaff and Van Doorslaer 1993).[b]

Most studies aim to measure violations of the horizontal version of "treatment according to need" rather than access according to need. This is because equality in access is hard to measure and because when access and treatment are systematically unequal, e.g. poorly educated individuals systematically opt out of the use of effective treatment, then that decision will often be dubious and may in itself be a result of an inequitable process. On the other hand, the treatment according to need would fail as an equity principle when systematic variation is due to individuals exercising their right to refuse needed health care if they so wish, for example for religious or ethical reasons. This is the reason why the WHO has advised against the use of this principle for equity in health care (Whitehead 1985). See Culyer and Wagstaff (1993) and Wagstaff and Van Doorslaer (2000) for a discussion.

Both equality in access and in treatment according to need require a measure of need. The concept of need depends crucially upon the chosen ethical perspective. If the interest lies in equity in health, then marginal need could be defined as the marginal contribution to health per monetary unit spent on health care. Total need may then be defined as the minimum amount of resources required to exhaust capacity to improve health adapted from Culyer and Wagstaff (1993). If the interest lies in equality of capabilities, and ones capabilities depend on ones functionings which in turn depend on health see (Sen 1992), then need could be defined as the minimum amount of resources required to exhaust capacity to improve functionings. As health improvements might differ in the way they affect functionings, both definitions might lead to different levels of need. Most work in this field has taken the equity in health perspective.

In practice measuring a concept of need as described above is complex as it involves the estimation of an attainable health gain for an extra monetary unit spent on each individual, which is often not feasible with the available data. Many researchers therefore revert to defining need as the predicted use based on an individuals health characteristics. This approach requires that (1) ill-health is related to capacity to benefit from health care in the same way across socio-economic classes and (2) costs per health improvement do not systematically vary between socio-economic classes. Both assumptions may be rather strong as the distribution of different diseases varies across socio-economic classes (Shaw, Dorling et al. 1999) and both cost and effectiveness of treatment vary with these diseases. Furthermore, disease outcomes vary systematically with socio-economic class (Shaw, Dorling et al. 1999), suggesting that the production of health with the same input results in differential outcomes and equal outcomes would require different inputs.

Even if access and received treatment were equal, the limitations of health care delivery may imply that it cannot entirely remove systematic health inequity; it may merely reduce it.[c] The more severe the limitations of health care, the less equal treatment will reduce the variation in health. The WHO approach could induce unequal health care such that those (groups) with lower health receive more health care than they would receive based on need alone (Murray and Frenk 2000). This may seem inequitable from an equal access viewpoint, but it would do more justice to the reason why the distribution of health care is studied at all.

In practice at this time, studying the causal contribution of national health care to the reduction of health inequities does seems rather challenging as, in most studies even at individual level, health care does not seem to increase health very much, if at all (e.g., see Newhouse and Insurance Experiment Group, 1993). Most of these studies evaluate the marginal health effects of health care. Differences between health states in developed countries are likely to lie at that margin and thus may not explain many

of the differences in health inequities. More is to be expected of studies at the micro-level which study the distributional impact of an intervention. Such impact could be weighted together with the cost-effectiveness ratio and other ethical considerations when considering the coverage of the intervention.

This approach of measuring need also implies that the system gets it right on average in its vertical (between those of unequal need) distribution of resources, even though this vertical standard may vary from country to country and within countries over the years. This raises questions with respect to vertical equity, but these are not addressed here. While many of the issues touched upon above require further study, for this thesis we have followed the convention of measuring violations of treatment according to need and coin them inequitable.

We were able however to get a better understanding of which group or factor contributes what amount to the observed health care inequity by developing a novel method to apply the factor decomposition when the underlying model is non-linear in its coefficients. This method is then applied to decompose horizontal inequities in health care as measured with the concentration index. Milanovic (1997) also proposed as decomposition that could be applied to the concentration index, which is introduced to the health (care) inequality literature in chapter 2. Chapter 6 also decomposes the concentration index into at least one visit and subsequent visits to physicians, following the already popular use of two part models to model health care use in health economics.

1.4 ORGANISATION OF THE THESIS

Health inequality and health care inequity measurement are undergoing permanent evolution. Many issues have been solved in the past, many are currently addressed and possibly even more are still to be raised. With this thesis we hope to contribute to the advancement of the measurement and the evidence base of both socioeconomic health inequality and socioeconomic health care inequity.

Chapter 2 responds to the complex interpretation of complex measures of inequality by presenting a novel way to interpret the concentration index in terms of the redistribution percentage that is required to obtain an equal distribution of health or health care. It also shows that the concentration index is a multiplication of a correlation coefficient between the rank of the ranking variable and the outcome variable and the coefficient of variation of the outcome variable, and that the relative index of inequality, as redefined by Mackenbach and Kunst (1997), is not mathematically similar to the concentration index, and thus has other properties.

The extent of the inequality may be easier to interpret if it is expressed in the relative index of inequality, as defined by Mackenbach and Kunst. However, when the outcome variable is categorical, it is often based on an odds ratio, making the interpretation much

more complex. Chapter 3 facilitates the interpretation of the relative index of inequality by introducing a novel method to compute the relative index of inequality based on prevalence rate ratios rather than odds-ratios.

More evidence of health inequality as measured by the concentration index is presented in chapter 4. These results are based upon the European Community Household Panel that was set up to provide more comparable data. To better understand what factors are associated with this health inequality, we provide an application of the decomposition approach introduced by Wagstaff, Van Doorslaer et al. (2003) in income related inequalities in health. We introduce bootstrap techniques to construct confidence intervals that incorporate more sources of bias. To find out to what extent greater health inequality is due to differences in health elasticities of these factors or to variation in the income distribution of these factors, we subsequently compare all countries with the country with the smallest health inequality, using the Oaxaca (1973) decomposition. For chapter 4 we use the third wave of a longitudinal data set. These panel data grow less representative due to (health-related) possibly drop out or attrition.

In chapter 5 we study whether drop out is health-related and whether the association between income and health is sensitive to correction for drop out using inverse probability weighting techniques. Some of the income related inequalities observed may result from income related inequity in health care use, which is why many EU member states have stated that health care ought to be distributed according to need, and not to income (Wagstaff and Van Doorslaer 1993). Using the data from the European Community Household Panel that is more comparable than data used before (Van Doorslaer, Wagstaff et al. 2000), we test whether health care is indeed distributed according to need, or if that income does play a role. To better understand the origins of inequity in health care we also compute concentration indices and apply a method to decompose GP and specialist visits inequalities in factors when the underlying model is non-linear in the coefficients. We also study both the probability of a visit, the number of visits for those who had at least one visit, and the total number of visits, as both may follow from different decision processes.

Private health care insurance may well improve access to health care and is likely to be concentrated among the better off in terms of income. The latter is confirmed by Van Doorslaer, Koolman et al. (2002) and in chapter 6. To carefully study the effect of voluntary health insurance we aim to go beyond the conditional distribution and present the causal effect of voluntary health insurance on health care consumption using an instrumental variable approach in chapter 7. This thesis is concluded with a discussion of both methods and findings and a summary.

Notes

a. Even though the unit of measurement may be the individual, which has lead to some confusion Wolfson, M. and G. Rowe (2001). On measuring inequalities in health. Bulletin of the World Health Organization 79(6): 553-560.
b. Revealed preferences often fail to correspond with stated preferences for equality in access, for example when governments allow for preferential treatment which is only accessible through substantial (co)payments or private insurance.
c. It could even increase it, for example because health production may be more efficient for the higher SES groups.

Ackerknecht, EH (1953). Virchow, R: doctor, statesman, anthropologist. Madison, University of Wisconsin Press.

Anand, P and P Dolan (2005). Equity, capabilities and health. Social Science and Medicine 60(2): 219-222.

Asada, Y (2005). A framework for measuring health inequity. Journal of Epidemiology and Community Health 59: 700-705.

Braveman, P and S Gruskin (2003). Defining equity in health. Journal of Epidemiology and Community Health 57(4): 254-258.

Braveman, P, N Krieger, et al. (2000). Health inequalities and social inequalities in health. Bulletin of the World Health Organization 78(2): 232-234.

Brouwer, WBF, NJA Van Exel, et al. (2005). Acceptability of less than perfect health states. Social Science and Medicine 60: 237-246.

Bryan, S and P Dolan (2004). Discrete choice experiments in health economics. European Journal of Health Economics 5: 199-202.

Carneiro, P, KT Hansen, et al. (2001). Removing the veil of ignorance. Swedish Economic Policy Review 8: 273-301.

Clarke, PM, UG Gerdtham, et al. (2002). On the measurement of relative and absolute income-related health inequality. Social Science and Medicine 55: 1923-1928.

Coleman, W (1982). Public health and political economy in early industrial France. Madison, University of Wisconsin Press.

Culyer, A J and A Wagstaff (1993). Equity and Equality in health and health care. Journal of Health Economics 12(4): 431-57.

Dworkin, R (1977). Taking rights seriously. Cambridge, Harvard University Press.

Farr, W (1839). Letter to the Registrar General. First annual report of the Registrar General. London, HMSO.

Feeny, DH, WJ Furlong, et al. (1995). Multi-attribute health status classification systems: Health Utilities Index. Pharmacoeconomics 7(6): 490-502.

Gakidou, EE, CJL Murray, et al. (2003). Quality and Equity: Preferences for Health System Outcomes. Health Systems Performance Assessment. CJL Murray and D B Evans. Geneva, World Health Organization.

Gakidou, EE, CJL Murray, et al. (2000). Defining and measuring health inequality: an approach based on the distribution of health expectancy. Bulletin of the World Health Organization 78(1): 42-54.

Gravelle, H (2003). Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics 12(10): 803-819.

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

Heckman, J (2001). Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture. Journal of Political Economy 109(4): 673-748.

Hernan, M, S Hernandez-Diaz, et al. (2004). A structural approach to selection bias. Epidemiology 15: 615-625.

Horsman, J, W Furlong, et al. (2003). The Health Utility Index (HUI): concepts, measurement properties and applications. Health and Quality of Life Outcomes 1(54).

Kakwani, N, A Wagstaff, et al. (1997). Socioeconomic inequalities in health: Measurement, computation and statistical inference. Journal of Econometrics 77: 87-103.

Mackenbach, J, A Cavelaars, et al. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe - Reply. Lancet 350(9076): 517-518.

Mackenbach, J and A Kunst (1997). Measuring the magnitude of socio-economic inequalities in health: An overview of available measures illustrated with two examples from Europe. Social Science and Medicine 44(6): 757-771.

Mackenbach, J, A Kunst, et al. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. Lancet 349(9066): 1655-1659.

Mackenbach, JP, MJ Bakker, et al. (2002). Socioeconomic inequalities in health in Europe; An overview. Reducing inequalities in health; A European perspective. JP Mackenbach and MJ Bakker. London, Routledge.

Milanovic, B (1997). A simple way to calculate the Gini coefficient, and some implications. Economics Letters 56: 45-49.

Murray, CJL and J Frenk (2000). A framework for assessing the performance of health systems. Bulletin of the World Health Organization 78(6): 717-731.

Murray, CJL, EE Gakidou, et al. (1999). Health inequalities and social group differences: what should we measure? Bulletin of the World Health Organization 77: 537-543.

Newhouse, JP and Insurance Experiment Group (1993). Free for all? Lessons from the RAND Health Insurance Experiment. Cambridge, Harvard University Press.

Oaxaca, RL (1973). Male-female wage differentials in urban labor markets. International Economic Review 14: 693-709.

Rao, VM (1969). Two decompositions of concentration ratio. Journal of Royal Statistical Society Series A 132(3): 418-425.

Sen, AK (1992). Inequality reexamined. New York.

Sen, AK (2001). The fundamental challenges of measurement and perceptions in health equity. Challenging inequities in health; From ethics to action. T Evans, M Whitehead, F Diderichsen, A Bhuiya and M Wirth. Oxford, Oxford University Press.

Shaw, M, D Dorling, et al. (1999). The Widening Gap: Health Inequalities and Policy in Britain (Studies in Poverty, Inequality & Social Exclusion). Bristol, The Policy Press.

Simon, J, J De Boer, et al. (2005). How is your health in general? A qualitative study on self-assessed health. European Journal of Public Health 15(2): 200-208.

Simon, JG (2002). How is your health in general? Department of Public Health. Rotterdam, Erasmus University Rotterdam.

Szklo, M and FJ Nieto (2000). Epidemiology: Beyond the basics. Gaithersburg, Aspen.

Torrance, GW, DH Feeny, et al. (1996). Multiattribute Utility Function for a Comprehensive Health Status Classification System. Medical Care 34(7): 702-722.

Townsend, P and N Davidson (1982). The Black Report. Harmondsworth, Penguin.

Van Doorslaer, E and AM Jones (2003). Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics 22(1): 61-87.

Van Doorslaer, E and X Koolman (2000). Income-related inequalities in health in Europe: evidence from the European Community Household Panel. ECuity Project II Working Paper 1. Erasmus University, Rotterdam, ECuity Project II Working Paper.

Van Doorslaer, E, X Koolman, et al. (2002). Equity in the use of physician visits in OECD countries: Has equal treatment for equal need been achieved? Measuring up: Improving health systems performance in OECD countries. Paris, OECD: 225-248.

Van Doorslaer, E, A Wagstaff, et al. (1997). Income-related inequalities in health: Some international comparisons. Journal of Health Economics 16(1): 93-112.

Van Doorslaer, E, A Wagstaff, et al. (2000). Equity in the delivery of health care in Europe and the US. Journal of Health Economics 19(5): 553-583.

Wagstaff, A (2002). Inequalities in health in developing countries: swimming against the tide? Washington, D.C.

Wagstaff, A, P Paci, et al. (1991). On the measurement of inequalities in health. Social Science and Medicine 33(5): 545-557.

Wagstaff, A and E Van Doorslaer (1993). Equity in the finance and delivery of health care: Concepts and definitions. Equity in the finance and delivery of health care: An international perspective. E Van Doorslaer, E Wagstaff and F Rutten. Oxford, Oxford University Press.

Wagstaff, A and E Van Doorslaer (1994). Measuring inequalities in health in the presence of multiple category morbidity indicators. Health Economics 3(4): 281-291.

Wagstaff, A and E Van Doorslaer (2000). Equity in health care financing and delivery. Handbook of Health Economics. A J Culyer and J P Newhouse, North Holland. 2: 1803-1862.

Wagstaff, A, E Van Doorslaer, et al. (1991). Horizontal equity in the delivery of health care: Reply. Journal of Health Economics 10(2): 251-256.

Wagstaff, A, E Van Doorslaer, et al. (1991). On the measurement of horizontal inequity in the delivery of health care. Journal of Health Economics 10: 169-205.

Wagstaff, A, E Van Doorslaer, et al. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics 112: 207-223.

Whitehead, M (1985). The concepts and principles of equity and health. Copenhagen, World Health Organisation Regional Office for Europe: 18.

Wolfson, M and G Rowe (2001). On measuring inequalities in health. Bulletin of the World Health Organization 79(6): 553-560.

CHAPTER 2

Interpretation of a concentration index

ABSTRACT

This paper aims to add a more intuitive understanding to the concept of a concentration index for measuring relative inequality with an application of health-related measures by income. An existing and a new redistribution interpretation of the Gini are presented and applied to the concentration index. Both indicate the share of the total amount of any variable that needs redistributing in a particular way from rich to poor (or vice versa) to achieve equality. The characteristics of these redistribution schemes are compared. The paper also draws attention to the relationship between a concentration index, a correlation coefficient with relative income rank and a coefficient of variation of the variable of interest. These relationships are illustrated using data on inequality in dental care utilisation in European countries taken from the European Community Household Panel survey.

2.1 INTRODUCTION

Concentration indices (CIs) and curves have now become fairly standard measurement tools in the health economics literature on equity and inequality in health and health care (Wagstaff and Van Doorslaer, 2000). They were first introduced by Wagstaff, Van Doorslaer and Paci (1989) and have since been used successfully to describe and measure the degree of inequality in various measures of health (e.g. Van Doorslaer et al., 1997), health care utilisation (e.g. Van Doorslaer et al., 2000) or in health care payments (e.g. Wagstaff et al., 1999). Wagstaff, Paci and Van Doorslaer (1991) have reviewed and compared the properties of the CI with alternative measures of health inequality and concluded that it shares the same properties as one of the two relative index of inequality measures that are used by epidemiologists but that concentration curves have an additional advantage in terms of their visual representation of the location of deviations from proportionality and the possibility to perform checks of dominance relationships (Van Doorslaer et al., 2000; Wagstaff, 2000).

2.2 REDISTRIBUTION SCHEMES AND THE GINI

While the Gini coefficient has proven the most popular measure of income inequality for decades, it lacks a straightforward intuitive interpretation. For this reason, other measures of inequality are sometimes presented alongside or instead of the Gini. One popular alternative is the Schutz-coefficient (Schutz, 1951; Lambert, 1993), also referred to as Pietra-ratio (Kondor, 1971), maximum equalisation percentage (Kondor, 1975), or Robin Hood Index (Kennedy et al., 1996) in the income literature. The Schutz-coefficient can be interpreted as the percentage of total income that needs transferring from the group with above average incomes to the group of individuals with below average incomes in order to reach equality. It is defined as half the relative mean deviation (Kondor, 1971):

$$S = \frac{1}{2\bar{y}n} \sum_{i=1}^{n} |y_i - \bar{y}| \qquad (2.1)$$

While easy to interpret, the Schutz-coefficient bears no unique relationship with the Gini, and may therefore lead to a different ranking when comparing groups. It also fails to meet the principle of transfers, i.e. that every transfer from higher to lower incomes results in a lower index of income inequality.

Blackburn (1989) has proposed an alternative (hypothetical) redistribution scheme that does have a unique relationship with the Gini, and can therefore serve to interpret the magnitude of a Gini. The redistribution requires a reallocation of a fixed lump sum amount from all units above the median income to those below the median income.

The fixed amount k is defined as:

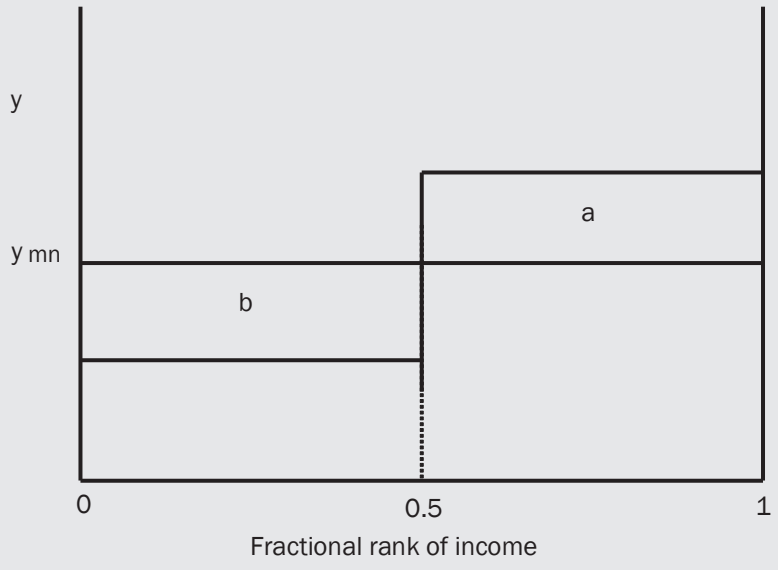
$$k=200\Delta G \tag{2.2}$$

where k is the percentage of the mean level of the variable of interest and ΔG is the required change in the Gini coefficient. As this amount will be taken from half the units, the total share that needs to be redistributed in order to obtain total equality is

$$R_b=100G \tag{2.3}$$

where R_b is the percentage that needs redistributing. The Blackburn redistribution is depicted by the reallocation of area a to area b in figure 2.1.

Figure 2.1 The Blackburn redistribution



Because the redistribution is a lump sum and is not related to the actual distribution of income, it does not remove all income inequality. The Blackburn redistribution only reduces the Gini to zero if one holds the income ranks constant to their pre-redistribution position. Although the Gini based on the post-redistribution income rank will be lower, it may not even be close to zero.

2.3 REDISTRIBUTION INTERPRETATIONS OF THE CONCENTRATION INDEX

The CI is derived from the Gini but differs, as the ranking variable and the variable of interest (for which the inequality is evaluated) are different. Hence, the CI is a bivariate

measure of inequality, measuring inequality in one variable related to the ranking of another. Like the Gini, the CI has the disadvantage of lacking a straightforward interpretation in natural units. As the CI is a bivariate measure, a redistribution of the variable of interest need not affect the ranking based on the other variable. Therefore, the critique that Blackburn's redistribution only holds if the ranking is preserved, is in the case of the CI of little importance.

2.3.1 A linear redistribution scheme

However, other possible redistribution schemes with different properties are conceivable. We propose a new redistribution scheme that makes use of some properties of the so-called 'convenient regression' that is often used to compute a CI (Kakwani et al., 1997). The convenient OLS regression is given by

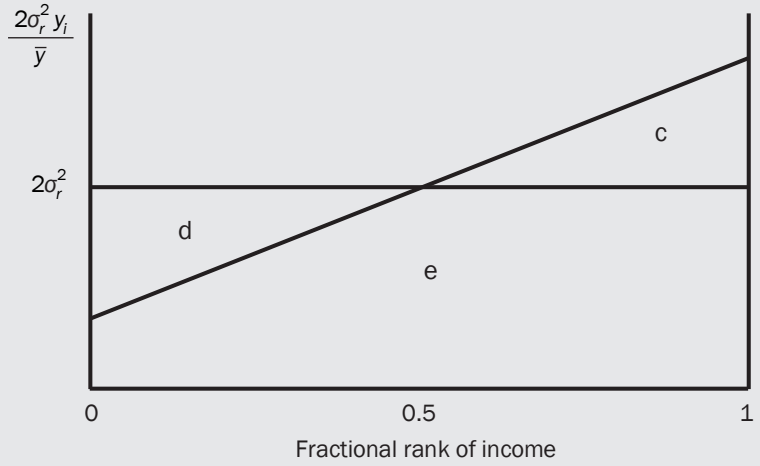
$$\frac{2\sigma_r^2}{\bar{y}} y_i = \alpha_1 + \beta_1 r_i + \varepsilon_{1,i} \tag{2.4}$$

where σ_r^2 is the variance of r, $\bar{y} = 1/n \sum_{i=1}^{n} y_i$, r is the fractional rank of income and $\hat{\beta}_1$ is an estimate of the CI.[a] This regression is a convenient device for computing a CI that does not require additional assumptions. This can be shown using the properties of the summation operator and the nth partial sum of r_i, which is equal to n/2 (n/1 + n/n):

$$\begin{aligned}
\beta_1 &= \frac{2}{n\bar{y}} \sum_{i=1}^{n} \left(y_i r_i - \bar{y} r_i - \bar{r} y_i + \bar{r} \cdot \bar{y} \right) \\
&= \frac{2}{n\bar{y}} \left[\sum_{i=1}^{n} y_i r_i - \bar{y} \sum_{i=1}^{n} r_i - \bar{y} n \bar{r} + \bar{y} n \bar{r} \right] \\
&= \frac{2}{n\bar{y}} \sum_{i=1}^{n} y_i r_i - \frac{2}{n\bar{y}} \bar{y} \left(\frac{1}{n} + \frac{n}{n} \right) \frac{1}{2} n \\
&= \frac{2}{n\bar{y}} \sum_{i=1}^{n} y_i r_i - 1 - \frac{1}{n} = C
\end{aligned} \tag{2.5}$$

From this result we can now derive a redistribution of the transformed y variable on the left hand side of equation 2.4 that reduces $\hat{\beta}_1$ to zero. Figure 2.2 shows the rank-predicted linear distribution of the transformed y variable which, by definition, passes through the means point $\left(0.5; 2\sigma_r^2\right)$. As a result, the sum of all deviations from the mean on the right half of this distribution (area *c*) exactly equals the sum of all deviations from the mean on the left hand side of the distribution (area *d*). Equality requires *c* (and *d*) to be zero. The area *c*/(*d* + *e*) is the estimated proportion of the (transformed) *y* variable that needs redistributing from the richest to the poorest half, and we will refer to it as the linear redistribution scheme (R_l). Note that area *c* is always equal to 1/8 times the area

Figure 2.2 The linear redistribution



$\hat{\beta}_1 (r_n - r_1)$. For large samples, the latter area is equal to the slope coefficient $\hat{\beta}_1$ because $p\lim(r_n - r_1) = 1$. Therefore, area *c* equals $\frac{1}{8}\hat{\beta}_1$. Using this result and equation 2.4, we now have

$$R_I = \frac{\frac{100}{8}\beta_1}{2\sigma_r^2} = \frac{100}{16\sigma_r^2}C \qquad (2.6)$$

Since $|c| < |e|$, R_I varies between -50% and 50% of (transformed) *y*.

The transformed *y* variable on the left hand side of equation 2.4 is, however, not a natural scale and is therefore still difficult to interpret. Therefore, we use $p\lim \sigma_r^2 = \frac{1}{12}$ to reduce the left hand side to $y_i/(6\bar{y})$. To find the percentage of all y that needs to be redistributed, we multiply area c by 6 and arrive at

$$R_I = \frac{300}{4}\beta_1 = 75C \qquad (2.7)$$

which holds approximately for large samples. An additional advantage of this result is that RI has an interpretation quite similar to the Schutz-coefficient and is directly related to the CI. In the appendix we show that the two are identical in this case as a result of the fixed shape of the predicted concentration curve when based on equation 2.4.

2.3.2 Lump sum and linear redistribution compared

When comparing Blackburn's constant amount redistribution and the linear redistribution as a function of relative rank, we see that the latter redistribution scheme is more 'effective' in the sense that the total amount of *y* that needs redistributing is smaller. This follows from the principle of transfers' property of the CI, i.e. that every transfer of *y*

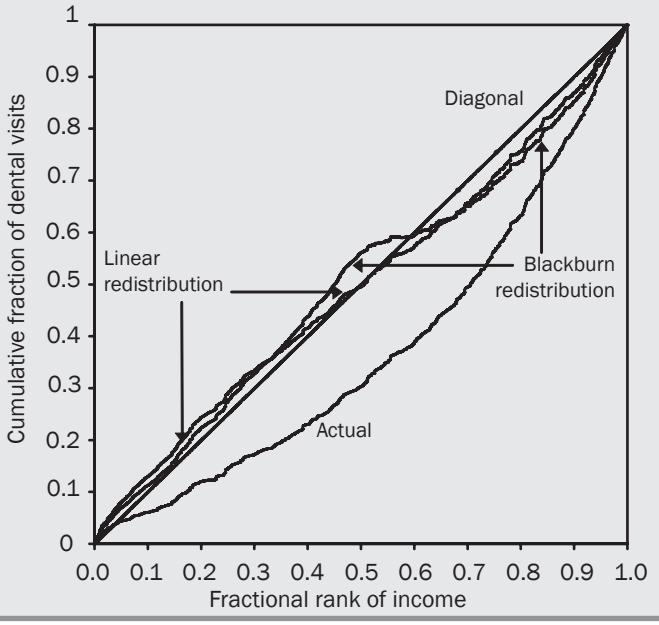
from higher to lower income households leads to a CI closer to zero if the CI is positive, and vice versa if the CI is negative. Since the linear redistribution transfers more from and to the extreme income households, it is therefore more effective. The principle of transfers' property can be shown using the equation for $\hat{\beta}_1$

$$\beta = \frac{\sum_{i=1}^{n} (r_i - \bar{r})(y_i - \bar{y})}{\sum_{i=1}^{n} (r_i - \bar{r})^2} \tag{2.8}$$

Redistributions will only affect the numerator and more so to the extent that deviations before and after affect the units that are further away from the median income unit. The most effective redistribution will affect only the richest and the poorest units and leave the rest of the distribution untouched. While this may help to obtain a zero CI, the resulting concentration curve will not match the diagonal, and income-related inequalities in parts of the income distribution will remain. Therefore, we introduce an 'appropriateness' characteristic (A) to compare redistribution schemes: the mean squared concentration curve deviations from perfect equality, where Lp denotes the post redistribution concentration curve

$$A = \frac{1}{n} \sum_{i=1}^{n} \left(L_i^p - r_i \right)^2 \tag{2.9}$$

Figure 2.3 Concentration curves of dental visits in Portgal before and after redistribution



This appropriateness measure does not allow areas above and below the diagonal to compensate each other and places greater weight on deviations further from the diagonal.

Figure 2.3 shows the distribution of *y* before and after the Blackburn and the linear redistribution for dental care visits in Portugal. The most appropriate redistribution scheme will redistribute from all those with a 'surplus' of dental care contacts to individuals with a 'shortage' of contacts (in relation to their income rank). It would remove all income-related inequality throughout the income distribution and equate the concentration curve with the diagonal, but the unique relationship with the CI would be lost.

From figure 2.3 it can be seen that the Blackburn redistribution is less appropriate than the linear redistribution as it shows large deviations from the diagonal because the units just left of the median are treated very differently from the ones just right of the median. For Portugal this results in very low post redistribution consumption of dental care for the households just above the median income. Appropriateness of both redistribution schemes can be compared using the ratio of A_{lin} divided by A_b. Table 2.1 shows that the linear redistribution is the more appropriate redistribution in eleven of the thirteen countries.

Three points are worth making. One, we now have two intuitive interpretations of a CI in more natural units. An index value of 0.10 means that a lump sum redistribution of 10% of the total amount of *y* (which can be health in utility units, health care use in doctor

Table 2.1 Inequality in annual number of dentist visits in 13 EU member countries (ECHP, 1996)

Country	$\bar{y}$	$\sigma_y / \bar{y}$	$\rho(y, r)$	σ_r^2	CI	t^*	R_l	A_{lin}/A_b
Germany	1.97	1.37	0.020	0.083	0.016	1.73	1.2%	0.74
Denmark	1.75	1.22	0.080	0.084	0.056	5.21	4.2%	0.73
Netherlands	1.63	1.03	0.099	0.080	0.056	7.94	4.4%	2.21
Belgium	1.34	1.80	0.057	0.081	0.057	4.02	4.4%	2.42
Luxembourg	1.73	1.96	0.019	0.083	0.022	1.14	1.6%	0.93
United Kingdom	1.45	1.33	0.096	0.081	0.072	6.24	5.5%	0.58
Ireland	0.66	2.20	0.171	0.079	0.205	11.91	16.2%	0.49
Italy	1.11	2.54	0.094	0.085	0.140	10.37	10.3%	0.36
Greece	0.76	2.69	0.078	0.084	0.121	7.58	9.0%	0.44
Spain	0.82	2.90	0.078	0.083	0.131	8.15	9.8%	0.24
Portugal	0.84	2.79	0.167	0.080	0.259	10.45	20.1%	0.67
Austria	1.91	1.41	0.019	0.084	0.015	1.23	1.1%	0.73
Finland	1.43	1.68	0.094	0.084	0.091	6.18	6.8%	0.54

* t-statistics were estimated using Newey-West estimator of $\hat{\sigma}_C^2$ (Kakwani et al., 1997)

visits or health care payments in Euro's) is required from the richest half to the poorest half of the population in order to equalise the distribution. Using the linear redistribution scheme, the amount to be redistributed is only 7.5%. Neither redistribution scheme will make the distribution completely equal. The amount to be redistributed under the linear scheme equals $\hat{y}_i$-$\overline{y}$ for every individual, where $\hat{y}_i$ is the rank-predicted value from an OLS regression. Such redistribution removes all income-related inequality by equalising the rank-predicted distribution, not the actual distribution.

Secondly, the redistribution interpretation also makes clear that both of these indices have ratio scale properties if the underlying *y* variable has ratio scale properties and the r -related variation in y is linear. In other words, when an index value doubles, also the degree of inequality doubles, or twice as many units of *y* need redistributing. This is important if these indices are going to be used as observations in further statistical analysis or comparisons.

Finally, expression (2.3) easily lends itself to a generalisation in the case where one is only interested in the degree of inequity, i.e. in the inequality remaining after having standardised or adjusted for demographic differences (as in Van Doorslaer et al., 1997) or in the degree of inequality of health care use after having standardised for need differences as proxied by morbidity variables (as in Van Doorslaer et al., 2000). We did not attempt any standardisation for differences in dental care needs here.

2.4 INEQUALITY, VARIATION AND CORRELATION

The above suggests that a CI basically records the association between a variable and the relative rank in the income distribution. However, it does more than just that. This can most easily be illustrated using a slightly manipulated result derived by Milanovic (1997) between a Gini, a coefficient of variation of y and a correlation coefficient $\rho(y,r)$ between income and rank. Replacing the Gini by the CI, the equivalent relationship for the CI becomes

$$C = \frac{12\sigma_r^2}{\sqrt{3}} \frac{\sigma_y}{\overline{y}} \rho(y,r) \qquad (2.10)$$

Again, for a large sample the first component in equation 2.1 is a constant and the difference between $\hat{C}$ and $\hat{\rho}$ only depends on the second component, the coefficient of variation of *y*. This means that, even if the correlation of two variables with rank is the same, income-related inequality will be greater for the one with the greater variability. A CI, unlike a correlation coefficient, but like a covariance, takes into account both the strength of the association and the magnitude of the differences. However, unlike a covariance but like a correlation, it is normalised in the interval [-1,1].

The relevance of the magnitude of the differences can be illustrated with the distribution of dental visits across income groups in European countries. Table 2.1 presents the mean number of dentist visits, the coefficient of variation, the correlation coefficient with rank in the income distribution, the variance of the fractional rank, the CI, and its t-value (Newey and West, 1987) (which is the same for the CI and the percentage to be linearly redistributed). Blackburn's redistribution percentage is not reported as it is simply the CI times hundred. The results are estimated for thirteen EU member countries using the ECHP wave 3 sample-weighted data (see EUROSTAT, 1999).

The number of visits to the dentist has a positive correlation with income rank, and therefore all CIs of dentist visits are positive and significant. The CIs vary substantially between 0.015 for Austria and 0.259 for Portugal. The percentage to be linearly redistributed to achieve equality equally varies greatly between 1.1 percent for Austria and 20.1 percent for Portugal. In most countries, except Portugal, Ireland and Italy, equalising dental care utilisation across the income distribution requires redistributing less than 10% of all visits.

Comparing results across countries, we see that the correlation coefficient and the CI do not provide identical pictures. Depending on the variability relative to the mean dental utilisation, as given by the coefficient of variation, a similar correlation can be translated into a higher or lower degree of inequality. Take the comparison between Ireland and Portugal. The correlation between dentist visits and income rank is very similar ($\hat{\rho}$=0.17) but because the variation in Portugal is much higher than in Ireland, its inequality in dentist use is also much higher. For a similar reason, inequality is twice as high in Spain as in The Netherlands, despite the fact that income rank is more highly correlated with dental visits in The Netherlands. This result could mimic the differences in the coefficients of variation in dental health in both countries. In other words, two systems could result in identical correlation coefficients and respond identically to need, yet present quite different inequality indices as a result of differing coefficients of variation in health. This also holds for inequity, as standardisation does not level out differences in coefficients of variance.

2.5 CONCLUSION

Concentration indices have gained popularity in research aimed at measuring and explaining degrees of relative inequality in health-related variables of interest. It has been argued that their appeal to policy-oriented users of such research is limited by their alleged difficult interpretation. This paper draws attention to the relationship between a CI and two other - more intuitive - measures of redistribution and correlation.

The first useful result is the relationship between a CI and redistribution schemes which remove income-related inequality in the variable of interest. We propose a new

(linear) redistribution scheme which bears some resemblance but also offers certain advantages over Blackburn's lump sum redistribution. We show that it is more effective than the lump sum and more appropriate. We also show that there is a very simple and straightforward relationship with the Schutz coefficient. The percentage to be linearly redistributed is approximately equal to three quarters of the Blackburn redistribution. Using these redistribution schemes, a CI value can easily be translated into a percentage redistribution required from rich to poor to make estimated income-related inequality equal to zero, but not, however, to obtain equality!

The second result draws attention to the difference between a CI and a simple correlation with rank. CIs can also be interpreted as variation-weighted correlation coefficients: the strength of the association with rank is 'weighted' by the coefficient of variation in producing the measured degree of inequality. While the contribution of the correlation with rank accounts for most of the variation in CIs, the coefficient of variation still varies by almost a factor three. Therefore, it too plays an important role in explaining the differences in CIs between countries. The distinction may also have ethical implications, as societies may strive for a zero CI through zero correlation between income and health care use but not through few will prefer zero variation in health care use. Finally, policy makers wanting to lower an income-related inequality may now target both components separately.

Note

a. By rescaling the dependent variable, we can alternatively calculate the CI from $y_i = \alpha_2 + \beta_2 r_1 + \varepsilon_{2,i}$, where CI = $\beta_1 = 2\sigma_r^2 \beta_2 / \bar{y}$. The so-called slope index of inequality (SII) equals β_1, while both $\beta_2 / \bar{y}$ (Pamuk, (1985) and $(\alpha_2+\beta_2)/\alpha_2$ (Mackenbach and Kunst, 1997)) have been referred to as the relative index of inequality (RII) often used by epidemiologists.

Proof that for large samples and with linear redistribution of *y* with respect to *r* the percentage to be linearly redistributed index equals 3/4 of the *CI*.

Let relative *y* of individual *i* be determined by

$$I(r_i) = \frac{y_i}{\bar{y}} = \alpha_4 + \beta_4 r_i + \varepsilon_{4,i} \text{ for } r \in (0,1) \tag{2.11}$$

where $\beta_4 = \frac{\beta_2}{\bar{y}}$ (2.12)

Using OLS we obtain a rank-predicted estimate of relative *y*, equal to $\hat{I}(r_i)$. The predicted concentration curve $\hat{L}(r_i)$ is obtained by integrating $\hat{I}(r_i)$ over r_i.

$$\hat{L}(r_i) = \int_0^1 \hat{I}(r_i)dr = \int_0^1 \frac{\hat{y}_i}{\bar{y}} dr = \hat{\alpha}_4 r_i + \tfrac{1}{2}\hat{\beta}_4 r_i^2 + q \tag{2.13}$$

where *q* equals zero because $\hat{L}(r_i)$ equals zero at *r*=0.
The percentage to be linearly redistributed (R_l) is defined as the maximum difference between the diagonal $D(r_i)$ and $\hat{L}(r_i)$ if and only if $\frac{d^2\hat{L}(r_i)}{dr^2} \leq 0$ or $\frac{d^2\hat{L}(r_i)}{dr^2} \geq 0$ for $r \in (0,1)$. In that case R can be estimated as

$$\hat{R} = \max\left[D(r_i) - \hat{L}(r_i)\right] = \max\left[r_i - \hat{\alpha}_4 r_i - \tfrac{1}{2}\hat{\beta}_4 r_i^2\right] \tag{2.14}$$

The maximum difference can be found by setting the first derivative of equation 2.11 with respect to *r* equal to zero, or

$$\frac{dD}{dr} - \frac{d\hat{L}}{dr} = 1 - \frac{\hat{y}_i}{\bar{y}} = 1 - (\hat{\alpha}_4 + \hat{\beta}_4 r_i) = 0 \tag{2.15}$$

Obviously, equation 2.15 will be zero only at the mean of $\hat{y}_i$, where $\hat{I}(r_i)$ equals 1, and this value of a linear function will only be reached at the mean of r. Since $p\lim \bar{r} = \frac{1}{2}$, substituting $r = \frac{1}{2}$ in equation 2.14 gives

$$\hat{R} = D(\tfrac{1}{2}) - \hat{L}(\tfrac{1}{2}) = \tfrac{1}{2} - \tfrac{1}{2}\hat{\alpha}_4 - \tfrac{1}{8}\hat{\beta}_4 \tag{2.16}$$

Using

$$\hat{\alpha}_4 = E\left[I(r_i)\right] - \hat{\beta}_4\bar{r} = \sum_{i=1}^{n} \frac{y_i}{n\bar{y}} - \hat{\beta}_4\bar{r} = 1 - \hat{\beta}_4\bar{r} \tag{2.17}$$

we obtain

$$\hat{R} = \tfrac{1}{2} - \tfrac{1}{2} + \tfrac{1}{4}\hat{\beta}_4 - \tfrac{1}{8}\hat{\beta}_4 = \tfrac{1}{8}\hat{\beta}_4 \tag{2.18}$$

Using equation 2.12 it follows that,

$$\hat{R} = \tfrac{1}{8}\hat{\beta}_4 = \frac{\hat{\beta}_2}{8\bar{y}} = \tfrac{3}{4}\frac{\hat{\beta}_2}{6\bar{y}} \approx \tfrac{3}{2}\sigma_r^2\frac{\hat{\beta}_2}{\bar{y}} \tag{2.19}$$

Since $p\lim\hat{\sigma}_r^2 = \frac{1}{12}$, we now have $p\lim R = \frac{3}{4}CI$.

Blackburn, ML (1989), Interpreting the magnitude of changes in measures of income inequality. Journal of Econometrics, 42, 21-25.

EUROSTAT (1999), European Community Household Panel (ECHP): selected indicators from the 1995 wave, European Commission, Eurostat, Luxemburg.

Kakwani, NC, Wagstaff, A, van Doorslaer, E, ((1997)). Socioeconomic inequalities in health: measurement, computation and statistical inference. Journal of Econometrics, 77(1), 87-103.

Kennedy, BP, Kawachi I, Prothrow-Stith D (1996), Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States , British Medical Journal, 312(7037): 1004-7.

Kondor, Y, (1971), An old-new measure of income inequality , Econometrica, 39(6), 1041-2.

Kondor, Y, (1975), Value judgements implied by the use of various measures of income inequality , Review of Income and Wealth; 21(3), 309-21.

Lambert, P (1993). The distribution and redistribution of income: a mathematical analysis, 2nd Edition, Manchester University Press.

Mackenbach, JH and AE Kunst, (1997), Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe, Social Science and Medicine, 44(6): 757-71.

Milanovic, B, (1997), A simple way to calculate the Gini coefficient, and some implications, Economics Letters, 45-49

Newey, W and K West, (1987), A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica, 55, 703-708.

Pamuk, E, (1985), Social class inequality in mortality from 1921-1972 in England and Wales, Population Studies, 39, 17-31.

Schutz, RR, (1951) On the measurement of income inequality, American Economic Review, 41, 107-22.

Van Doorslaer, E, A Wagstaff, H Bleichrodt, S Calonge, Ulf-G Gerdtham, M Gerfin, J Geurts, L Gross, U Häkkinen, R Leu, O O'Donnell, C Propper, F Puffer, M Rodriguez, G Sundberg and O Winkelhake, (1997), Income-related inequalities in health: some international comparisons, Journal of Health Economics, 16(1), 93-112.

Van Doorslaer, E, A Wagstaff, H van der Burg, T Christiansen, D De Graeve, I Duchesne, U-G Gerdtham, M Gerfin, J Geurts, L Gross, U Häkkinen, J John, J Klavus, RE Leu, B Nolan, O O'Donnell, C Propper, F Puffer, M Schellhorn, G Sundberg, O Winkelhake, (2000), Equity in the delivery of health care in Europe and the US, Journal of Health Economics 19(5), 553-83.

Wagstaff, A, E van Doorslaer, H van der Burg, S Calonge, T Christiansen, G Citoni, U Gerdtham, M Gerfin, L Gross, U Häkkinen, P Johnson, J John, J Klavus, C Lachaud, J Lauritsen, R Leu, B Nolan, E Perán, J Pereira, C Propper, F Puffer, L Rochaix, M Rodriguez, M Schellhorn, G Sundberg and O Winkelhake, (1999), Equity in the finance of health care: some further international comparisons, Journal of Health Economics, 18, 263-290.

Wagstaff, A, Paci, P, Van Doorslaer, E, (1991). On the measurement of inequalities in health. Social Science and Medicine, 33, 545-557.

Wagstaff, A, (2000), Socioeconomic inequalities in child mortality: comparisons across nine developing countries, Bulletin of the World Health Organization, 78(1): 19-29.

Wagstaff, A and E van Doorslaer, (2000), Equity in health care financing and delivery, In: AJ Culyer and JP Newhouse (Eds.) Handbook of Health Economics, North Holland, 1803-62.

Wagstaff, A, Van Doorslaer, E, Paci, P, (1989). Equity in the finance and delivery of health care: some tentative cross-country comparisons. Oxford Review of Economic Policy 5(1), 89-112.

CHAPTER 3

Estimating relative risks for categorical outcomes

ABSTRACT

This paper presents a method of estimating relative risks, rather than odds ratios, when conducting logistic regression. Odds ratios and relative indices of inequality based on odds ratios are popular instruments when measuring health inequality. However, measures based on odds ratios are hard to interpret when the prevalence of an outcome is greater than a few percent, which is often the case in health inequality research. In this paper we propose a procedure that uses the estimates derived from logistic and other types of regression to compute a (standardized) relative risk. These measures have a more intuitive interpretation and thus allow a better appreciation of the extent of the relative health inequality. We apply the procedure using logistic regression to compute the relative risks between the lowest and highest income quintiles and relative indices of health inequality based on ranking all individuals according to income. In addition, we provide related methods to compute their absolute counterparts and suggest a method for statistical inference. Empirical illustrations are derived from self-assessed health of the elderly in European Union member states. Results show that both magnitude and ranking vary markedly between the relative risk and odds ratio-based methods. This method of relative risk estimation can be applied to other categorical outcome data, statistical models and fields of research.

3.1 INTRODUCTION

In research on socio-economic health inequalities, relative risks (RR) and related measures such as relative index of inequality (RII) are the most commonly used measures of relative health differences (Anand, Diderichsen, Evans, Shkolnikov, and Wirth, 2001; Huisman, Kunst, and Mackenbach, 2003; Mackenbach and Kunst, 1997). These measures have been used to express the magnitude of health inequalities in single, meaningful terms that facilitate comparisons between studies, places and periods. Recent papers stressed that these relative measures need to be complemented by measures of absolute health inequalities, such as rate differences and the slope index of inequality (SII). Nonetheless, measures of relative health inequalities are acknowledged as key instruments for monitoring and explanatory research.

When measures of relative health inequalities are used, the choice for a specific measure is commonly determined by the nature of the health outcome variable that is studied. When count data are used, such as in incidence and mortality analyses, researchers often apply log-linear regression analyses that yield RR estimates. On the other hand, when categorical data are used, such as prevalence rates and cumulative incidence rates, researchers commonly apply logistic regression analyses that produce Odds Ratios.

In fact, the OR offers little more information than the untransformed coefficient of the logistic regression, the log OR, itself. The log OR also has the advantage that when the study groups are interchanged only the sign will differ, enhancing comparability of results across studies. Nevertheless, in epidemiological studies the OR is nearly always preferred over the untransformed coefficients. This suggests the use of the OR as an approximation of the RR in the interpretation of the measured inequality. If the probability of an event is close to zero, such as in population studies of rare diseases, the OR will overestimate the RR only slightly. However, in many epidemiological studies, and especially in studies of socio-economic inequality in health, outcomes are common and the OR may no longer resemble the RR.

The tendency to use OR for categorical outcomes also effected the use of a measure that was especially developed within health inequalities research, the RII (Wagstaff, Paci, and Van Doorslaer, 1991; Mackenbach and Kunst, 1997; Koolman and Van Doorslaer, 2004). The RII has been introduced as the ratio of the probability of self-reported morbidity of the individual at the very bottom of the social hierarchy compared with that of the individual at the very top of the hierarchy. These predicted probabilities are based on the systematic association between morbidity and socio-economic position of all individuals, and are usually adjusted for demographic factors. As such the interpretation of the RII is like a RR. While it may be hard for non-researchers to be on top of the

intricacies of the assumptions involved in the different statistical models, we believe that the general idea is sufficiently intuitive.

In studies with a categorical outcome, however, the RII is typically measured as an OR. This OR can be described as the ratio of the odds of having a health problem in the very bottom of the educational or income hierarchy compared to the very top of the hierarchy, see (Huisman, Kunst, and Mackenbach, 2003). Other studies equally correctly describe the OR, yet the OR remains hard to interpret, let alone when used in combination with the RII concept. This may seriously conflict with the researchers' wish to express the magnitude of health inequalities in concrete terms that can be intuitively grasped by a broad audience.

This paper presents a method to construct RRs in studies with categorical outcomes, irrespective of the statistical model used. First, we will clarify the relationship between the OR and the RR with a description of the mathematical relationship and a numerical example. Next, we will provide a procedure to construct a RR and a RII based on a RR using a logistic regression and suggest methods for statistical inference and estimating absolute counterparts of these relative measures (risk difference and slope index of inequality). We will illustrate these methods by estimating the income related inequality in health of the elderly population of the EU. Finally, we will discuss the pros and cons of presenting RRs versus ORs.

3.2 THE RELATIONSHIP BETWEEN THE ODDS RATIO AND THE RELATIVE RISK

The odds of an event (1) of group *a* is the ratio of the probability of an event over the probability of a non-event (0),

$$Odds = \frac{p(y=1 \mid a=1)}{p(y=0 \mid a=1)} = \frac{\pi_a^1}{\pi_a^0} \tag{3.1}$$

The OR is the ratio of the odds of an event in group *b* divided by the odds of an event in group *a*, as is shown in the first part of equation (3.2). The formal relationship between the OR and RR is given by:

$$OR = \underbrace{\frac{\pi_b^1/\pi_b^0}{\pi_a^1/\pi_a^0}}_{\text{I}} = \underbrace{\frac{\pi_b^1}{1-\pi_b^1} \times \frac{1-\pi_a^1}{\pi_a^1}}_{\text{II}} = \underbrace{\frac{1-\pi_a^1}{1-\pi_b^1}}_{\text{III}} \times RR \tag{3.2}$$

From the last part of equation (3.2) it follows that the OR is equal to the RR multiplied by term III. The latter reduces to one when the prevalence approaches zero. This confirms that the OR best approximates the RR if the prevalence of an event is low. Term III shows

that the OR becomes infinitely larger than the RR if the probability of an event in group b (π_b^1) becomes more common while it does not for group a (π_a^1).

Note that part I of equation (3.2) shows an attractive feature of the OR compared to the RR. Regardless of whether the OR estimates inequality in events or non-events, the magnitude of the inequality and, thus, the ranking of groups is not affected. This is because the OR based on events is the inverse of the OR based on non-events. The RR does not share this property.

Table 3.1 illustrates the relationship between the OR and the RR with a numerical example based on two imaginary countries (X and Y) and two groups (A and B). In the first country the prevalence is low (5%) and in the second the prevalence is high (60%). The OR approximates the RR in the first country, but is much higher in the second. In fact, if one mistakenly interprets the OR of country B as a RR, then the risk difference (RD) between both groups is more than 100%.[a] Furthermore, it is clear that when prevalence levels vary between countries, so may inequality rankings. Note that this reranking will not take place in a study where all groups are compared to one reference group, i.e. the prevalence within the reference groups does not vary.

The above illustrates that in studies where the outcome is common, RRs and ORs-based measures are quite distinct. In practice, researchers often appear not to make an explicit choice between the two, by presenting the typical output that follows from the appropriate statistical model. For example RRs are frequently presented when the outcome measure is survival time or a count and ORs are commonly presented when the outcome is categorical. Below we will present a general method to compute RRs irrespective of the statistical model used.

Table 3.1 Example relationship relative risk and odds ratio

	Probability of event	
	Country X	Country Y
Prevalence (%)	0.05	0.6
Group A	0.02	0.3
Group B	0.08	0.9
Relative risk[a]	4.0	3.0
Odds ratio[a]	4.3	21.0
Risk Difference[a]	0.06	0.60
Risk Difference[a,b]	0.06	1.09

[a] Group A is the reference group

[b] Risk difference when odds ratio is interpreted as relative risk and average prevalence levels are held constant

3.3 A METHOD TO ESTIMATE RELATIVE RISKS

The procedure to compute a RR varies with independent variable of interest. We will begin with a dichotomous dependent variable and a dichotomous independent (dummy) analysed using a logistic regression, which is the most common model used to compute ORs. It is popular when compared to other regression models because (1) fitted probabilities always remain within the [0,1] interval; (2) the partial effect of a characteristic on the outcome is modified by other characteristics of the individual included in the model and depends on the level of the variable when the variable is not dichotomous (due to the non-linear transformation: the link function).[b] (3) coefficients are log ORs which can easily be transformed into ORs. The first two advantages remain in the following procedure.

The procedure to compute a RR for a dummy is presented in four steps:

1. Estimate the logistic regression and retain its coefficients;
2. Predict the outcome for each individual while fixing the category of interest at one;
3. Repeat Step 2 but now fix the dummy at zero;
4. Divide the average outcome of Step 2 by the average outcome of Step 3.

This will result in a RR. This RR is directly comparable to the OR that follows from a logistic regression.[c, d]

To obtain an adjusted RR one may add standardising variables to the regression model. More formally, when using a logistic regression Step 2 and 3 can be calculated using

$$\hat{p}(y|d,x) = \frac{1}{1+e^{-(\hat{\alpha}+\hat{\beta}d+\hat{\delta}_1 x_{i1}+\ldots\hat{\delta}_k x_{ik})}} \qquad (3.3)$$

where y is the outcome variable; d is the dummy of the independent variable of interest, $x_{i1}\ldots x_{ik}$ are the individual specific values for the k standardising variables and $\hat{\alpha}, \hat{\beta}$ and $\hat{\delta}_{1\ldots k}$ are the corresponding estimated coefficients from the logistic regression. This formula is used to predict the probability of an event for each individual in the sample while holding d constant at one. The distribution of the standardising variables as observed in the study population should not be altered during Step 2 and 3. The resulting predictions of Step 2 can be interpreted as the predicted probability of an event had each individual been part of the group where d equals one. Step 3 is similar except term $\hat{\beta}d$ is now left out. The average predicted probabilities needed for Step 4 are defined as $\bar{\hat{p}}_d = 1/n \sum_{i=1}^{n} \hat{p}_{i,d}$. The above shows that this RR is based on what is sometimes referred to as the average partial effect. This procedure implicates that the resulting RR is based on the functional form imposed by the logistic regression, or whichever regression model is chosen. Please note that (all) regression based

standardisation requires a structural model, such that estimated coefficients have a causal interpretation.

If the variable of interest is categorical then all relevant categories need to be manipulated analogous to the adjustment of the dummy in the above procedure. If the variable of interest is a rank variable with ranks varying between zero and one, then the rank should be set to one in Step 2 and to zero in Step 3. The latter procedure produces a relative index of inequality based on a RR and can be interpreted as the average RR of each individual reporting an event had he or she moved from the very highest to the very lowest rank. For count or continuous variables, one may estimate similar partial effects and use them to construct a RR.

For statistical inference, we suggest using resampling techniques (Efron and Tibshirani, 1994). Resampling techniques require no distributional assumptions and can be adjusted to take the complex sampling frame of many modern surveys into account. We prefer the bootstrap method rather than the jackknife method, as the jackknife can be sensitive to non-smooth distribution of the statistic of interest. While this is not problematic in most instances, it may in some.

We can also manipulate this procedure to estimate the absolute counterparts of the RR and the relative index of inequality: the risk difference (RD) and the slope index of inequality (SII). This would amount to subtracting the average probabilities that result from Step 2 from those of Step 3. The RD is the average predicted difference in risk of an event between two groups, adjusted for all other variables in the analysis. Similarly, the SII represents the absolute difference in the probability of reporting an event between the person with the lowest rank and the highest rank. The SIIs are based on the same model assumptions that underlie their relative counterparts such as those related to the functional form. This SII, therefore, does not impose a linear relationship between the rank and the outcome variable. The SII reflects the average standardised difference in outcome for everyone in the sample had he/she either the highest or the lowest rank.

The above procedures may seem straightforward; its application can prove cumbersome. To overcome this drawback, we provide two solutions. First, on request we provide fully automated and annotated syntax for the statistical packages SPSS, SAS and Stata. Secondly, we present and discuss a more convenient alternative method to compute a RR based on an individual with average characteristics.

3.4 RELATIVE RISK BASED ON AN AVERAGE INDIVIDUAL

Statisticians and econometricians sometimes construct an average individual and estimate the effect of a change in the variable of interest on the outcome for this individual, rather then predicting differences in for each individual in the sample. This is

sometimes referred to as the partial effect evaluated at the mean. We will refer to the RR based on this procedure as the RR evaluated at the mean (RRm).

For this method we introduce an average individual who has the average standardising variable characteristics. The procedure is as follows:

1. Estimate the logistic regression on the entire sample to obtain its coefficients;
2. Compute the averages for all standardising covariates that describe the characteristics of the average individual;
3. Predict the outcome for the average individual whilst in the lowest socio-economic class;
4. Repeat Step 2 for the highest socio-economic class;
5. Dividing the outcome of Step 2 by that of Step 3.

Formally, when using logistic regression Step 2 can be calculated using:

$$\hat{p}(y|d,x) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 d_1+\hat{\delta}_1\bar{x}_1+\ldots\hat{\delta}_k\bar{x}_{ik})}} \qquad (3.4)$$

where $x_{1\ldots k}$ are the standardising variables that are now kept at their means. Setting dichotomous covariates (dummies) at their averages implies setting the value of the average individual to the proportion of individuals with that characteristic. For example, assigning the value 0.532 for the dummy (female) when 53.2 percent of the (weighted) sample is female.

The relative index of inequality based on PEM is straightforward to compute, but comes at a price. First, like the popular conversion presented by (Zhang and Yu, 1998) it fails to take full account of the way the distribution of the standardising variables affect the partial effect. Secondly, the average individual does not exist and the method is therefore sometimes argued to be unappealing. Consequently, some researchers prefer to replace the means of the standardising variables with their most likely value, in order to create a reference individual with more likely characteristics. In the example above this individual would be female. As a result, men are not represented in the RR, even though sex was adjusted for in the regression. Likewise researchers sometimes prefer to use median values for continuous variables to hold the standardising variables constant. The median may be a better representative of the populations if the continuous variable has a skewed distribution.

3.5 EMPIRICAL ILLUSTRATION

To show the effect of the methods we have re-estimated the results from Huisman, Kunst, and Mackenbach (2003). As we have used data from the same source (European

Community Household Panel, ECHP) and the same year (1994), we refer the reader to Huisman, Kunst, and Mackenbach (2003) for a discusssion of the data, methods and results. Because the provider of the data (Eurostat) does not allow users to keep a copy of the data, our estimators are based upon an updated version of the database (2002) and are, therefore, slightly different.

As suggested above, we used the bootstrap method for statistical inference. The standard bootstrap procedure hinges on the assumption that the observed distribution is a random sample of the underlying population distribution and that individuals within the sample are independent. This assumption does not hold for the complex stratified multi-stage sampling designs used in many countries to gather ECHP survey data. To mimic the complex sample design, for each replication we stratified each country sample equal to the original stratification. Then, we drew a random sub-sample with replacement of the primary sampling unit and subsequently from the secondary sampling unit (the household) each of a size equal to the original sample size, and included all individuals in the selected households in each bootstrap sample. For each of the 2000 bootstrap samples we computed the RR using the above procedure. Finally, we based the 95% confidence intervals on the percentile scores. The bootstrap procedure was adjusted to take account of different sampling frames used and for countries that did not make the relevant information available (see table 3.3).

Table 3.2 shows the prevalence, rate difference, OR and RR of reporting less than good health of the lowest compared to the highest income quintile for EU countries in 1994. The analysis is performed on individuals age 60 and above and is stratified for both men and women. The results are directly comparable to Table 4 of Huisman, Kunst, and Mackenbach (2003). The prevalence of reporting less than good health varies markedly from country to country and within country between sexes. The lowest observed prevalence is 41 percent for Danish men and the highest is 86 percent for Portuguese women. For the comparison of groups we chose the highest socio-economic group is as the reference group. The RD is positive for most groups, which indicates that individuals in the lowest socio-economic group have a higher risk of reporting less than good health in all countries for both men and women. Consequently, most ORs are greater than one. It is striking to see how much ORs vary between countries. For males, the lowest inequality measured in ORs was observed among Belgian men (OR = 1.2) and the highest inequality among Irish males (OR = 3.5). While for females, the lowest for inequality was observed for Belgium (OR = 0.9) and the highest for Greece (OR = 2.7). Throughout, we obtain clearly lower RRs and observe less variation. Should one mistakenly interpret ORs as RRs, then one would ‘overestimate’ the relative difference between 1.7 times for Belgian men, to 7.4 times for Portuguese women. Even though the

Table 3.2 Prevalence, risk difference, odds ratio and relative risk of reporting less than good health of the lowest compared to the highest income quintile, EU countries, 1994, men and women 60+

Country	No. obs.	Prevalence (%)	Risk Difference	Odds Ratio	Relative Risk	(OR-1)[a] / (RR-1)
Men						
Belgium	740	44%	5%	1.22	1.12	1.7
Denmark	644	41%	15%	1.87	1.46	1.9
France	1527	60%	12%	1.69	1.24	2.9
Germany	979	57%	17%	2.00	1.34	2.9
Greece	1646	60%	11%	1.60	1.22	2.7
Ireland	1000	43%	28%	3.49	2.13	2.2
Italy	1744	71%	15%	2.15	1.26	4.5
Netherlands	927	46%	12%	1.64	1.32	2.0
Portugal	1543	76%	21%	3.19	1.37	5.9
Spain	2082	65%	17%	2.17	1.33	3.6
United Kingdom	1257	48%	19%	2.22	1.58	2.1
Women						
Belgium	936	56%	-1%	0.94	0.97	2.0
Denmark	817	50%	8%	1.41	1.21	2.0
France	1939	67%	14%	1.88	1.26	3.4
Germany	1174	64%	8%	1.43	1.13	3.2
Greece	1960	68%	21%	2.66	1.40	4.2
Ireland	1017	46%	18%	2.16	1.58	2.0
Italy	2068	80%	10%	1.91	1.14	6.7
Netherlands	1126	51%	9%	1.48	1.22	2.2
Portugal	1898	86%	4%	1.36	1.05	7.4
Spain	2675	74%	14%	2.03	1.21	4.9
United Kingdom	1604	50%	14%	1.78	1.40	2.0

[a] Displays the OR's 'overestimation' of RR compared to no inequality (OR = RR = 1).

extremes for males remain the same, the ranking of countries clearly depends upon the choice for either OR or RR.

Table 3.3 presents a SII and RII based on OR and RR with confidence intervals of reporting less than good health of men and women in the lowest income compared to the highest income for EU countries in 1994. A SII is estimated using the average partial effects. RIIs based on ORs deviate considerably more from one (equality) than those based on RRs. The differences are now even greater than those between RRs and ORs (table 3.2) because the SIIs are greater than the RDs.

Table 3.3 Prevalence, slope index of inequality and relative index of inequality in reporting less than good health, EU countries, 1994, men and women 60+

Country	Prevalence	Slope Index of Inequality	Relative Index of Inequality			
			Odds Ratio	Relative Risk	95% CI of RR	Relative Risk Me
Men						
Belgium[a]	44%	8%	1.38	1.20	[0.90-1.66]	1.20
Denmark[b]	41%	22%	2.54	1.74	[1.25-2.51]	1.75
France[c]	60%	17%	2.04	1.33	[1.15-1.56]	1.33
Germany[d]	57%	20%	2.27	1.42	[1.17-1.74]	1.43
Greece[a]	60%	24%	1.95	1.31	[1.13-1.52]	1.30
Ireland	43%	15%	4.92	2.51	[1.91-3.40]	2.67
Italy[c]	71%	36%	2.83	1.36	[1.22-1.54]	1.35
Netherlands[d]	46%	20%	2.05	1.47	[1.15-1.92]	1.48
Portugal	76%	17%	4.34	1.43	[1.28-1.61]	1.41
Spain	65%	25%	3.07	1.48	[1.32-1.67]	1.47
United Kingdom	48%	24%	2.74	1.69	[1.36-2.13]	1.71
Women						
Belgium[a]	56%	3%	1.11	1.05	[0.85-1.30]	1.05
Denmark[b]	50%	14%	1.78	1.33	[1.04-1.72]	1.33
France[c]	67%	20%	2.52	1.36	[1.21-1.54]	1.36
Germany[d]	64%	12%	1.73	1.22	[1.05-1.43]	1.22
Greece[a]	68%	16%	3.34	1.48	[1.32-1.69]	1.47
Ireland	46%	25%	2.72	1.73	[1.35-2.26]	1.74
Italy[c]	80%	23%	2.25	1.18	[1.09-1.28]	1.17
Netherlands[d]	51%	13%	1.65	1.28	[1.05-1.59]	1.28
Portugal	86%	12%	1.37	1.04	[0.98-1.12]	1.04
Spain	74%	4%	2.61	1.28	[1.18-1.40]	1.28
United Kingdom	50%	18%	1.88	1.37	[1.15-1.65]	1.37

[a] Strata were not made available/unreliable.

[b] Original households were sampled directly.

[c] Primary sampling units were not made available/unreliable.

[d] Strata and primary sampling units were not made available.

[e] Relative risk is evaluated for a individual with mean characteristics.

Table 3.4 shows the effect the choice of measure has on the ranking of countries according to income related inequality in reporting less than good health. The RII based on an OR results in a much lower position for Portuguese men than that based on either RR or RRm. Similarly we observe a drop for the women of the United Kingdom of several

positions when in the inequality measure is based upon a RR. When we compare RR and RRm-based measures of inequality, we see that the ranks are mostly comparable, with only few countries trading places.

3.6 DISCUSSION

We set out to present a procedure to construct relative risks (RR) irrespective of the statistical model used. This procedure allows researchers to consider both odds ratios (ORs) and RRs, whereas previously, a categorical outcome variable almost automatically implied the presentation of an OR. As this procedure is somewhat cumbersome to compute, we offer a straightforward procedure based on a hypothetical individual with average characteristics and example programs for popular statistical packages. We also propose ways to construct the absolute difference counterparts of these relative measures and suggest a method to perform statistical inference. In an empirical illustration we apply these methods to measure income related health inequality in the elderly population of the European Union.

The results show that the RR and the RR evaluated for an individual with (sample) mean characteristics (RRm) show little difference. There are two explanations to why the observed differences are so small. First, there may be a small effect of the curvature of the relationship between income rank and self-reported less than good health. This may not be surprising as curvature plays less of a role when estimating sample indices with a prevalence of around 50 percent and because our income variable is already transformed into relative ranks. Secondly, other characteristics may not influence the partial effect on the outcome very much in this example. This too, seems reasonable as

Table 3.4 Order of EU countries from the lowest to the highest income related inequality of self assessed health as measured by several methods, 1994, men and women 60+

Men				Women		
RII-OR	RII-RR	RII-RRm		RII-OR	RII-RR	RII-RRm
Belgium	Belgium	Belgium	1	Belgium	Portugal	Portugal
Greece	Greece	Greece	2	Portugal	Belgium	Belgium
France	France	France	3	Netherlands	Italy	Italy
Netherlands	Italy	Italy	4	Germany	Germany	Germany
Germany	Germany	Portugal	5	Denmark	Netherlands	Spain
Denmark	Portugal	Germany	6	UK	Spain	Netherlands
UK	Netherlands	Spain	7	Italy	Denmark	Denmark
Italy	Spain	Netherlands	8	France	France	France
Spain	UK	UK	9	Spain	UK	UK
Portugal	Denmark	Denmark	10	Ireland	Greece	Greece
Ireland	Ireland	Ireland	11	Greece	Ireland	Ireland

we adjusted for age only and all age groups are mutually exclusive. Had we used several categories of standardising variables, which is often the case with health care utilisation research, then the differences could have been more significant. Nonetheless, we think that RRm-based methods will generally make a good approximation of the RR-based methods and have the advantage of being much less cumbersome to compute.

The results show that the OR cannot be viewed as an approximation of a RR. That the relationship is demonstrably weaker as the prevalence is higher. That the ranking of countries varies between the two measures. Clearly the choice between both measures is an important one and can now be based on theoretical rather than distributional considerations. Therefore, the question arises whether RR should be preferred above OR-based measures in research on socio-economic inequalities in health? We think there is no definitive answer to this question. Underneath we will list some considerations when choosing between the two.

First and foremost, the preferred measure should capture relative inequality best, i.e. matches best with societal notions about relative inequality. A RR of three means that the relative difference of two groups/individuals is a factor three, irrespective of the overall prevalence. While in such a case, the OR varies between just above three to infinity as prevalence levels go up. In other words, we need to answer the question whether society is more worried about relative inequality when the prevalence is higher. And if so, whether this worry is correctly represented by the weight the OR attaches to it. Although an answer may not be available at this time, an infinite weight when the prevalence level approaches one is certainly questionable. Incidentally, the above also explains why ORs are hard to interpret by non-researchers. For that reason alone, they are unsuitable for descriptions of health inequalities directed at the general public or other uninitiated audiences. One way to deal with the issue of weighting prevalence is to present both the RR and its absolute counterpart which are related through overall prevalence.

When the dependent variable is ordered and the choice of cut-off is subjective, then this choice will not affect the OR under the assumption of proportional odds (also referred to as the parallel regression assumption). However, as prevalence levels will vary with the choice of cut-off so will RRs. By implication -under the proportional odds assumption- ORs will be more comparable across countries or over time if cut-points between self-assessed health categories vary between countries or changes over time, as they may (Sen, 2002). In practice though, tests as provided by (Brant, 1990) often show that the assumption of proportional odds is violated (Scott Long, 1997), in which case the OR is sensitive to the choice of cut-point.

Furthermore, the RR is sensitive to the choice of outcome category that is defined as an event (1) and which as a non-event (0) (i.e. whether one is evaluating the inequality

in health or ill-health). As this choice influences both magnitude and ranking, it can never be made arbitrarily. The RR thus forces the researcher to be explicit in the choice for a positive or a negative definition of the health outcome indicator. The RR shares this feature with the concentration index (Clarke, Gerdtham, Johannesson, Bingefors, and Smith, 2002; Koolman and Van Doorslaer, 2004). This drawback may become an advantage when both positive and negative indicators are of interest, but are not equally important or have different implications for research or policy.

The OR of inequality in health is the inverse of the OR of inequality in ill-health. Thus ranking never varies with the choice of which category to label as an event. In fact, the log OR (the untransformed coefficient) only changes sign when the choice of event is reversed, facilitating comparisons over time or place. Because of this and the fact that the log OR does not tempt people to interpret it as a RR, we recommend presentation of the log OR above the OR, especially when the outcome is not infrequent (roughly above 5%).

Although the choice between a RR and a OR may depend on many factors, the freedom to choose between both of them when the dependent variable is categorical is a step forward and allows a careful choice of methods to describe and display socio-economic inequalities in health.

Notes

a. The bottom right figure was derived assuming the OR in interpreted as a RR, i.e. Rb/Ra=21 (E1a), which can be rewritten as Rb = 21Ra (E1b), while the average prevalence equals 0.6, i.e. Rb + Ra = 1.2 = 66/55 (E2). Substituting equation E1b in equation E2 gives 22Ra = 66/55. Solving gives Ra = 3/55 and Rb = 63/55. To obtain the RD we subtract Ra from Rb, which equals 60/55, which we rounded to 1.09.
b. See for a discussion on concepts of interaction the chapter by the same title in Rothman and Greenland (1998).
c. The exponentiation of the log link (rather than the logistic link) function coefficients directly give RRs. A drawback of this link function is that the predicted probabilities are no longer confined to the [0,1] inter val.
d. As expected, when this procedure is applied using Poisson regression, the result of Step 4 provides the exact RR normally obtained by taking the antilog of the coefficient.

Anand, S., Diderichsen, F., Evans, T., Shkolnikov, V.M., & Wirth, M. (2001). Measuring Disparities in Health: Methods and Indicators. In T. Evans,M. Whitehead, F. Diderichsen, A. Bhuiya, & M. Wirth (Eds.), Challenging Inequities in Health; From Ethics to Action. Oxford: Oxford University Press.

Brant (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics, 46, 1171-1178.

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.

Efron, B., & Tibshirani, R. (1994). An Introduction to the Bootstrap Boca Raton: CRC Press

Eurostat (2002). European Community Household Panel Users Database.

Huisman, M., Kunst, A.E., & Mackenbach, J.P. (2003). Socioeconomic inequalities in morbidity among the elderly; a European overview. Social Science and Medicine, 57(5), 861-873.

Koolman, X., & Van Doorslaer, E.K.A. (2004). On the interpretation of a concentration index of inequality. Health Economics, 13, 649-656.

Mackenbach, J.P., & Kunst, A.E. (1997). Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. Social Science and Medicine, 44(6), 757-771.

Rothman, K.J. & Greenland, S., 1998, Modern epidemiology, Lippincott Williams & Wilkins

Scott Long, J. (1997). Regression models for categorical and limited dependent variables Thousand Oaks: Sage publications

Sen, A. (2002). Health: perception versus observation; Self reported morbidity has severe limitations and can be extremely misleading. BMJ, 324, 860-861.

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

Wooldridge, J.M. (2002). Econometric analysis of cross section and panel data Cambridge: Massachusetts Institute of Technology

Zhang, J., & Yu, K.F. (1998). What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA, 280(19), 1690-1691.

CHAPTER 4

Health inequalities across European countries decomposed

ABSTRACT

This paper provides new evidence on the sources of differences in the degree of income-related inequalities in self-assessed health in thirteen European Union member states. It goes beyond earlier work by measuring health using an interval regression approach to compute concentration indices and by decomposing inequality into its determining factors. New and more comparable data were used, taken from the 1996 wave of the *European Community Household Panel*. Significant inequalities in health (utility) favouring the higher income groups emerge in all countries, but are particularly high in Portugal and – to a lesser extent – in the UK and in Denmark. By contrast, relatively low health inequality is observed in the Netherlands and Germany, and also in Italy, Belgium, Spain Austria and Ireland. There is a positive correlation with income inequality *per se* but the relationship is weaker than in previous research. Health inequality is not merely a reflection of income inequality. A decomposition analysis shows that the (partial) income elasticities of the explanatory variables are generally more important than their unequal distribution by income in explaining the cross-country differences in income-related health inequality. Especially the relative health and income position of non-working Europeans like the retired and disabled explains a great deal of "excess inequality". We also find a substantial contribution of regional health disparities to socio-economic inequalities, primarily in the Southern European countries.

4.1 INTRODUCTION

Persistent differences in health by socio-economic status (SES) have long been a serious health policy concern in many European countries and have recently also been put at the forefront of the European Union's National Action Plans as agreed upon at the Lisbon European Council. In the recent Atkinson Report's (Atkinson et al., 2002) recommendations on Indicators for Social Inclusion in the EU, a less unequal distribution of self-reported health by income quintiles is seen as an intrinsic part of the broad goals of social inclusion and cohesion endorsed by the European Union. Comparative research which exploits cross-national variation but uses comparable data from all member states offers the prospect of providing insights into the reasons for cross-country differences.

In previous work (Van Doorslaer et al., 1997), we have employed concentration indices and curves to test for differences across eight European countries and the US in the extent to which self-reported (ill-) health was unequally distributed across income. We then found that significant income-related health inequalities were present in all countries but also found substantial differences across countries. A country's degree of (income-related) health inequality appeared to be particularly closely associated with its degree of income inequality. While a fair degree of comparability was achieved, it was still limited by the fact that the comparison had to be based on secondary analysis of country-specific health interview and socioeconomic surveys. Inevitably, and despite all efforts at harmonization and improved comparability, differences in income and health variable definitions constrained the type of comparisons that could be performed. A similar study for six European countries using comparable methods also found that income-related inequality in self-assessed health was strongly associated with income inequality (Cavelaars et al., 1997). They noted that the measured degree of income-related health inequality decreased substantially when other socio-economic indicators, notably education level, occupational status and especially employment status were controlled for.

This paper updates and substantially extends the evidence on income-related inequalities in Europe by using both new data sources and new methods. First, we use new data from the third wave (1996) of the *European Community Household Panel* (ECHP) Study, which were recently released by the European Commission's Statistical Office (EUROSTAT). It provides a rich new source of comparable household level data on income, health and various other socio-economic characteristics from all European Union member states. Second, we measure and explain self-reported health using a recently proposed method based on interval group regression (Van Doorslaer and Jones, 1997). Third, we explore potential causes of cross-country differences in income-related health inequality using a new method for decomposing the estimated inequality into the contributions of

various determinants (Wagstaff et al., 2003). Fourth, we examine the causes of 'excess inequality' by decomposing the differences with the European country with the lowest degree of inequality. Finally, we perform statistical inference based on a bootstrapping procedure of the decomposition method.

The paper is organised as follows. In the next section we set out the methods used to measure, explain and decompose the sources of inequality. Section three describes the data and variables used and section four gives the results of the analysis. We end with conclusions and discussion in section 4.5.

4.2 METHODS

4.2.1 Measurement of health

As in previous work, we mainly use respondents' self-assessments as our measure of general health status. While this measure may seem simple and subjective, it has been shown to be a powerful predictor of subsequent mortality (Idler and Benyamini, 1997) and, more importantly, its predictive power does *not* appear to vary systematically by SES (Burström and Fredlund, 2001) which means that inequalities in self-assessed health (SAH) also have predictive power for inequalities in mortality (Van Doorslaer and Gerdtham, 2003). It basically provides an ordinal ranking of individuals' self-perception of their health status. However, in contrast to previous work we now have empirical distributions of SAH based on the same question "How is your health in general?" and the same five response categories ranging from "very good" to "very bad". This is the WHO recommended wording for self-perceived health questions in European health interview surveys (WHO, 1996). This health ranking variable basically provides an ordinal ranking to which many previous researchers have responded by dichotomising it into two categories: individuals reporting to be in good or very good health versus those in less-than-good-health. This practice has been shown to have the undesirable property of leading to rank reversals when comparing health inequality over time in the Netherlands (Wagstaff and Van Doorslaer, 1994) or across countries in Europe (Van Doorslaer and Koolman, 2000).

One approach used to exploit all information contained in the 5-point scale, was to assume that underlying these responses is a latent self-assessed health variable with a skewed, standard lognormal distribution to obtain latent ill-health scores for each of the response categories (Van Doorslaer et al., 1997; Wagstaff and Van Doorslaer, 1994). While this approach makes the arbitrary assumption that the distribution of latent ill-health is identical (i.e. the standard lognormal one) in all countries, its joint distribution with income can still vary and the relative inequality associated with income around this identical mean can still be examined. The restrictive assumption of a common

underlying health distribution was necessitated by lack of a cardinal scale for the ill-health scores and by the differing questions and response categories across countries. The availability of one SAH measure with an identical wording of question and response categories in the ECHP (albeit translated and asked in the various European languages) makes this restrictive standard lognormal assumption redundant. A more promising strategy is to use information on the empirical distribution of generic health measures with well-tested properties. While preference-based valuations of health states are now available also for generic instruments like the SF-36 (Brazier et al, 1998) and the EuroQol (Dolan, 1997), the Canadian Health Utilities Index (HUI) Mark III (Feeny et al., 1995; Feeny et al., 2002) is the only instrument which is now routinely included in a general population health survey. Scoring the SAH levels using external information on an instrument such as the HUI therefore has the attraction of obtaining a more 'natural' index for the SAH scores as utilities between 0 and 1.

Appropriate econometric analysis of an ordered categorical dependent variable, such as SAH, is typically based on the ordered probit or logit model or, if information on the scaling of the variable is available, the interval (or grouped data) regression model. Interval, or grouped data, regression provides a more efficient alternative to the ordered probit model when the values of the boundaries of the intervals are known (Jones, 2000). Our approach here is to use the empirical distribution function of HUI scores in the 1994 Canadian *National Population Health Survey* sample obtained in Van Doorslaer and Jones (1997) to scale the intervals of SAH for all European countries. To do this we assume that there is a stable mapping from HUI to the (latent) variable that determines reported SAH and that this applies not only to Canadian but also to European individuals. We compute the cumulative frequency of observations for each category of SAH and then find the thresholds μ_j of the empirical distribution function (EDF) for HUI that match these frequencies. Formally,

$$\mu_j = F^{-1}(G_j) \qquad (4.1)$$

where $F^{-1}(.)$ is the inverse of the EDF of HUI and G_j is the cumulative frequency of observations for category j of SAH.

Because we use HUI thresholds to scale SAH, the linear index $x_i\beta$ for the interval regression model gives us a prediction of each individual's level of health utility as derived from the observed SAH level. It is the predicted level of HUI knowing that an individual has characteristics x. The prediction is both continuous and linear in the x_i's. Linearity is a useful property which implies that concentration indices calculated using the predictions are suitable for decomposition analysis (cf. section 2.3). In effect, the interval regression technique exploits the between-SAH category variation to generate some within-SAH category variation in HUI, while HUI itself is unobserved. Moreover,

by incorporating external information to scale the categorical observations of SAH, the predictions are measured on the same scale as HUI and do not require *ex post* re-scaling, as is often done with ordered probit predictions.

4.2.2 Measurement of inequality

As before, we use the *health concentration index* as our measure of *relative* income-related health inequality (Wagstaff et al., 2003).[a] Suppose we have a continuous cardinal measure of health (utility) y_i. A concentration curve $L(s)$ plots the cumulative proportion of the population (ranked by income, beginning with the lowest incomes) against the cumulative proportion of health. If $L(s)$ coincides with the diagonal, everyone enjoys the same health. If, by contrast, $L(s)$ lies *below* the diagonal, inequalities in health exist and favour the richer members of society. The further $L(s)$ lies from the diagonal, the greater the degree of inequality. The health concentration index, C, is defined as twice the area between $L(s)$ and the diagonal. C takes a value of zero when $L(s)$ coincides with the diagonal and is negative (positive) when $L(s)$ lies above (below) the diagonal. The minimum and maximum values of C using individual-level data are -1 and +1 respectively: these occur when all the population's ill-health is concentrated in the hands of the most and least disadvantaged persons respectively.

Since the estimation and comparison of inequality estimates across countries requires representative and therefore suitably weighted sample data, the computation formula for C given by Kakwani et al. (1997) can be modified to accommodate sample weighting as follows:

$$C = \frac{2}{N\mu} \sum_{i=1}^{N} w_i y_i R_i - 1 \qquad (4.2)$$

where,

$$\mu = \frac{1}{N} \sum_{i=1}^{N} w_i y_i \qquad (4.3)$$

is the (weighted) mean health of the sample, N is the sample size, w_i is the sampling weight of individual i (with the sum of w_i equal to N), and R_i is the fractional rank (for weighted data) of the ith individual. The latter is defined as Lerman and Yitzhaki (1989):

$$R_i = \frac{1}{N} \sum_{j=1}^{i-1} w_j + \tfrac{1}{2} w_i \quad \text{where } w_0 = 0 \qquad (4.4)$$

and thus indicates the weighted cumulative proportion of the population up to the midpoint of each individual weight.

C can be computed conveniently using the weighted covariance of μ and the fractional rank (for weighted data) as Kakwani et al. (1997):

$$C = \frac{2}{N\mu} \sum_{i=1}^{N} w_i \left(y_i - \mu\right)\left(R_i - \tfrac{1}{2}\right) = \frac{2}{\mu}\mathrm{cov}_w \left(y_i, R_i\right) \qquad (4.5)$$

where cov_w denotes the weighted covariance.

4.2.3 Decomposing inequality

A straightforward way of decomposing the measured degree of inequality into the contributions of explanatory factors, proposed in Wagstaff et al. (2003), requires the specification of a linear additive regression model of health such as

$$y_i = \alpha + \sum\nolimits_k \beta_k x_{ki} + \varepsilon_i \qquad (4.6)$$

where y is the health measure, the x_k variables are health determinants and ε is a disturbance term. One could think of this equation as a reduced form of a demand for health equation where all x_k are exogenous determinants. Given the relationship between y_i and x_{ki} in equation (4.6), the concentration index for y, C, can be written as:

$$C = \sum\nolimits_k (\beta_k \overline{x}_k / \mu) C_k + GC_\varepsilon / \mu \qquad (4.7)$$

where μ is the mean of y, $\overline{x}_k$ is the mean of x_k, C_k is the concentration index for x_k (defined analogously to C) and GC_ε is the generalized concentration index for ε_i. Equation (4.7) shows that C is equal to a weighted sum of the concentration indices of the k regressors, where the weight or "share" for, say, x_k, is the elasticity of y with respect to x_k. The residual component — captured by the last term — reflects the inequality in health that is not explained by systematic variation across income groups in the x_k. In the case of the interval regression approach, no residuals can be computed and the decomposition reduces to the first term in equation (4.7).

If we define the estimated health elasticity of determinant k as

$$\hat{\eta}_k \equiv \hat{\beta}\, \overline{x}_k / \mu_k \qquad (4.8)$$

and using estimated concentration indices, we can rewrite the decomposition as

$$\hat{C} = \sum\nolimits_k \hat{\eta}_k \hat{C}_k \qquad (4.9)$$

In other words, estimated health inequality is just a weighted sum of the inequality in each of its determinants, with the weights equal to the health elasticities of the determinants. As a result, total inequality can be partitioned into what has been labeled (cf. Kakwani et al., 1997) 'potentially avoidable' and 'unavoidable' health inequality. The unavoidable part of the inequalities is comparable to, for instance, the age-sex

expected health inequality by income due to demographics. Using the method of indirect standardisation (Kakwani et al., 1997), one can compute the age-sex expected inequality as C^* and subtract this from C to obtain an estimate of 'potentially avoidable' inequality as $I^* = C - C^*$. Note that if y is predicted *only* on the basis of age and sex (as in (Van Doorslaer et al., 1997)), then $\hat{C}$ in equation (4.9) is identical to C^*, i.e. the degree of inequality to be expected simply on the basis of the unequal distribution of age and sex across income groups. The current approach, based on 'full' equations rather than 'auxiliary' standardising regressions for age and sex only, has two advantages. First, we standardise only for the purely 'demographic', i.e. partial effects on health, not the 'total' effects of age and gender. If other relevant variables are omitted in the standardising equations, these 'partial' effects run the risk of omitted variable bias (Schokkaert and Van de Voorde, 2003; Gravelle, 2001). Secondly, the unavoidability of, for instance, gender-related differences in income-related health inequality is contentious and it seems therefore preferable to decompose the total relative inequality in health into its various determinants, including demographic structure, leaving it to the user of the results to standardise for whatever background characteristics are deemed appropriate.

The decomposition also makes clear how each determinant k's separate contribution to total income-related health inequality can be decomposed into two meaningful parts: (i) its impact on health, as measured by the health elasticity (η_k), and (ii) its degree of unequal distribution across income, as measured by the (income) concentration index (C_k). This decomposition method therefore not only allows to separate the contributions of the various determinants, but also to identify the importance of each of these two components within each factor's contribution. This property makes it a powerful tool for unpacking the mechanisms contributing to a country's degree of health inequality.

4.2.4 Statistical inference

When inequality has been measured and decomposed into its sources, it is useful to be able to statistically test for cross-country differences. Given the complexity of the survey designs of the ECHP samples and the complex composition of the contribution terms in equation (4.9), we have opted to use a "bootstrap" method (Efron and Tibshirani, 1993; Deaton, 1997) to assess sampling variability and to obtain standard errors for the estimates of both C and for $\eta_k C_k$, for each k. A bootstrap procedure hinges on the assumption that the observed distribution is a random sample of the underlying population distribution, and that individuals within the sample are independent. In our bootstrap we have corrected for differences in sampling probability, but not for the different types of the multi-stage sampling designs used in the ECHP. The latter was

impossible because the necessary information — such as the primary sampling units — was not provided for all countries. Our bootstrap estimators of the t-statistics were computed using a five-step approach. First, we have inflated our sample size to allow for the differences in sampling probability by multiplying the sampling weights by the inverse of the smallest weight and rounded to the nearest integer. Second, from this expanded sample we have drawn a random sub-sample of the size of the original sample with replacement. Third, we have run the entire procedure to obtain estimates of the factor contributions and I^*, including the interval regression, the construction of a fractional rank and a covariance matrix. This procedure differs from the calculation of the point estimates in that it is not weighted, as the differences in sampling probability are taken into account in the inflation of the sample. Fourth, by repeating this whole process, we generated 1000 resample data sets, each providing us with estimates of the contributions. Fifth, using these datasets we computed the standard deviations and t-statistics for each factor's contribution and for I^*.

4.2.5 Decomposing inequality differences between countries

More interesting than a mere country-by-country decomposition of inequality is the decomposition of *between-country differences* in health inequalities into (a) the differences in inequality in the determinants of health, and (b) differences in the health effects of these determinants across countries. This is not straightforward because some differences (e.g. in the mean of x_k) may be offset by other differences (e.g. in C_k). These changes will, in general, not be independent of one another and the decomposition method has to take these interdependencies into account. One approach of dealing with this problem is to apply a decomposition method proposed in Wagstaff et al. (2003). If we denote by η_{ki} the elasticity of y with respect to x_{ki} for country i, and by C_{ki} the concentration index of determinant k in country i, then we can apply Oaxaca's (1973) method by choosing country i for reference values of η_{ki} and country j for values of C_{ki} to obtain:

$$\Delta C = C_i - C_j = \sum_k \eta_{ki}\left(C_{ki} - C_{kj}\right) + \sum_k C_{kj}\left(\eta_{ki} - \eta_{kj}\right), \qquad (4.10)$$

with the alternative being:

$$\Delta C = C_i - C_j = \sum_k \eta_{kj}\left(C_{ki} - C_{kj}\right) + \sum_k C_{ki}\left(\eta_{ki} - \eta_{kj}\right). \qquad (4.11)$$

Then ΔC_k, the contribution of any variable x_k to ΔC, equals the sum of two terms, i.e.

$$\Delta C_k = \eta_{ki}\left(C_{ki} - C_{kj}\right) + C_{kj}\left(\eta_{ki} - \eta_{kj}\right) \qquad (4.12)$$

Both terms consist of a product of an elasticity and a concentration index. Note that while equations (4.10) and (4.11) produce identical ΔC's and ΔC_k's, their decomposition

in equation (4.12) is not unique and may lead to slightly differing results depending on the choice of *i* or *j* as 'index country' for the estimates of η_{ki} and C_{ki}. To facilitate some assessment of the relative importance of the inequality versus the elasticity component of ΔC_k, we have computed, for each determinant, the *relative excess* elasticity compared to the Netherlands, i.e. $\left(\eta_{ki} - \eta_{kj}\right) / \eta_{kj}$ and the *relative excess* inequality, i.e. $\left(C_{ki} - C_{kj}\right) / C_{kj}$, for j = NL.

4.3 DATA AND VARIABLE DEFINITIONS

The data used in this paper were taken from the third wave (held in 1996) of the *European Community Household Panel* Users Database (Eurostat, 1999). The ECHP is a survey based on a standardised questionnaire that involves annual interviewing of a representative panel of households and individuals 16 years and older in each EU member state. It covers a wide range of topics including demographics, income, social transfers, health, housing, education, employment, etc.

Our health measure is the answer to the question "How is your health in general?" rated in 5 categories (Very Poor, Poor, Fair, Good to Very Good).[b] The relative frequencies of the European-wide 1996 ECHP responses were (2.4%, 7.9%, 24.3%, 42.2% and 23%). These are remarkably close to the relative frequencies for the response categories (Poor, Fair, Good, Very Good and Excellent) in the Canadian 1994 NPHS, which were (2.4%, 8.6%, 27%, 37.2% and 24.8%) despite the different wording. For the estimation of the interval regression we have used the HUI thresholds obtained in Van Doorslaer and Jones (1997): 0, 0.428, 0.756, 0.897, 0.947 and 1.

Our income measure (i.e. our ranking variable) is disposable (i.e. after-tax) household income per equivalent adult, using the modified OECD equivalence scale. This gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 4 in the household. Total household income includes all the net monetary income received by the household members during the reference year (which is 1995 for the 1996 wave). It includes income from work (employment and self-employment), private income (from investments and property and private transfers to the household), pensions and other direct social transfers received. No account has been taken of indirect social transfers (e.g. reimbursement of medical expenses), receipts in kind and imputed rent from owner-occupied accommodation. All incomes were converted from national monetary units into a common reference unit (the "purchasing power standard") (Eurostat, 1999).

Other health determinants included in the analysis are: (i) Education level, i.e. the highest level of general or higher education completed is available at three levels: recognised third level education (ISCED 5-7), second stage of secondary level of education (ISCED 3) and less than second stage of secondary education (ISCED 0-2)); (ii) Marital status

Table 4.1 Means of variables per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
HUI predicted	0.8851	0.9077	0.9040	0.9027	0.8947	0.8692	0.8972	0.9236	0.8668	0.9077	0.8822	0.8183	0.8983
Log Income	9.4491	9.4927	9.3840	9.4261	9.9107	9.3738	9.4004	9.0952	9.0058	8.7821	8.9365	8.6797	9.5084
M30-44	0.1375	0.1483	0.1619	0.1445	0.1576	0.1373	0.1339	0.1482	0.1205	0.1295	0.1310	0.1329	0.1404
M45-59	0.1232	0.1194	0.1265	0.1048	0.1176	0.1156	0.1078	0.1066	0.1090	0.1089	0.0975	0.1065	0.1090
M60-69	0.0662	0.0507	0.0512	0.0686	0.0690	0.0567	0.0596	0.0479	0.0718	0.0749	0.0621	0.0617	0.0514
M70+	0.0442	0.0559	0.0401	0.0473	0.0384	0.0510	0.0551	0.0433	0.0552	0.0608	0.0499	0.0473	0.0420
F16-29	0.1002	0.1198	0.1130	0.1168	0.1139	0.1076	0.1141	0.1379	0.1153	0.1157	0.1364	0.1213	0.1217
F30-44	0.1443	0.1373	0.1656	0.1501	0.1520	0.1497	0.1574	0.1584	0.1215	0.1346	0.1326	0.1395	0.1446
F45-59	0.1245	0.1207	0.1227	0.1160	0.1086	0.1166	0.1241	0.1054	0.1195	0.1208	0.1061	0.1192	0.1192
F60-69	0.0735	0.0525	0.0576	0.0790	0.0726	0.0733	0.0696	0.0515	0.0871	0.0765	0.0688	0.0760	0.0606
F70+	0.0923	0.0744	0.0567	0.0738	0.0628	0.0806	0.0858	0.0612	0.0787	0.0718	0.0764	0.0709	0.0821
Second Educ	0.4707	0.3568	0.5218	0.2651	0.2602	0.3046	0.3305	0.3051	0.2808	0.2508	0.1830	0.1145	0.5921
Higher Educ	0.1781	0.2723	0.1587	0.2302	0.1341	0.1624	0.2069	0.1073	0.0560	0.1602	0.1398	0.0447	0.0594
Part-time empl	0.0089	0.0027	0.0210	0.0137		0.0081	0.0123	0.0023	0.0036	0.0011	0.0060	0.0026	0.0031
Self-employed	0.0388	0.0399	0.0367	0.0556	0.0473	0.0518	0.0748	0.0914	0.0993	0.1836	0.0856	0.1511	0.0710
Student	0.0454	0.0800	0.0837	0.0836	0.0663	0.0793	0.0275	0.0730	0.0824	0.0546	0.0886	0.0788	0.0515
Unemployed	0.0424	0.0543	0.0707	0.0847	0.0174	0.0637	0.0353	0.0763	0.0646	0.0612	0.1138	0.0452	0.0331
Housework	0.1049	0.0150	0.2470	0.1159	0.2026	0.1069	0.1415	0.2616	0.1991	0.2055	0.2035	0.0822	0.1520
Econ inactive	0.0014	0.0113	0.0850	0.0214	0.0064	0.0123	0.0316	0.0285	0.0242	0.0135	0.0641	0.0504	0.0049
Retired	0.2782	0.2592	0.1616	0.2539	0.2261	0.2403	0.2298	0.0972	0.2507	0.2243	0.1418	0.2113	0.2159

Table 4.1 (continued) Means of variables per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
Divorced/sep	0.0562	0.0654	0.0460	0.0646	0.0413	0.0467	0.0764	0.0258	0.0144	0.0190	0.0143	0.0205	0.0443
Widowed	0.0924	0.0763	0.0604	0.0821	0.0773	0.0854	0.0934	0.0763	0.0981	0.0922	0.0800	0.0868	0.0867
Unmarried	0.1761	0.1887	0.1849	0.1972	0.2078	0.2190	0.1631	0.3145	0.2664	0.1998	0.2846	0.2417	0.2140
Born oth Euro		0.0133		0.0509	0.2804	0.0356	0.0159	0.0432	0.0056		0.0064	0.0066	0.0205
Born non-Euro		0.0258		0.0354	0.0337	0.0571	0.0401	0.0056	0.0103		0.0080	0.0206	0.0077
Region 2				0.5806		0.1760	0.0936	0.3094	0.0198	0.1967	0.1003	0.1736	0.2314
Region 3				0.3212		0.0667	0.0676		0.1270	0.3932	0.1260	0.3588	0.3482
Region 4						0.0799	0.0355		0.0624	0.0893	0.1329	0.0591	
Region 5						0.1288	0.3071		0.1054		0.2759	0.0364	
Region 6						0.1018	0.0853		0.0905		0.2096	0.0098	
Region 7						0.1179	0.0970		0.0332		0.0414	0.0206	
Region 8						0.1226	0.1128		0.1033				
Region 9							0.0422		0.1152				
Region 10							0.0848		0.0838				
Region 11							0.0323		0.0334				
N	8535	4952	8934	5656	1891	11166	6023	6663	16703	10549	15184	10520	7042
% of pop 60+	27.6%	23.4%	20.6%	26.9%	24.3%	26.2%	27.0%	20.4%	29.3%	28.4%	25.7%	25.6%	23.6%
% higher educ	64.9%	62.9%	68.0%	49.5%	39.4%	46.7%	53.7%	41.2%	33.7%	41.1%	32.3%	15.9%	65.2%

distinguishes between married, separated/divorced, widowed and unmarried (including co-habiting); (iii) Activity status includes full-time employed, part-time employed, self-employed, student, unemployed, retired, doing housework and 'other economically inactive'.[c] Region of residence uses the EU's NUTS 1 level (Nomenclature of Statistical Territorial Units), except for countries where such information was withheld for privacy reasons (NL, D) or because the country is too small (DK, L). Sample sizes before and after deletion of cases with missing observations, as well as the means for all variables are presented in Table 4.1. Most country's sample sizes are between 7000 and 11000 adults, but some are larger (Spain, Italy) and some are smaller (Denmark and Luxembourg). Cross-sectional sample weights at the individual level were applied in all analyses.

4.4 RESULTS

4.4.1 Measuring and decomposing inequality by country

The country means of the predicted health and explanatory variables presented in Table 4.1 provide an interesting basis for simple cross-country comparisons. The predicted HUI means (using the interval regressions from Table 4.2) show average health utility values above 0.9 for some of the richer countries such as Belgium, Denmark and the Netherlands, but also for poorer countries like Ireland and Greece. By far the lowest mean health is predicted for Portugal (0.82 only) but also France and Italy have relatively low mean health scores.

Luxembourg is the richest country in the EU with by far the highest disposable income per equivalent adult, but not the best health. Greek and the Irish adults, on the other hand, report better health than would be expected on the basis of their mean incomes. The countries' demographic structures, as presented by the age-sex dummies, do not differ dramatically but illustrate that Ireland and the Netherlands are relatively younger countries with just 20% of adults over 60, while in most other countries this age group represents almost a quarter or more of their population (see bottom of table). While cross-country differences in education are somewhat more difficult to compare due to inter-country differences in definitions of educational qualifications, the low proportion of Portuguese with secondary level education is striking. Among the activity status variables, most notable are the high percentages of self-employed in Greece and Portugal, the low percentage of retired in Ireland, and the high variability of proportions reporting to be doing housework or to be in the non-economically-active category.

The analysis of inter-individual variation in health requires the specification of a theoretical model explaining health behaviour. The models estimated below are not derived from a formal model of health production and investment but could be thought

Table 4.2 Health equations: interval regression coefficients per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
Constant	**0.8765**	**0.9380**	**0.8986**	**0.8685**	**0.7768**	**0.8510**	**0.8297**	**0.9014**	**0.8615**	**0.8682**	**0.8559**	**0.7731**	**0.8765**
Log Income	**0.0047**	0.0003	**0.0031**	**0.0072**	**0.0150**	**0.0077**	**0.0088**	**0.0054**	**0.0073**	**0.0112**	**0.0069**	**0.0156**	**0.0099**
M30-44	**-0.0162**	**-0.0179**	**-0.0123**	**-0.0186**	**-0.0164**	**-0.0357**	**-0.0154**	**-0.0085**	**-0.0277**	**-0.0181**	**-0.0174**	**-0.0341**	**-0.0185**
M45-59	**-0.0445**	**-0.0281**	**-0.0236**	**-0.0285**	**-0.0339**	**-0.0517**	**-0.0271**	**-0.0217**	**-0.0484**	**-0.0416**	**-0.0406**	**-0.0626**	**-0.0581**
M60-69	**-0.0164**	0.0029	-0.0121	**-0.0316**	-0.0055	**-0.0527**	-0.0114	**-0.0269**	**-0.0685**	**-0.0610**	**-0.0534**	**-0.0829**	**-0.0309**
M70+	**-0.0483**	-0.0015	-0.0129	**-0.0487**	**-0.0547**	**-0.0725**	-0.0049	**-0.0549**	**-0.1348**	**-0.1122**	**-0.0929**	**-0.1124**	**-0.0912**
F16-29	**-0.0057**	-0.0019	-0.0033	**-0.0105**	**-0.0102**	**-0.0083**	-0.0011	0.0006	**-0.0073**	0.0007	-0.0027	-0.0021	0.0032
F30-44	**-0.0114**	**-0.0189**	**-0.0133**	**-0.0193**	**-0.0217**	**-0.0277**	-0.0055	**-0.0103**	**-0.0259**	**-0.0144**	**-0.0183**	**-0.0409**	**-0.0107**
F45-59	**-0.0377**	**-0.0284**	**-0.0282**	**-0.0336**	**-0.0524**	**-0.0461**	**-0.0202**	**-0.0162**	**-0.0631**	**-0.0449**	**-0.0578**	**-0.0909**	**-0.0462**
F60-69	**-0.0237**	-0.0079	**-0.0219**	**-0.0500**	**-0.0479**	**-0.0475**	0.0004	**-0.0389**	**-0.1030**	**-0.0760**	**-0.0923**	**-0.1253**	**-0.0574**
F70+	**-0.0564**	-0.0322	**-0.0426**	**-0.0724**	**-0.0797**	**-0.0775**	**-0.0225**	**-0.0627**	**-0.1421**	**-0.1240**	**-0.1195**	**-0.1389**	**-0.1034**
Second Educ	**0.0073**	**0.0168**	**0.0081**	0.0049	**0.0199**	**0.0101**	**0.0102**	0.0040	**0.0155**	**0.0100**	**0.0107**	**0.0164**	**0.0176**
Higher Educ	**0.0154**	**0.0224**	**0.0115**	**0.0100**	**0.0206**	**0.0187**	**0.0215**	**0.0076**	**0.0207**	**0.0118**	**0.0117**	**0.0352**	**0.0210**
Part-time em	0.0037	-0.0344	**-0.0152**	**-0.0958**		-0.0058	-0.0022	-0.0014	-0.0006	-0.0065	-0.0046	-0.1045	0.0085
Self-employe	0.0001	-0.0028	-0.0005	**0.0085**	0.0022	-0.0040	-0.0041	-0.0008	**0.0050**	**0.0042**	0.0008	-0.0020	**-0.0104**
Student	**0.0103**	**0.0082**	**0.0064**	0.0059	**0.0176**	**0.0121**	-0.0066	0.0039	**0.0144**	**0.0097**	**0.0074**	0.0077	**0.0182**
Unemployed	**-0.0373**	**-0.0212**	**-0.0245**	**-0.0232**	**-0.1039**	**-0.0136**	**-0.0172**	**-0.0106**	-0.0042	-0.0012	0.0019	**-0.0185**	**-0.0281**
Housework	**-0.0118**	**-0.0234**	**-0.0174**	-0.0077	0.0038	**-0.0608**	**-0.0351**	**-0.0153**	**-0.0086**	**-0.0170**	**-0.0084**	**-0.0412**	**-0.0200**
Econ inactive	-0.0207	**-0.1144**	**-0.0616**	**-0.0792**	**-0.0783**	-0.0029	**-0.1967**	**-0.1466**	**-0.0964**	**-0.2080**	**-0.1158**	**-0.1814**	**-0.2390**
Retired	**-0.0557**	**-0.1014**	-0.0073	**-0.0162**	**-0.0544**	**-0.0372**	**-0.0526**	**-0.0205**	**-0.0264**	**-0.0515**	**-0.0341**	**-0.1175**	**-0.0484**

Note: Regression coefficients which differ significantly from zero (at P<0.05) are in bold typeface. Constant term not reported.

Table 4.2 (continued) Health equations: interval regression coefficients per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
Divorced/sep	**-0.0160**	**-0.0169**	**-0.0165**	-0.0100	-0.0215	**-0.0231**	-0.0068	-0.0117	-0.0014	-0.0206	-0.0091	-0.0039	-0.0119
Widowed	0.0044	-0.0015	-0.0076	-0.0046	**0.0228**	-0.0113	-0.0132	-0.0091	-0.0101	**-0.0189**	0.0109	0.0134	0.0020
Unmarried	**0.0067**	-0.0008	-0.0015	-0.0001	**0.0145**	**-0.0138**	-0.0010	-0.0022	-0.0011	**-0.0065**	-0.0019	**-0.0132**	-0.0050
Born oth Euro		0.0067		-0.0058	-0.0065	0.0012	-0.0157	-0.0033	-0.0029		**0.0193**	0.0080	0.0041
Born non-Euro		-0.0131		-0.0083	**-0.0195**	-0.0085	-0.0123	-0.0302	-0.0030		0.0007	0.0069	-0.0100
Region 2				0.0064		-0.0033	0.0075	0.0010	0.0020	0.0002	**0.0221**	0.0013	-0.0012
Region 3				**-0.0085**		-0.0010	0.0129		0.0058	**-0.0057**	**0.0101**	0.0006	**0.0067**
Region 4						0.0005	**0.0247**		-0.0021	**-0.0112**	**0.0089**	**0.0172**	
Region 5						0.0003	**0.0201**		-0.0045		**0.0175**	**0.0228**	
Region 6						0.0001	**0.0222**		-0.0014		**0.0103**	**0.0286**	
Region 7						0.0022	**0.0154**		-0.0033		**0.0097**	**0.0569**	
Region 8						-0.0077	0.0132		-0.0063				
Region 9							-0.0056		-0.0147				
Region 10							**0.0194**		-0.0013				
Region 11							0.0150		-0.0108				
Adj R^2	0.142	0.180	0.119	0.159	0.159	0.112	0.126	0.171	0.218	0.298	0.263	0.330	0.236

Note: Regression coefficients which differ significantly from zero (at P<0.05) are in bold typeface. Constant term not reported.

of as reduced-form estimates of a static model of the demand for health. By definition, these reduced form estimates do not provide any guidance as to *how* individual choice behaviour affects health but they do allow for the estimation of the impact of partial changes in exogenous health determinants, some of which may be amenable to policy intervention.

We estimated an interval regression equation per country using the Canadian HUI thresholds for the SAH variable and including the following explanatory variables: (i) the logarithm of equivalent disposable household income (to capture well-known concavity in this relationship), (ii) ten age-sex categories indicating the age groups 16-29, 30-44, 45-59, 60-69 and 70+ for both males and females, (iii) three education level categories, (iv) eight categories of activity status, (v) four categories of marital status, and, where available, (vi) three categories of country of birth (dummies for other European and other non-European country) and (vii) where relevant and available, the region of residence. In the case of the dummies, one reference category was omitted, of course. The profile of the omitted reference category was the young, highly educated, employed, married, male with average income, born in the country in question and (if applicable) living in the omitted region (usually the capital region).

The results of the interval regressions are presented in Table 4.2. No causal interpretation can be given to the coefficient estimates because they are purely cross-sectional and because the exogenous status of some of the explanatory variables (in particular household income and activity status) is contentious. Some general observations emerge. First, the (logarithm of) household income has a significantly positive coefficient in all countries except one: only in Denmark no significant partial association between health and income remains when these other factors are controlled for. The estimated coefficients are directly comparable because both income and health are measured in the same units in all countries. In general, and as expected, the income elasticity of health is lower in countries with higher income (Luxembourg is the exception). Secondly, the demographic effects are largely as expected: lower health for females and decreasing health with rising age. However, there seems to be sufficient variation across countries for homogeneity of (partial) age-gender effects across Europe to be rejected. Third, without any exception, higher levels of education are significantly associated with better health in all countries. Countries with the lowest 'health return' to education are Ireland and Belgium. Fourth, those who are married or cohabiting generally report better health than those who are not (or not any more), but there is substantial variation across countries in this pattern. Fifth, there is little or no clear health effect of being born in another country, European or other. Sixth, not surprisingly, there is important variation in reported health by activity status. Being unemployed, retired or otherwise not economically active (i.e. because of disability status) is associated with significantly

lower health, but generally – and more surprisingly – the same is true for those doing housework. Especially the partial association between retirement and other non-active status (and the difference between these two probably differs across countries) and health turns out to be strong. Notable exceptions include the Netherlands (for retirement status) and France and Germany (for inactivity status). Finally, for a number of countries it was possible to include dummies indicating region of residence to allow for regional variation in health status. The names of the corresponding regions are given in Tabl4.5. In some countries – France in particular – none of the region dummies is significant. In some other European countries, however, significant partial regional health effects emerge. Regions with, *ceteris paribus*, relatively worse health appear to be: the south of Belgium (Wallonia), the North West of the UK (including Northern England, Yorkshire and Humberside, East Midlands, the North West, Northern Ireland and Wales), the South of Italy, the non-Athens regions in Greece, the North-West of Spain, the North and the Lisbon area in Portugal and the South East of Austria.

The concentration indices, computed using equation (4.5) and presented in Table 4.3, tell an interesting story about income-related inequalities in themselves. First of all, in *all* EU member states self-reported health is unequally distributed in favour of the higher income groups. But there are substantial differences between countries. Portugal has far greater income-related health inequality than the UK and Greece. The Netherlands and Germany have very low health inequality, but also Italy, Belgium and Austria have relatively low CIs. To some degree, the pattern of income inequality (as measured by the CI of log income) follows the pattern of income-related health inequality, but exceptions are Denmark and Luxembourg, both of which show higher health inequality than expected on the basis of their income inequality. In terms of age groups, it is striking to see older women concentrated in lower income groups everywhere, but least of all in Spain and Italy. By contrast, older males tend to rank low in income only in Denmark, the UK, Ireland, Greece and Portugal. In all countries, higher educated are strongly concentrated amongst the richest, but most of all in Portugal and least of all in Germany. Retired are more concentrated among the lower income groups and the same is true for unemployed in all countries but Denmark. Immigrants seem worse off in most countries, but not in Austria and Portugal. The concentration indices of the regional dummies also clearly show which are the relatively wealthy and which are the less well-off regions. Regional income disparities are particularly pronounced in Italy, Spain and Portugal.

We now turn to the more important question: how do these various characteristics contribute – or not – to the observed income-related health inequality in European countries? It is clear that a determinant's contribution to inequality can be either positive or negative, depending on the sign of its health effect and its distribution by income,

Table 4.3 Concentration indices of dependent and independent variables per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
	DE	DK	NL	BE	LU	FR	UK	IE	IT	GR	ES	PT	AT
HUI predicted	0.0043	0.0094	0.0034	0.0071	0.0104	0.0075	0.0129	0.0077	0.0063	0.0119	0.0066	0.0218	0.0073
Log Income	0.0318	0.0262	0.0332	0.0321	0.0285	0.0324	0.0360	0.0365	0.0403	0.0428	0.0397	0.0465	0.0291
M30-44	0.0862	0.1463	0.0374	0.1142	0.0868	0.0399	0.1324	0.1087	0.1091	0.1208	0.0803	0.0529	0.0500
M45-59	0.1468	0.2236	0.1503	0.0958	0.0396	0.1409	0.1968	0.0650	0.0352	0.0751	0.0117	0.0920	0.1354
M60-69	-0.0123	-0.1448	0.0395	-0.0607	0.0074	0.0288	-0.0041	-0.0123	0.0461	-0.0322	-0.0219	-0.0652	-0.0032
M70+	-0.0032	-0.3305	-0.0571	-0.1593	-0.0665	-0.0263	-0.2701	-0.1819	-0.0176	-0.2425	-0.0367	-0.2487	-0.1245
F16-29	-0.1054	-0.1044	-0.0996	-0.0241	-0.0102	-0.1182	-0.0523	0.0112	-0.0992	-0.0295	-0.0395	0.0554	-0.0431
F30-44	-0.0127	0.1400	-0.0291	0.0236	-0.0138	-0.0073	0.0231	0.0020	0.0264	0.1045	0.0553	0.0274	-0.0226
F45-59	0.0952	0.1580	0.1150	0.0584	0.0112	0.1387	0.1276	0.0505	0.0417	0.0517	-0.0075	0.0491	0.1246
F60-69	-0.1001	-0.2515	-0.0503	-0.0915	-0.0438	-0.0222	-0.1654	-0.1170	-0.0344	-0.1029	-0.0126	-0.1319	-0.1303
F70+	-0.1790	-0.3281	-0.1795	-0.2213	-0.1641	-0.1906	-0.3600	-0.3274	-0.0650	-0.2468	-0.0988	-0.2730	-0.2404
Second Educ	0.0000	-0.0136	-0.0045	0.0223	0.1502	0.0569	0.0556	0.1733	0.1601	0.1555	0.1452	0.3192	0.0822
Higher Educ	0.2176	0.2521	0.3676	0.3034	0.4951	0.4503	0.3688	0.5503	0.4495	0.4210	0.4315	0.7940	0.3409
Part-time empl	-0.0141	-0.0929	0.0745	-0.2190		-0.0875	0.0443	0.2194	-0.0223	0.0050	-0.2097	-0.1368	0.1111
Self-employed	0.2131	0.1306	0.1363	-0.0567	0.0845	0.0912	0.3071	0.1861	-0.0503	0.0335	-0.0508	-0.1630	-0.0662
Student	-0.2023	-0.2217	-0.3104	-0.0633	-0.1434	-0.0851	-0.1997	-0.0607	-0.0889	-0.0121	-0.0033	0.1052	-0.0973
Unemployed	-0.2161	-0.0017	-0.2931	-0.2777	-0.4209	-0.3212	-0.3160	-0.3428	-0.3674	-0.1989	-0.2326	-0.1007	-0.1153
Housework	-0.0953	-0.3005	-0.1241	-0.1875	-0.1066	-0.2641	-0.2672	-0.2357	-0.1964	-0.0850	-0.1462	-0.1853	-0.2455
Econ inactive	-0.2455	-0.0674	-0.1079	-0.3104	-0.2132	-0.0503	-0.3196	-0.3116	-0.2024	-0.0808	-0.1140	-0.2408	-0.0222
Retired	-0.1053	-0.2941	-0.0980	-0.1058	-0.0959	-0.0386	-0.2146	-0.1249	0.0397	-0.1317	-0.0041	-0.1431	-0.0476

Table 4.3 (continued) Concentration indices of dependent and independent variables per country

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
	DE	DK	NL	BE	LU	FR	UK	IE	IT	GR	ES	PT	AT
Divorced/sep	-0.1887	-0.1828	-0.1758	-0.1603	-0.0851	-0.0869	-0.2990	-0.4044	0.1643	0.0412	-0.0720	-0.1396	-0.0470
Widowed	-0.1447	-0.2974	-0.1029	-0.1805	-0.0464	-0.2149	-0.3706	-0.2723	-0.0706	-0.1169	-0.0616	-0.1804	-0.1999
Unmarried	-0.0359	-0.1685	-0.1238	-0.0074	0.0437	-0.0781	-0.0042	-0.0166	-0.0299	0.0340	0.0022	0.0358	0.0120
Born oth Euro		-0.1354		-0.0508	-0.0670	-0.1755	0.1191	-0.0288	0.0437		-0.0823	0.0693	0.1077
Born non-Euro		-0.2421		-0.1862	-0.1661	-0.1820	-0.0516	-0.0383	-0.0188		-0.0431	0.2101	0.2047
Region 2				0.0187		-0.0561	-0.1116	0.1015	0.2658	-0.0938	0.1402	-0.1613	-0.0575
Region 3				-0.0405		-0.1804	-0.0178		0.1976	0.1549	0.2890	0.1733	-0.0468
Region 4						-0.0099	-0.1228		0.2298	-0.0335	-0.1540	-0.1376	
Region 5						-0.0837	0.1514		0.1039		0.0995	-0.1770	
Region 6						-0.1240	0.0242		0.0704		-0.1986	-0.2608	
Region 7						0.0150	-0.0376		-0.0574		-0.2253	-0.2993	
Region 8						-0.0625	-0.0284		-0.2591				
Region 9							-0.2104		-0.2907				
Region 10							-0.0960		-0.3017				
Region 11							-0.1335		-0.3322				

Table 4.4 Region dummies by country

	Belgium	France	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
Region 1	Brussels	Île de France	North	Non-Dublin	Nord Ovest	Voreia Ellada	Noroeste	Norte	Ost
Region 2	Flanders	Bassin Parisien	Yorkshire / Humberside	Dublin	Lombardia	Kentriki Ellada	Noreste	Centro (P)	Süd
Region 3	Wallonia	Nord-Pas-de-Calais	East Midlands		Nord Est	Attiki	Comunidad de Madrid	Lisboa e Vale do Tejo	West
Region 4		Est	East Anglia		Emilia-Romagna	Nisia Aigaiou, Kriti	Centro (E)	Alentejo	
Region 5		Ouest	South East		Centro (I)		Este	Algarve	
Region 6		Sud-Ouest	South West		Lazio		Sur	Açores	
Region 7		Centre-Est	West Midlands		Abruzzo-Molise		Canarias	Madeira	
Region 8		Méditerranée	North West		Campania				
Region 9			Wales		Sud				
Region 10			Scotland		Sicilia				
Region 11			Northern Ireland		Sardegna				

Source: EUROSTAT, User's Database Documentation

as reflected in its concentration index. Table 4.5 summarizes the decomposition of each country's income-related health inequality into the percentage contributions of the regressor variables. A positive (negative) x% contribution of variable X is to be interpreted as follows: income-related health inequality would, *ceteris paribus*, be x% lower if variable X were equally distributed across the income range, or if variable X had a zero health elasticity. The second row presents estimates of $I^* = C - C^*$, defined as the health inequality *not* due to demographics, i.e. the degree of inequality which would be observed if age and sex were equally distributed by income, or had no effect on health. It is computed by subtracting the contributions of age and gender from total inequality and it is comparable to the degree of potentially avoidable inequality as it was labelled in past work using the indirect method of standardisation (e.g. Van Doorslaer et al., 1997). The important difference is, however, that the age-sex contributions here are now estimated as *partial* effects, i.e. while controlling for income and other influences on health. We can see that I^* is often but not always smaller than C. In other words: the health and income distribution of age and gender can both increase and decrease observed health inequality.

Figure 4.1 shows the countries' ranking by I^* with the confidence intervals obtained from the bootstrapped standard errors. In all countries, the degree of non-age-sex related - and therefore potentially avoidable - inequality I^* is highly significant and different from

Figure 4.1 Degree of income-related inequality in health utility, by country, with 95% confidence intervals (ECHP, 1996, age-sex standardised)

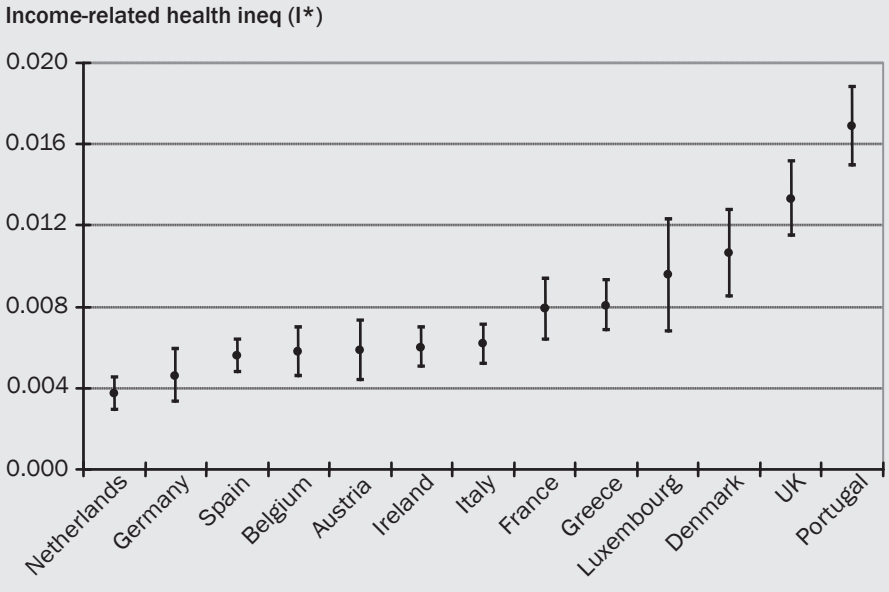
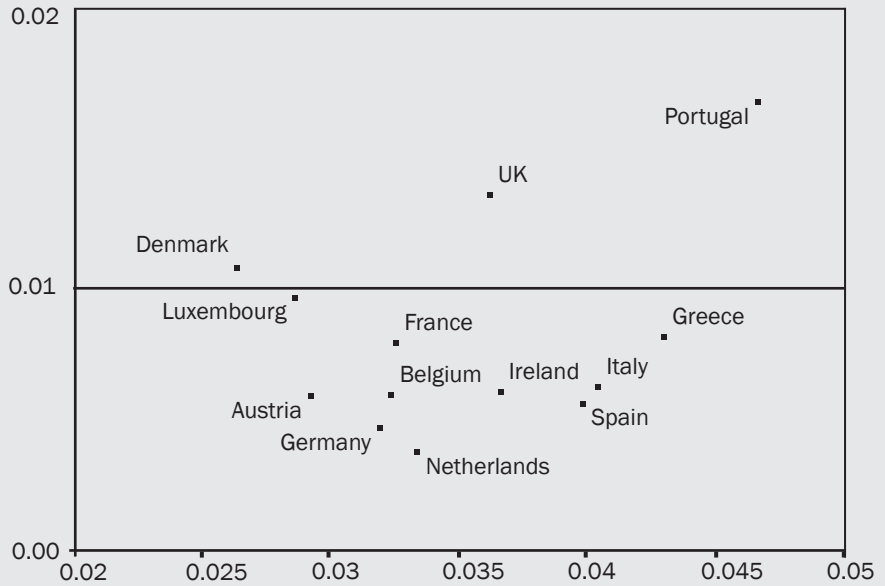


Figure 4.2 Income inequality and health inequality (r(G,c)=0.64; r(G,I*)=0.47)



zero. But there are large differences between countries. In the Netherlands, inequality is significantly lower than in any other country except Germany. In Portugal, on the other hand, inequality is significantly greater than in any other country. Table 4.A.1 summarizes the results of all pair wise comparisons by showing the t statistics of the between-country differences in I* values.

Figure 4.2 shows a simple scatter diagram of health inequality (as measured by *I**) and income inequality (as measured by the Gini coefficient of income). The correlation is positive, but weak and not significant (r=0.47; P=0.10). The correlation with unstandardised inequality *C* is higher and significant (r=0.66; P=0.019) but also much lower than the r=0.87 reported in Van Doorslaer et al. (1997). The positive correlation is mainly due to the outlier position of Portugal, a role played by the US in Van Doorslaer et al. (1997). Especially Denmark and the UK show a much higher health inequality than would be expected on the basis their income inequality. The decompositions will allow us to explore these findings in greater detail.

Some general findings emerge from the decomposition results presented in Table 4.5. First, in all countries except Denmark, income itself accounts for a significant and sizeable contribution: between 25% and 40% of all measured inequality. The peculiar Danish result does not arise because Denmark has the lowest income inequality in the EU, but because of the complete lack of any (partial) linear association between household income and adult health. Apart from income itself, age, education and activity

status are the most important 'contributors' to health inequality. While older females contribute significantly to higher health inequality, middle aged males - with worse health but better incomes than young males - reduce health inequality. Higher educational differences invariably and significantly contribute to inequality. The contributions of the activity status variables are less straightforward to generalise. This may be related to the variation in social security schemes across European countries leading to differences in meaning and uses of this categorization. In some countries (e.g. Denmark, Germany), it is mainly the (health and income distribution of the) retired which contributes to inequality, while in others (Netherlands, UK, Ireland and Spain), the 'other economically inactive' group has a greater contribution. In France the largest contribution comes from those reporting to be doing housework.[d] The important thing to note for a proper interpretation is that all of these contributions are *partial*, i.e. after having controlled for demographics and income. This means that, for instance, the retired report worse health than others of the same age and income. It must therefore mainly reflect the disadvantaged position of the *early* retired. In the next section we will further zoom in on the differences between the countries.

Marital status and immigration status generally show only minor contributions to health inequality. The contribution of region is mixed and again depends also on the regional detail available and the choice of the reference region. In general, the estimates show that region of residence does matter in the Mediterranean countries, especially in Spain and Italy, and to a lesser extent in Portugal and Greece. There is also some regional contribution in the UK but - perhaps surprisingly - none at all in France.[d]

4.4.2 Decomposing "excess inequality"

The next interesting question is why some countries appear to be able to enjoy much lower health inequality than others. We have further explored the factors driving these differences between countries using the Oaxaca-type decomposition method presented in equations (4.10) and (4.12). In principle, with 13 countries and two index options, there are 24 possible ways of decomposing the inequality differences. With the information in Tables 1-5, the reader can reproduce all of these possible comparisons. In Table 4.6 we present the ΔC_k estimates for just one of these options, with all countries compared to the country with the lowest inequality, i.e. the Netherlands, as the 'index' country. In effect, this shows a decomposition of the 'excess inequality' compared to the Netherlands and the (column) percentage contributions presented here are comparable to the ones presented in Table 4.5. Generally, the same variables as in Table 4.5 (i.e. income itself, female elderly, retirement and other non-active status) account for most of the differences, although the relative shares differ because of the relatively favourable position of older females and retired in the Netherlands. Note that virtually

Table 4.5 Health inequality contributions of regressors per country (in % of HUI concentration index, and with bootstrapped t-values)

	Germany		Denmark		Netherlands		Belgium	
	CI contr	t-val	CI contr	t-val	CI contr	t-val	CI contr	t-val
C HUI pred	0.00434		0.00938		0.00337		0.00710	
I* = C - C*	0.00461	7.05	0.01062	9.85	0.00372	9.22	0.00579	9.37
Ln (Inc)	36.7%	2.68	0.9%	0.10	31.6%	2.74	33.9%	4.49
M30-44	-5.0%	-3.90	-4.6%	-4.29	-2.4%	-2.42	-4.8%	-3.82
M45-59	-21.0%	-7.10	-8.8%	-5.47	-14.8%	-6.39	-4.5%	-3.50
M60-69	0.3%	0.47	-0.3%	-0.29	-0.8%	-1.13	2.1%	1.79
M70+	0.2%	0.12	0.3%	0.13	1.0%	1.21	5.7%	3.85
F16-29	1.6%	2.15	0.3%	0.56	1.2%	1.39	0.5%	1.01
F30-44	0.5%	0.82	-4.3%	-4.09	2.1%	1.94	-1.1%	-1.22
F45-59	-11.6%	-5.18	-6.4%	-4.66	-13.1%	-5.52	-3.5%	-2.42
F60-69	4.5%	2.84	1.2%	0.72	2.1%	1.81	5.6%	3.10
F70+	24.3%	5.42	9.2%	2.35	14.3%	4.35	18.4%	5.73
Second Educ	0.0%	0.00	-1.0%	-1.12	-0.6%	-0.74	0.5%	1.27
Higher Educ	15.6%	4.67	18.1%	5.81	22.1%	4.89	10.9%	3.99
Part-time empl	0.0%	-0.11	0.1%	0.34	-0.8%	-1.59	4.5%	2.33
Self-employed	0.0%	0.02	-0.2%	-0.55	-0.1%	-0.14	-0.4%	-1.20
Student	-2.5%	-3.01	-1.7%	-2.46	-5.5%	-2.65	-0.5%	-1.28
Unemployed	8.9%	4.51	0.0%	0.05	16.7%	5.37	8.5%	4.78
Housework	3.1%	3.01	1.2%	2.08	17.6%	5.92	2.6%	1.88
Econ inactive	0.2%	0.64	1.0%	0.92	18.6%	4.64	8.2%	4.23
Retired	42.5%	6.86	90.8%	10.54	3.8%	1.47	6.8%	3.30
Divorced/sep	4.4%	2.52	2.4%	2.47	4.4%	2.88	1.6%	1.95
Widowed	-1.5%	-0.73	0.4%	0.15	1.6%	1.53	1.1%	0.71
Unmarried	-1.1%	-1.78	0.3%	0.23	1.1%	0.69	0.0%	0.01
Born oth Euro			-0.1%	-0.62			0.2%	0.74
Born non-Euro			1.0%	1.42			0.9%	1.08
Region 2							1.1%	1.43
Region 3							1.7%	1.75
Region 4								
Region 5								
Region 6								
Region 7								
Region 8								
Region 9								
Region 10								
Region 11								

Table 4.5 (continued) Health inequality contributions of regressors per country (in % of HUI conc index, and with bootstrapped t-values)

	Luxembourg		France		UK		Ireland	
	CI contr	t-val	CI contr	t-val	CI contr	t-val	CI contr	t-val
C HUI pred	0.01036		0.00745		0.01286		0.00769	
I* = C - C*	0.00955	6.80	0.00788	10.25	0.01332	14.39	0.00600	12.49
Ln (Inc)	45.8%	3.80	36.1%	3.30	25.8%	4.44	25.1%	4.13
M30-44	-2.4%	-1.87	-3.0%	-2.94	-2.4%	-3.42	-1.9%	-2.48
M45-59	-1.7%	-0.97	-13.0%	-7.50	-5.0%	-5.05	-2.1%	-2.86
M60-69	0.0%	-0.07	-1.3%	-1.26	0.0%	0.12	0.2%	0.41
M70+	1.5%	0.91	1.5%	1.06	0.6%	0.54	6.1%	4.31
F16-29	0.1%	0.24	1.6%	2.74	0.1%	0.27	0.0%	0.17
F30-44	0.5%	0.40	0.5%	0.58	-0.2%	-0.86	0.0%	-0.11
F45-59	-0.7%	-0.28	-11.5%	-6.71	-2.8%	-3.80	-1.2%	-2.11
F60-69	1.6%	0.95	1.2%	1.11	0.0%	-0.05	3.3%	3.31
F70+	8.9%	2.66	18.4%	6.64	6.0%	2.60	17.7%	6.99
Second Educ	8.4%	3.97	2.7%	3.22	1.6%	2.98	3.0%	2.23
Higher Educ	14.7%	3.31	21.1%	6.01	14.2%	7.30	6.3%	3.14
Part-time empl			0.1%	0.47	0.0%	-0.19	0.0%	-0.08
Self-employed	0.1%	0.21	-0.3%	-0.90	-0.8%	-1.25	-0.2%	-0.33
Student	-1.8%	-1.88	-1.3%	-2.40	0.3%	1.29	-0.2%	-1.33
Unemployed	8.2%	2.21	4.3%	2.86	1.7%	2.56	3.9%	3.81
Housework	-0.9%	-0.56	26.5%	8.81	11.5%	7.01	13.2%	5.21
Econ inactive	1.2%	1.37	0.0%	0.29	17.2%	6.10	18.3%	6.42
Retired	12.7%	3.01	5.3%	3.02	22.5%	7.26	3.5%	3.14
Divorced/sep	0.8%	0.78	1.4%	2.29	1.3%	1.32	1.7%	1.75
Widowed	-0.9%	-0.90	3.2%	1.77	4.0%	1.92	2.7%	1.64
Unmarried	1.4%	1.55	3.6%	3.20	0.0%	0.07	0.2%	0.68
Born oth Euro		1.27	-0.1%	-0.19	-0.3%	-1.40	0.1%	0.48
Born non-Euro		1.26	1.4%	1.53	0.2%	0.90	0.1%	0.33
Region 2			0.5%	0.96	-0.7%	-0.94	0.5%	0.62
Region 3			0.2%	0.22	-0.1%	-0.55		
Region 4			0.0%	-0.05	-0.9%	-2.04		
Region 5			-0.1%	-0.08	8.1%	2.97		
Region 6			0.0%	-0.02	0.4%	0.91		
Region 7			0.1%	0.38	-0.5%	-1.25		
Region 8			0.9%	1.62	-0.4%	-0.97		
Region 9					0.4%	0.54		
Region 10					-1.4%	-2.18		
Region 11					-0.6%	-1.44		

Table 4.5 (continued) Health inequality contributions of regressors per country (in % of HUI conc index, and with bootstrapped t-values)

	Italy		Greece		Spain		Portugal		Austria	
	CI contr	t-val	CI contr	t-val	CI contr	t-val	CI contr	t-val	CI contr	t-val
C HUI pred	0.00626		0.01190		0.00663		0.02180		0.00730	
I* = C - C*	0.0062	12.51	0.0081	12.72	0.0056	13.37	0.0169	17.07	0.0058	7.91
Ln (Inc)	48.9%	5.63	39.1%	7.61	42.1%	6.58	35.4%	8.66	41.9%	4.05
M30-44	-6.7%	-6.57	-2.6%	-5.09	-3.1%	-4.76	-1.3%	-3.45	-2.0%	-2.67
M45-59	-3.4%	-2.64	-3.2%	-4.36	-0.8%	-0.82	-3.4%	-4.98	-13.1%	-6.29
M60-69	-4.2%	-2.86	1.4%	1.61	1.2%	1.24	1.9%	2.64	0.1%	0.11
M70+	2.4%	1.02	15.3%	7.79	2.9%	2.00	7.4%	6.83	7.3%	3.18
F16-29	1.5%	3.67	0.0%	-0.31	0.3%	1.58	-0.1%	-0.69	-0.3%	-1.21
F30-44	-1.5%	-2.07	-1.9%	-3.98	-2.3%	-3.61	-0.9%	-1.96	0.5%	1.27
F45-59	-5.8%	-3.36	-2.6%	-3.23	0.8%	0.54	-3.0%	-3.00	-10.5%	-5.65
F60-69	5.7%	2.44	5.5%	4.60	1.4%	0.77	7.0%	5.88	6.9%	3.90
F70+	13.4%	4.44	20.3%	9.08	15.4%	6.23	15.1%	9.19	31.1%	7.64
Second Educ	12.8%	10.06	3.6%	4.62	4.9%	6.18	3.3%	5.75	13.0%	6.38
Higher Educ	9.6%	7.10	7.4%	5.07	12.1%	6.24	7.0%	8.84	6.5%	4.29
Part-time empl	0.0%	0.03	0.0%	-0.01	0.1%	0.58	0.2%	1.47	0.0%	0.45
Self-employed	-0.5%	-1.88	0.2%	1.88	-0.1%	-0.36	0.3%	0.66	0.7%	1.76
Student	-1.9%	-4.36	-0.1%	-0.49	0.0%	-0.21	0.4%	2.34	-1.4%	-2.88
Unemployed	1.8%	1.55	0.1%	0.44	-0.8%	-1.03	0.5%	3.05	1.6%	2.53
Housework	6.2%	3.21	2.7%	4.74	4.3%	2.99	3.5%	5.64	11.4%	4.89
Econ inactive	8.7%	5.56	2.1%	1.59	14.5%	6.38	12.3%	7.73	0.4%	0.21
Retired	-4.8%	-4.40	14.1%	7.13	0.3%	0.38	19.9%	10.20	7.6%	3.18
Divorced/sep	-0.1%	-0.28	-0.1%	-0.83	0.2%	0.94	0.1%	0.47	0.4%	1.03
Widowed	1.3%	2.09	1.9%	2.97	-0.9%	-1.89	-1.2%	-1.79	-0.5%	-0.29
Unmarried	0.2%	0.44	-0.4%	-1.70	0.0%	-0.18	-0.6%	-2.49	-0.2%	-0.71
Born oth Euro	0.0%				-0.2%	-1.14	0.0%	0.52	0.1%	0.51
Born non-Euro	0.0%				0.0%	-0.05	0.2%	1.01	-0.2%	-0.84
Region 2	0.2%	0.37	0.0%	-0.08	5.3%	5.79	-0.2%	-0.36	0.2%	0.42
Region 3	2.7%	2.29	-3.2%	-2.79	6.3%	3.45	0.2%	0.23	-1.7%	-2.41
Region 4	-0.6%	-0.58	0.3%	1.49	-3.1%	-2.65	-0.8%	-2.92		
Region 5	-0.9%	-1.56			8.2%	5.74	-0.8%	-2.81		
Region 6	-0.2%	-0.44			-7.3%	-3.42	-0.4%	-2.09		
Region 7	0.1%	0.74			-1.6%	-2.12	-2.0%	-5.36		
Region 8	3.1%	2.34								
Region 9	9.1%	4.70								
Region 10	0.6%	0.43								
Region 11	2.2%	2.23								

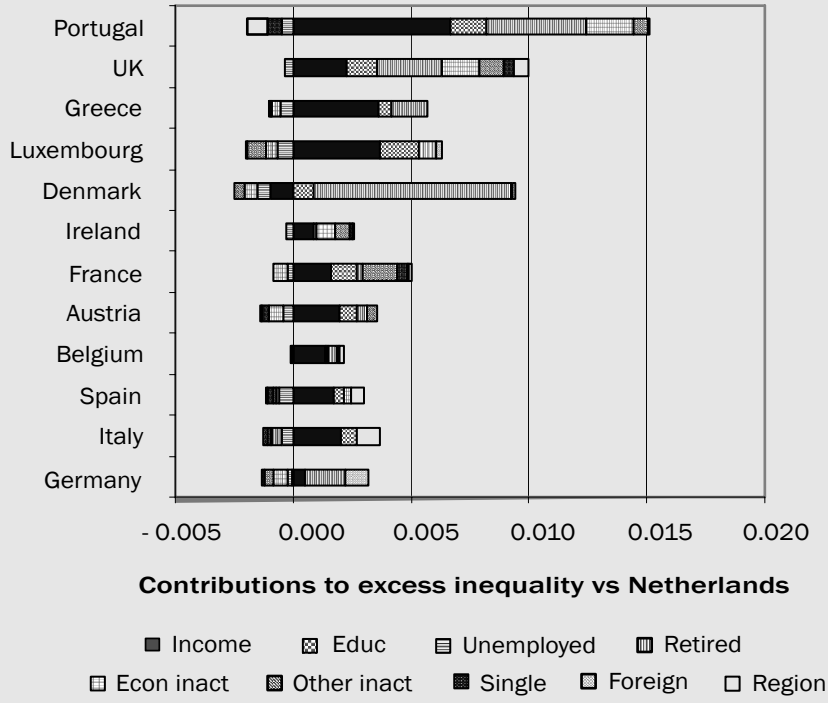
Table 4.6 Contributions of regressors to excess health inequality per country versus Netherlands (in % of excess concentration index of HUI in first row)

	Germany	Denmark	Netherl	Belgium	Luxemb	France	UK
Excess ineq							
CI - CI $_{NL}$	0.00097	0.00602		0.00373	0.00700	0.00409	0.00949
Ln (Inc)	54.3%	-16.2%		36.0%	52.6%	39.8%	23.7%
M30-44	-13.8%	-5.8%		-6.9%	-2.4%	-3.5%	-2.3%
M45-59	-42.4%	-5.5%		4.8%	4.6%	-11.6%	-1.5%
M60-69	4.3%	0.1%		4.6%	0.3%	-1.8%	0.3%
M70+	-2.6%	0.0%		10.0%	1.8%	1.9%	0.5%
F16-29	2.7%	-0.3%		-0.2%	-0.4%	1.9%	-0.4%
F30-44	-4.9%	-7.8%		-3.9%	-0.3%	-0.9%	-1.0%
F45-59	-6.7%	-2.6%		5.0%	5.3%	-10.2%	0.9%
F60-69	13.0%	0.7%		8.8%	1.4%	0.4%	-0.8%
F70+	58.8%	6.4%		22.2%	6.3%	21.8%	3.1%
Second Educ	2.2%	-1.1%		1.4%	12.7%	5.5%	2.4%
Higher Educ	-7.1%	15.8%		0.7%	11.2%	20.3%	11.4%
Part-time empl	2.6%	0.6%		9.2%		0.8%	0.3%
Self-employed	0.3%	-0.2%		-0.7%	0.5%	-0.5%	-1.1%
Student	8.1%	0.4%		4.0%	-2.6%	2.2%	2.4%
Unemployed	-18.1%	-9.3%		1.2%	14.8%	-5.9%	-3.7%
Housework	-47.1%	-7.9%		-10.8%	-9.3%	33.9%	9.3%
Econ inactive	-63.4%	-8.8%		-1.1%	-6.7%	-15.2%	16.7%
Retired	176.1%	139.4%		9.5%	9.9%	6.6%	29.1%
Divorced/sep	4.6%	1.2%		-0.9%	-0.9%	-1.0%	0.3%
Widowed	-12.3%	-0.2%		0.6%	-2.1%	4.5%	4.8%
Unmarried	-8.9%	-0.2%		-1.0%	1.6%	5.7%	-0.4%
Born other Euro		-0.2%		0.4%	2.0%	-0.2%	-0.3%
Born non-Euro		1.5%		1.6%	1.7%	2.5%	0.3%
Region 2				2.1%		0.9%	-0.9%
Region 3				3.3%		0.3%	-0.2%
Region 4						0.0%	-1.3%
Region 5						-0.1%	11.0%
Region 6						0.0%	0.5%
Region 7						0.1%	-0.7%
Region 8						1.7%	-0.5%
Region 9							0.6%
Region 10							-1.9%
Region 11							-0.8%

Table 4.6 (continued) Contributions of regressors to excess health inequality per country versus Netherlands (in % of excess concentration index of HUI in first row)

	Netherl	Ireland	Italy	Greece	Spain	Portugal	Austria
Excess ineq							
CI - CI $_{NL}$		0.00433	0.00290	0.00853	0.00326	0.01844	0.00394
Ln (Inc)		20.1%	69.1%	42.1%	53.0%	36.1%	50.7%
M30-44		-1.5%	-11.6%	-2.7%	-3.8%	-1.1%	-1.6%
M45-59		7.7%	9.7%	1.4%	13.6%	-1.4%	-11.6%
M60-69		1.0%	-8.1%	2.2%	3.3%	2.4%	0.8%
M70+		10.1%	4.1%	21.0%	4.9%	8.6%	12.6%
F16-29		-0.9%	1.9%	-0.5%	-0.8%	-0.3%	-1.5%
F30-44		-1.7%	-5.8%	-3.4%	-6.8%	-1.4%	-0.8%
F45-59		8.0%	2.7%	1.5%	15.1%	-1.1%	-8.2%
F60-69		4.2%	9.8%	6.9%	0.6%	7.9%	11.0%
F70+		20.3%	12.4%	22.7%	16.6%	15.2%	45.5%
Second Educ		5.8%	28.4%	5.3%	10.5%	4.1%	24.7%
Higher Educ		-6.0%	-5.0%	1.5%	1.8%	4.2%	-6.9%
Part-time empl		0.6%	0.9%	0.3%	1.0%	0.4%	0.8%
Self-employed		-0.3%	-0.9%	0.4%	0.0%	0.3%	1.4%
Student		3.8%	2.2%	2.1%	5.6%	1.4%	2.1%
Unemployed		-6.1%	-15.4%	-6.4%	-18.9%	-2.5%	-11.2%
Housework		9.9%	-6.9%	-3.1%	-9.4%	1.0%	6.1%
Econ inactive		18.1%	-2.8%	-4.4%	10.2%	11.2%	-15.1%
Retired		3.3%	-14.9%	18.1%	-3.2%	22.9%	10.8%
Divorced/sep		-0.4%	-5.2%	-1.9%	-4.2%	-0.7%	-3.1%
Widowed		3.5%	1.0%	2.0%	-3.5%	-1.7%	-2.3%
Unmarried		-0.6%	-1.0%	-1.0%	-1.2%	-1.0%	-1.3%
Born other Euro		0.1%	0.0%	0.0%	-0.4%	0.0%	0.3%
Born non-Euro		0.2%	0.0%	0.0%	0.0%	0.2%	-0.4%
Region 2		0.8%	0.4%	0.0%	10.8%	-0.2%	0.4%
Region 3			5.8%	0.0%	12.8%	0.3%	-3.1%
Region 4			-1.2%	0.0%	-6.3%	-0.9%	
Region 5			-2.0%		16.7%	-1.0%	
Region 6			-0.4%		-14.8%	-0.5%	
Region 7			0.3%		-3.2%	-2.3%	
Region 8			6.7%				
Region 9			19.6%				
Region 10			1.3%				
Region 11			4.8%				

Figure 4.3 Contributions to concentration indices of income-related inequality, by country, by source



all of the excess inequality in Denmark is due to the worse health and income position of the Danish retired compared to the Dutch retired.

Another interesting question is: which of the two components of any contribution - the health elasticity or the inequality by income - is the most important contributor to excess inequality? This distinction is not unimportant from a policy perspective, since in many cases health policies cannot directly alter the distribution of these characteristics by income, but they may be able to influence the health elasticity of some of these characteristics. As explained in section 4.2, the relative magnitudes of their contributions cannot easily be ascertained from comparing the two composite terms in each ΔC_k described in equation (4.12) since they are both a product of a concentration index and an elasticity. Rather, a comparison of the relative differences $\left(C_{ki} - C_{kj}\right)/C_{kj}$ and $\left(\eta_{ki} - \eta_{kj}\right)/\eta_{ki}$ can provide some insight. These differences are presented in Table 4.7 as excess percentages versus the Dutch concentration indices and elasticities (Figure 4.3).

Again, some general observations emerge. In many cases, the elasticity differences appear to dominate the inequality differences. Take the contribution of income: in all countries, without exception, the relative excess elasticity is greater - often much

Table 4.7 Relative 'excess' elasticity and inequality (vs NL) of determinants per country in %

	DE		DK		BE		LU		FR		UK	
	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind
Ln (Inc)	56.5	-4.3	-89.8	-21.1	134.4	-3.4	420.4	-14.2	159.6	-2.5	188.0	8.3
M30-44	14.4	130.6	33.3	291.4	35.4	205.4	31.6	132.3	156.7	6.8	4.4	254.1
M45-59	87.4	-2.3	11.9	48.8	0.0	-36.3	34.7	-73.6	108.1	-6.2	-1.5	31.0
M60-69	78.9	-131.2	-123.9	-466.5	250.4	-253.6	-38.6	-81.3	401.6	-27.2	10.4	-110.4
M70+	321.9	-94.4	-83.4	478.4	347.2	178.7	310.7	16.4	645.5	-54.0	-47.0	372.6
F16-29	54.0	5.8	-39.9	4.8	227.2	-75.8	213.3	-89.7	145.6	18.7	-67.0	-47.5
F30-44	-23.9	-56.4	17.2	-580.4	31.8	-181.1	51.4	-52.7	95.7	-74.9	-60.5	-179.2
F45-59	38.5	-17.2	-1.5	37.4	12.7	-49.2	66.1	-90.3	61.5	20.7	-26.9	11.0
F60-69	40.9	98.8	-67.2	399.5	212.7	81.7	178.0	-12.9	186.1	-56.0	-102.0	228.5
F70+	119.8	-0.3	-1.4	82.8	121.3	23.3	109.5	-8.6	169.0	6.2	-19.6	100.5
Second Educ	-17.1	-99.4	40.8	200.9	-69.4	-592.3	23.6	-3413.3	-24.2	-1356.0	-19.8	-1326.5
Higher Educ	53.3	-40.8	232.1	-31.4	25.5	-17.5	52.3	34.7	72.7	22.5	144.3	0.3
Part-time empl	-110.5	-119.0	-71.1	-224.7	311.8	-394.0			-84.7	-217.5	-91.7	-40.6
Self-employed	-116.4	56.3	544.9	-4.2	-2887.8	-141.6	-133.3	13.5	1165.6	-33.1	1695.4	125.2
Student	-11.8	-34.8	20.5	-28.6	-7.9	-79.6	-6996.1	-205.2	84.7	-72.6	-133.6	-35.7
Unemployed	-6.9	-26.3	-33.8	-99.4	13.6	-5.2	-437.8	35.6	-48.0	9.6	-64.8	7.8
Housework	-70.7	-23.2	-91.8	142.0	-79.2	51.0	-144.8	-63.6	57.0	112.8	16.1	115.2
Econ inactive	-99.5	127.5	-75.4	-37.6	-67.6	187.7	-88.2	71.7	-99.3	-53.4	19.5	196.2
Retired	1238.1	7.5	2115.2	200.2	249.5	8.0	137.3	-11.1	686.2	-60.6	930.2	119.0
Divorced/sep	21.1	7.3	45.1	3.9	-15.1	-8.8	18.3	-51.6	47.0	-50.6	-31.5	70.0
Widowed	-190.8	40.6	-74.9	189.0	-18.3	75.4	-486.3	-54.9	117.2	108.8	170.3	260.2
Unmarried	-528.6	-71.0	-47.0	36.1	-94.9	-94.0	-1186.4	-135.3	1017.8	-36.9	-43.3	-96.6

Table 4.7 (Continued) Relative 'excess' elasticity and inequality (vs NL) of determinants per country in %

	IE		IT		GR		ES		PT		AT	
	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind	Elast	Conc ind
Ln (Inc)	65.6	9.8	138.0	21.1	239.9	28.8	119.8	19.5	419.1	39.8	228.5	-12.4
M30-44	-37.7	190.9	75.0	191.8	17.4	223.1	17.4	114.9	151.7	41.5	31.2	33.9
M45-59	-24.2	-56.8	84.1	-76.6	51.0	-50.0	35.8	-92.2	146.3	-38.8	113.2	-9.9
M60-69	103.6	-131.1	727.2	16.6	635.0	-181.5	447.8	-155.3	811.9	-265.0	157.6	-108.0
M70+	350.5	218.3	1404.2	-69.3	1216.2	324.3	820.5	-35.8	1037.9	335.2	645.9	118.0
F16-29	-122.7	-111.2	134.0	-0.4	-122.0	-70.4	1.9	-60.3	-26.9	-155.6	-205.0	-56.7
F30-44	-27.4	-107.0	49.0	-190.6	-12.6	-458.7	12.9	-290.0	185.6	-194.1	-29.5	-22.4
F45-59	-51.7	-56.1	127.4	-63.7	55.9	-55.0	81.7	-106.5	245.7	-57.3	60.2	8.4
F60-69	55.2	132.5	639.8	-31.7	357.9	104.4	414.9	-75.1	732.2	162.1	176.5	158.9
F70+	55.5	82.4	382.8	-63.8	266.9	37.5	286.9	-45.0	350.2	52.1	253.5	33.9
Second Educ	-71.7	-3923.2	7.1	-3631.2	-41.2	-3528.6	-52.5	-3302.7	-51.1	-7139.2	147.4	-1912.8
Higher Educ	-56.5	49.7	-34.1	22.3	2.6	14.5	-8.2	17.4	-5.0	116.0	-31.4	-7.3
Part-time empl	-99.0	194.6	-99.3	-129.9	-97.8	-93.3	-91.1	-381.5	-5.5	-283.7	-108.3	49.2
Self-employed	325.5	36.5	-3115.9	-136.9	-4552.1	-75.5	-528.0	-137.3	1845.9	-219.5	4250.9	-148.5
Student	-47.7	-80.5	129.3	-71.4	-2.3	-96.1	24.7	-98.9	24.4	-133.9	75.0	-68.6
Unemployed	-54.4	17.0	-83.6	25.4	-95.7	-32.1	-112.5	-20.6	-46.7	-65.6	-45.9	-60.7
Housework	-9.2	89.8	-58.3	58.2	-19.3	-31.5	-59.3	17.8	-13.1	49.2	-29.0	97.7
Econ inactive	-22.0	188.8	-53.5	87.6	-46.4	-25.1	45.3	5.6	93.0	123.1	-77.6	-79.5
Retired	65.0	27.5	483.3	-140.5	873.8	34.4	320.0	-95.9	2222.2	46.1	789.9	-51.5
Divorced/sep	-61.2	130.0	-97.3	-193.4	-48.8	-123.4	-82.4	-59.0	-88.4	-20.6	-30.3	-73.3
Widowed	47.1	164.6	123.5	-31.4	277.6	13.6	-294.9	-40.1	-378.9	75.3	-138.2	94.3
Unmarried	140.5	-86.6	8.7	-75.9	359.5	-127.5	94.7	-101.8	1153.8	-128.9	284.2	-109.7

greater- than the relative excess inequality. This implies that it is not so much the differences in income inequality *per se*, but in the partial association between income and health that matter for income-related health inequality. The same observation holds for another important influence on such inequality: the partial association between health and retirement status appears in all countries as more influential than the degree to which retired are ranked lower in the income distribution. This is an important finding because, as we said before, it implies that reducing health inequalities seems more a matter of reducing these associations through appropriate health related policies than a matter of redistributing income. The observation does not hold for all other variables, but on the basis of Table 4.7, policy makers from each of the other countries can learn where the greatest opportunities lie for reducing income-related health inequalities if the goal is to come closer to the, apparently achievable, low degree of health inequality in the Netherlands.

4.5 CONCLUSION AND DISCUSSION

This paper adds considerably to the existing knowledge on inter-individual health disparities by income in Europe. First, by using a new data set with better and far more comparable measures of income and health, it provides more reliable estimates of the cross-country differences in health inequalities. Secondly, by using an interval regression approach to estimating more fully specified health (utility) equations, it achieves a more consistent and reliable estimate of the degree of health inequalities which is not due to demographic differences. Thirdly, by using a new decomposition technique, it allows to decompose the total observed income-related health inequality into the contributions of the health elasticity and the inequality by income for all health determinants included in the analysis. Fourth, by bootstrapping the entire estimation and decomposition procedure, it was possible to not only estimate but also statistically test for cross-country differences in the factors which drive inequality. Finally, an Oaxaca-type decomposition of cross-country differences enabled the identification of the relative contribution of 'excess elasticity' and 'excess inequality' in each of the determinants compared to a low-inequality country.

We find that especially Portugal, but also the UK and Denmark, show up with a high degree of such inequality, while countries like the Netherlands, Germany, but also Italy, Belgium, Spain, Austria and Ireland show a relatively low level of health inequality. The correlation with income inequality is positive, but weaker than in previous research. The decomposition shows that (the health effect and distribution of) income itself is the most important but not the only driving factor. When abstracting from the contributions of demographic variables, mainly (the health effect and distribution of) education, labour force status and region are the prime other contributors to health inequality.

This is in line with findings for Canada (Van Doorslaer and Jones, 1997). But there are exceptions. Denmark's relatively high level of income-related health inequality is *not* due to its income inequality (which is low) but almost entirely to the fact that early retired individuals have much worse health and are strongly concentrated among the lower income groups. Early retirement in this country may be used as a vehicle for those withdrawing from the labour force (early) for health reasons. In most other countries, it is either the health and income position of the retired, of other categories of non-labour force participants (e.g. the other economically inactive), or of those with the lowest education, which are the most important contributors after income itself. In the southern European countries, a non-negligible portion of the income-related inequality is also due to regional health inequalities.

But apart from the similarities, there are also some dramatic differences between European countries. Some are achieving a much lower degree of health inequality than others. The decomposition of these cross-country differences into the excess inequalities and elasticities of the determinants in comparison to the country with the lowest inequality (the Netherlands) brings out that in general the excess elasticities play a much greater part than the excess inequalities by income. This finding has very important policy implications. It means that reducing health inequalities appears to be more a matter of health policy than of income redistribution.

Of course, the latter conclusion crucially hinges on a causal interpretation of the health equations and this brings us to the limitations of this study. As explained in section 4.1, the estimated health equations do not generally allow for such a causal interpretation. The partial cross-sectional association between income and health as measured in the regression coefficient of (log) income may also to some extent reflect reverse causality or joint determination by some unobserved underlying factor. The same applies to some of the other variables, notably the variables reflecting labour force participation status. In this paper, we did not attempt to account for endogeneity or unobserved heterogeneity. In the near future, when sufficient waves of the ECHP panel will become available, we intend to use panel data techniques to address this limitation of the present study. A second caveat concerns the external validity of using the Canadian threshold values for health utility to scale the SAH categories in the interval regression. While we do not expect this to be very influential for the inequality results obtained, ideally this ought to be tested using some European source of health utility thresholds and applying the validity tests applied in Van Doorslaer and Jones (1997) on the Canadian data. Recent results from similar tests (and decompositions) using Finnish and Danish data (Lauridsen et al., 2002) indicate that extremely similar results for both

countries are obtained when instead of the Canadian HUI threshold values, thresholds based on the distribution of the Finnish distribution of another generic health measure: the EQ15D. It therefore appears unlikely that our results are very sensitive to the choice of generic measure for the threshold values.

The new empirical evidence on cross-European differences in income-related health inequalities provided in this paper also generates some important new policy insights. Some European countries achieve much lower degrees of such inequality than others. These are not necessarily the healthiest or the richest countries. While a lower degree of income inequality will certainly help in reducing health inequalities, it is certainly not the sole or the main driver of the inequality differences. What appears crucial is the following: (i) the strength of the partial association between income and health; measures which can reduce either the health-harming effects of income losses or the income consequences of health losses (e.g. through replacement incomes) will definitely help; (ii) the associations between health and income (rank) of being outside of the labour force; only if it is very hard or nearly impossible to reduce the health elasticities by improving or maintaining the health of the non-active (e.g. in the case of the disabled or aged), then the only option to reduce these inequalities may be through appropriate income policies; (iii) regional differences in health and income; if there were no systematic regional disparities, clearly income-related inequality would also be lower. While the paper does not tell us what could be done and how to change the inequality components, it does show where the greatest potential for inequality reductions lie. Future research which can exploit longitudinal changes in income and health could help to further unravel the underlying causal pathways which generate the patterns of income-related health differences which have emerged from this study.

Notes

a. In this paper we often use the word inequality for socioeconomic inequality, which is only one part of total health inequality. For a more detailed discussion of the relationships between these two concepts of inequality, see e.g. Wagstaff, A. and E. Van Doorslaer, 2003, Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations, *Health Economics* (Forthcoming)
b. There is one important exception: the French questionnaire asked respondents to indicate the satisfaction with their general health, ranging from very satisfied to very dissatisfied.
c. Due to an apparent coding error in the data, the status ‘retired’ for the Netherlands had to be r econstructed from age and income source information.
d. This peculiarity may have to do with the fact that only in France the question was one of a series asking about satisfaction with work, income, housing, health and leisure.

Table 4.A.1 Pairwise comparisons of I* estimates between countries: t-test statistics

	Netherl	Germany	Spain	Belgium	Austria	Ireland	Italy	France	Greece	Luxemb	Denmark	UK
Netherl												
Germany	1.03											
Spain	**2.64**	0.96										
Belgium	**2.46**	1.06	0.24									
Austria	**2.27**	1.04	0.26	0.04								
Ireland	**3.07**	1.34	0.55	0.22	0.15							
Italy	**3.26**	1.49	0.77	0.38	0.30	0.20						
France	**4.35**	**2.72**	**2.37**	1.80	1.58	1.83	1.64					
Greece	**5.09**	**3.07**	**2.85**	**2.10**	1.82	**2.21**	**1.99**	0.14				
Luxemb	**3.85**	**2.94**	**2.61**	**2.28**	**2.13**	**2.28**	**2.16**	0.95	0.90			
Denmark	**5.66**	**4.24**	**4.10**	**3.48**	**3.19**	**3.62**	**3.46**	1.79	1.83	0.47		
UK	**8.83**	**6.66**	**7.05**	**5.92**	**5.37**	**6.37**	**6.17**	**3.81**	**4.09**	1.72	1.51	
Portugal	**11.52**	**9.05**	**9.80**	**8.40**	**7.67**	**9.05**	**8.83**	**6.12**	**6.61**	**3.29**	**3.43**	**2.16**

Note: Countries ranked by *I**. Values significantly different from zero (at P<0.05) in bold typeface.

Atkinson, A, B. Cantillon, E. Marlier and B. Nolan. Social indicators: the EU and social inclusion, Oxford University Press, 2002, Oxford.

Brazier, J, T Usherwood, R Harper and K Thomas, Deriving a preference-based single index from the UK SF-36 Health Survey, J Clinical Epidemiology, 1998, 51 (11): 1115-1128

Burström, B. & Fredlund, P. Self rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? Journal of Epidemiology and Community Health, 2001, 55 (11), 836-840

Cavelaars, AEJM, AE Kunst, JJM Geurts, R Crialesi , L Grötvedt, U Helmert et al, Differences in self-reported morbidity by educational level: a comparison of 11 western European countries, Journal of Epidemiology and Community Health, 1998, 52: 219-27

Deaton, A., The analysis of household surveys: a microeconometric approach to development policy, Johns Hopkins University Press, 1997, Baltimore.

Dolan, P, Modelling valuations for EuroQol health states, Medical Care, 1997, 35 (11): 1095-1108

Efron, B. and R.J. Tibshirani, An introduction to the bootstrap, Chapman and Hall, 1993, London.

EUROSTAT, European Community Household Panel (ECHP): selected indicators from the 1995 wave, Theme 3, Sub-theme "Population and social indicators", OPOCE, Luxembourg, 1999, 260pp

Feeny, D, W Furlong, M Boyle and G Torrance, 1995, Multi-attribute health status classification systems: Health Utilities Index, Pharmacoeconomics, 7 (6): 490-502

Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M., Multiattribute and single-attribute utility functions for the health utilities index mark 3 system, Medical Care, 2002, 40(2):113-28

Gravelle, H. (2001). Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics, 12(10): 803-19

Idler, E.L. and Benyamini, Y. Self-rated health and mortality: a review of twenty-seven community studies, Journal of Health and Social Behavior, 1997, 38(1): 21-37.

Jones, A.M. Health Econometrics, in A.J. Culyer and J.P. Newhouse (eds.) Handbook of Health Economics. Elsevier, 2000, Amsterdam.

Kakwani, N.C., A. Wagstaff and E. van Doorslaer, Socioeconomic inequalities in health: measurement, computation and statistical inference. Journal of Econometrics, 1997, 77(1): 87-104

Lauridsen, J, T Christiansen and U Häkkinen, Determinants of inequality in self-reported health: the Danish/Finnish case, Working Paper, Department of Statistics and Demography, University of Southern Denmark, 2002, Odense.

Lerman, RI and S Yitzhaki. Improving the accuracy of estimates of Gini coefficients, Journal of Econometrics, 1989, 42(1) , 43-47

Oaxaca, RL, Male-female wage differentials in urban labor markets. International Economic Review, 1973, 14, 693-709.

Schokkaert, E and C van de Voorde, Risk selection and the specification of the risk adjustment formula, Journal of Health Economics, 2004; 23:1237-1259

Van Doorslaer, E. and U.-G. Gerdtham, Does inequality in self-assessed health predict inequality in survival by income? ¯Evidence from Swedish data, Social Science and Medicine, 2003; 57:1621-1629

Van Doorslaer, E, A Wagstaff, H Bleichrodt et al, Income-related inequalities in health: some international comparisons, Journal of Health Economics, 1997, 16: 93-112

Van Doorslaer, E. and A.M. Jones. Inequalities in self-reported health: validation of a new approach to measurement. Journal of Health Economics, 2003, 22: 61-87

Wagstaff, A, van Doorslaer, E., Watanabe, N. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam, Journal of Econometrics, 2003, 112 (1), 207-223

Wagstaff, A. and van Doorslaer, E., Measuring inequalities in health in the presence of multiple-category morbidity indicators, Health Economics, 1994, 3, 281-291.

Wagstaff, A. and van Doorslaer, Income-Related Inequalities in Health in Europe: Evidence from the European Community Household Panel, ECuity II Project Working Paper #1, Department of Health Policy and Management, Erasmus University, 2000, Rotterdam.

Wagstaff, A., P. Paci and E. van Doorslaer, On the measurement of inequalities in health, Social Science and Medicine, 1991, 33, 545-557.

WHO. Health interview surveys: towards international harmonisation of methods and instdruments. WHO Regional Publications, European Series No. 58, 1996, Copenhagen

CHAPTER 5

Health-related non-response in panel data

ABSTRACT

This paper considers health-related non-response in the first eleven waves of the British Household Panel Survey (BHPS) and the full eight waves of the European Community Household Panel (ECHP) and explores its consequences for dynamic models of the association between socioeconomic status and self-assessed health (SAH). We describe the pattern of health-related non-response revealed by the BHPS and ECHP data. We both test and correct for non-response in empirical models of the impact of socioeconomic status on self-assessed health. Descriptive evidence shows that there is health-related non-response in the data, with those in very poor initial health more likely to drop out, and variable addition tests provide evidence of non-response bias in the panel data models of SAH. Nevertheless a comparison of estimates - based on the balanced sample, the unbalanced sample and corrected for non-response using inverse probability weights (IPW) – shows that, on the whole, there are not substantive differences in the average partial effects (APE) of the variables of interest. The main differences are between unweighted and one form of IPW-weighted estimates for the APE of income and education in those countries that have fewer than eight waves of data. Similar findings have been reported concerning the limited influence of non-response bias in models of various labour market outcomes; we discuss possible explanations for our results.

5.1 INTRODUCTION

The objective of this paper is to explore the existence of health-related non-response in panel data and its consequences for modelling the association between socioeconomic status (SES) and self-assessed health (SAH). Using panel data, such as the British Household Panel Survey (BHPS) or European Community Household Panel (ECHP), to analyse longitudinal models of health creates a risk that the results will be contaminated by bias associated with longitudinal non-response. There are drop-outs from the panels at each wave and some of these may be related directly to health: due to deaths, serious illness and people moving into institutional care. In addition, other sources of non-response may be indirectly related to health, for example divorce may increase the risk of non-response and also be associated with poorer health than average. The long-term survivors who remain in the panel are likely to be healthier on average compared to the sample at wave 1. The health of survivors will tend to be higher than the population as a whole and their rate of decline in health will tend to be lower. Also, the socioeconomic status of the survivors may not be representative of the original population who were sampled at wave 1. Failing to account for non-response may result in misleading estimates of the relationship between health and socioeconomic characteristics. To address this issue we describe the pattern of health-related non-response revealed by the BHPS and ECHP data and we test and correct for non-response in empirical models of self-assessed health (SAH).

There are many recent studies that have used the BHPS, ECHP and other similar panels to estimate models involving measures of health and that have used regression analyses based on balanced or unbalanced panels which may be prone to problems of non-response. Examples include: Benzeval and Judge (2001) who analyse health and SES with the BHPS; Meer et al, (2003) who analyse health and SES with the US PSID; Buckley et al (2004) who analyse SAH and SES with the Canadian SLID; Contoyannis et al (2004a) who analyse health limitations in the BHPS; Wildman (2003) who analyses the relationship between mental health and SES in the BHPS; and Riphahn (1999) who analyses retirement and health with the GSOEP.

The paper adopts a broad definition of longitudinal non-response, that encompasses any observations that "drop-out" from the original wave 1 sample over the subsequent T waves. To borrow the taxonomy of reasons for non-participation used by Nicoletti and Peracchi (2005), non-response can arise due to:

1. Demographic events such as death.
2. Movement out of scope of the survey (e.g. institutionalization or emigration).
3. Refusal to respond at subsequent waves.
4. Absence of the person at the address.
5. Other types of non-contact.

To these points, we would add item non-response for any of the variables used in the model of health, which eliminates these observations from the sample. The notion of attrition, commonly used in the survey methods literature, is usually restricted to points 3, 4 and 5. However our concern is with any longitudinal non-response that leads to missing observations in the panel data regression analysis. In fact it is points 1 and 2 – death and incapacity – that are likely to be particularly relevant as sources of health-related non-response. The original sample consists of those who provide a full interview and usable information on SAH at the first wave of the panels. Non-response encompasses all of those who fail to provide usable observations for the model of SAH at subsequent waves.

Our aim is to estimate models that focus on the relationship between health (SAH) and socioeconomic status (SES). We take a representative sample of individuals at wave 1 and follow them for the 11 or 8 years of the BHPS and ECHP panels. The sample of interest is those n original individuals observed over a full T-year period (T=11 for BHPS and T=8 for ECHP). A fully observed sample from this population would consist of nT observations. Due to non-response we only observe $\sum_{i=1}^{n} T_i$ observations. The reasons for having incomplete observations include attrition (as conventionally defined in the survey methods literature) as well as individuals becoming ineligible, due to incapacity or death. This creates a problem of *incidental truncation*: we are interested in the association between SAH and SES for our n individuals over the full T waves. However the more frail individuals are more likely to die or drop-out before the end of the observation period, and their levels of SAH and SES are unobservable. This means that the remaining observed sample of survivors may contain less frail individuals – this is the source of potential bias in the relationship between SAH and SES across our sample of n individuals.

We apply variable addition tests for attrition bias (Verbeek and Nijman, 1992) and inverse probability weighting to adjust for non-response in estimation of pooled models (Robins *et al.* 1995; Fitzgerald *et al.*, 1998; Moffitt *et al.*, 1999; Wooldridge, 2002a). Descriptive evidence shows that there is health-related non-response in the data, with those in poor initial health more likely to drop out, and variable addition tests provide evidence of non-response bias in panel data models of SAH. Nevertheless a comparison of estimates - with and without correcting for non-response using inverse probability weights - does not show substantive differences in the average partial effects of the variables of interest. So, while health-related non-response exists, it does not appear to distort the magnitudes of the estimated effects of socioeconomic status. Similar findings have been reported concerning the limited influence of non-response bias in models of various labour market outcomes; we discuss possible explanations for our results.

The structure of the paper is as follows. Section 5.2 introduces the BHPS and ECHP datasets. Section 5.3 presents a descriptive analysis of health-related non-response in both surveys. In Section 5.4 we introduce the empirical models for self-assessed health and describe the estimation strategy. Section 5.5 reports and discusses the results for the models of socioeconomic status and self-assessed health and a conclusion is provided in Section 5.6.

5.2 DATA

5.2.1 BHPS

The sample

We first exploit the panel data available in the first eleven waves (1991-2001) of the British Household Panel Survey (BHPS). The BHPS is a longitudinal survey of private households in Great Britain that provides rich information on socio-demographic and health variables. It was designed as an annual survey of each adult (16+) member of a nationally representative sample of more that 5,000 households, with a total of approximately 10,000 individual interviews. The first wave of the survey was conducted between 1st September 1990 and 30th April 1991. The initial selection of households for inclusion in the survey was performed using a two-stage clustered systematic sampling procedure designed to give each address an approximately equal probability of selection (Taylor *et al.*, 1998). The same individuals are re-interviewed in successive waves and, if they split off from their original households are also re-interviewed along with all adult members of their new households. In this analysis we use both *balanced samples* of respondents, for whom information on all the required variables is reported at each wave, and *unbalanced samples,* that exploit all available observations for wave 1 respondents. Both samples do not include new entrants to the BHPS; they only track all of those who were observed at wave 1. In this sense, the analysis treats the sample as a cohort consisting of all those present at wave 1. To be included in the analysis individuals must be original sample members (OSMs) who were aged 16 or over and who provided a valid response for the health measure at wave 1. Our broad definition of non-response encompasses all individuals who are missing at subsequent waves.

Measures of health

The principal health outcome is self-assessed health (SAH), defined by a response to: 'Please think back over the last 12 months about how your health has been. Compared to people of your own age, would you say that your health has on the whole been excellent/good/fair/poor/very poor?' SAH should therefore be interpreted as indicating

a perceived health status relative to the individual's concept of the 'norm' for their age group. SAH has been used widely in previous studies of the relationship between health and socioeconomic status (e.g., Ettner, 1996; Deaton and Paxson, 1998; Smith, 1999; Benzeval *et al.*, 2000; Salas, 2002; Adams *et al.*, 2003; Frijters *et al.*, 2003; Contoyannis *et al.,* 2004b) and of the relationship between health and lifestyles (e.g., Kenkel, 1995; Contoyannis and Jones, 2004). SAH is a simple subjective measure of health that provides an ordinal ranking of perceived health status. However it has been shown to be a powerful predictor of subsequent mortality (see e.g., Idler and Kasl, 1995; Idler and Benyamini, 1997) and its predictive power does not appear to vary across socioeconomic groups (see e.g., Burström and Fredlund, 2001). Socioeconomic inequalities in SAH have been a focus of research (see e.g., Van Doorslaer *et al.*, 1997; Van Ourti, 2003; Van Doorslaer and Koolman, 2004) and have been shown to predict inequalities in mortality (see e.g., Van Doorslaer and Gerdtham, 2003). Categorical measures of SAH have been shown to be good predictors of subsequent use of medical care (see e.g., Van Doorslaer *et al.*, 2000; Van Doorslaer *et al.,* 2004).

Unfortunately there was a change in the wording of the SAH question at wave 9 of the BHPS. For waves 1-8 and 10-11, the SAH variable represents "health status over the last 12 months". However, the SF-36 questionnaire was included in wave 9. In this questionnaire, the SAH variable for wave 9 represents "general state of health", using the question: "In general, would you say your health is: excellent, very good, good, fair, poor?". Note that the question is not framed in terms of a comparison with people of one's own age and the response categories differ from the other waves. Item non-response is greater for SAH at wave 9 than for the other waves and these factors would complicate the analysis of non-response rates. Hernandez *et al.* (2004) have explored the sensitivity of ordered probit models of SAH to this change in the wording, but for simplicity we exclude wave 9 from the analysis.

Other indicators of morbidity are used to describe health-related non-response and as predictors of non-response. The BHPS variable HLLT measures self-reported functional limitations and is based on the question "does your health in any way limit your daily activities compared to most people of your age?" Respondents are left to define their own concepts of health and their daily activities. In contrast, for the variable measuring specified health problems (HLPRB), respondents are presented with a prompt card and asked, "do you have any of the health problems or disabilities listed on this card?" The list is made up of problems with arms, legs, hands, etc; sight; hearing; skin conditions/ allergies; chest/breathing; heart/blood pressure; stomach/digestion; diabetes; anxiety/ depression; alcohol/drug related; epilepsy; migraine and other (cancer and stroke were added as separate categories in wave 11 but are not included here). Also respondents are asked to report whether they are registered as a disabled person (HLDSBL).

Socioeconomic status

Two dimensions of socioeconomic status are included in our models of SAH: income and education. Income is measured as equivalised and RPI-deflated annual household income (INCOME). This variable is transformed to natural logarithms to allow for concavity of the relationship between health and income (e.g., Ettner, 1996; Frijters *et al.*, 2003; Van Doorslaer and Koolman, 2004; Contoyannis *et al.,* 2004a,b). Education is measured by the highest educational qualification attained by the end of the sample period in descending order of attainment (DEGREE, HND/A, O/CSE). NO-QUAL (no academic qualifications) is the reference category for the educational variable. In addition to income and education, variables are included to reflect individuals' demographic characteristics and stage of life: age, ethnic group, marital status and family composition. Marital status distinguishes between WIDOW, SINGLE (never married) and DIVORCED/SEPARATED, with married or living as a couple as the reference category. Similarly, we include an indicator of ethnic origin (NON-WHITE), the number of individuals living in the household including the respondent (HHSIZE), and the numbers of children living in the household at different ages (NCH04, NCH511, NCH1218). Age is included as a fourth-order polynomial, (AGE, AGE2 = $AGE^2/100$, AGE3 = $AGE^3/10000$, AGE4 = $AGE^4/1000000$). A vector of wave dummies is included to account for aggregate health shocks, time-varying reporting changes, and any effects of age which are not captured by the polynomial.

5.2.2 ECHP

The sample

The detailed analysis of the BHPS is complemented by a second source of data: the full eight waves, 1994-2001, of the *European Community Household Panel User Database* (ECHP-UDB) designed and coordinated by Eurostat, the European Statistical Office. This puts the UK data in the context of a broader analysis of patterns of health-related non-response across European countries. The ECHP is a standardised multi-purpose annual longitudinal survey carried out at the level of the European Union (Peracchi, 2002). The survey is based on a standardised questionnaire that involves annual interviewing of a representative panel of households and individuals of 16 years and older in each of the participating EU member states. It covers a wide range of topics including demographics, income, social transfers, health, housing, education and employment. We use data for the following fourteen member states of the EU for the full number of waves available for each: Austria (waves 2-8), Belgium (1-8), Denmark (1-8), Finland (3-8), France (1-8), Germany (1-3), Greece (1-8), Ireland (1-8), Italy (1-8), Luxembourg (1-3), Netherlands (1-8), Portugal (1-8), Spain (1-8) and the United Kingdom (1-3). Sweden did not take part

in the ECHP although the living conditions panel is included with the UDB. The ECHP-UDB also includes comparable versions of the BHPS and German Socioeconomic Panel (GSOEP) and descriptive evidence is provided for these.

Table 5.1 SAH sample size, drop outs, re-joiners, survival rate (%) and drop-out rates (%) by wave and previous period health status, BHPS

					drop out rates						
Wave	No. Ind.	Drop outs	Re-joiners	Survi-val rate	Raw	Net	EX at t-1	GOOD at t-1	FAIR at t-1	POOR at t-1	VPOOR at t-1
Men											
1	4832	652									
2	4180	428	0	86.5	13.5	13.5	12.2	13.5	14.2	14.6	26.9
3	3878	285	126	80.3	10.2	7.2	8.9	9.5	11.5	14.6	24.0
4	3675	283	82	76.1	7.4	5.2	6.7	7.4	7.3	8.5	14.5
5	3464	156	72	73.8	7.7	5.7	5.4	7.4	9.6	9.7	23.0
6	3408	159	100	70.5	4.5	1.6	3.6	3.1	4.8	12.2	25.4
7	3280	175	31	67.9	4.7	3.8	3.3	4.5	4.6	9.7	11.5
8†	3137	-	32	64.9	5.3	4.4	4.1	4.4	6.4	7.0	22.9
10‡	2899	137	-	60.0	-	-	-	-	-	-	-
11	2820		58	58.4	4.7	2.7	4.5	4.5	4.8	5.2	10.4
Women											
1	5424										21.4
2	4777	647	0	88.1	11.9	11.9	10.8	11.8	12.1	13.2	16.4
3	4532	373	128	83.6	7.8	5.1	7.2	7.0	8.2	11.3	14.0
4	4303	305	76	79.3	6.7	5.1	6.7	5.8	6.3	11.6	12.0
5	4106	268	71	75.7	6.2	4.6	7.1	5.2	6.6	8.3	14.3
6	4016	179	89	74.0	4.4	2.2	2.6	3.4	5.2	9.3	7.1
7	3882	166	32	71.6	4.1	3.3	3.0	3.3	4.9	8.3	12.6
8†	3775	151	44	69.6	3.9	2.8	2.7	3.3	4.5	5.2	-
10‡	3522	-	-	64.9	-	-	-	-	-	-	13.5
11	3403	183	64	62.7	5.2	3.4	4.3	3.7	6.8	7.4	

Drop-outs – respondents at wave t-1, non-respondents at wave t.

Re-joiners – non-respondents at wave t-1, respondents at wave t.

* Raw drop-out rates exclude re-joiners; Net drop-out rates include re-joiners.

† At wave 9 the self-assessed health question was changed to one based on the SF-36 questionnaire. SF-36 questionnaire response rates appear lower than those for hlstat and therefore are not used as a basis for calculating drop-out rates.

‡ Drop-out rates conditional on previous wave reporting of self-assessed health are not possible due to the change in the self-assessed health question at wave 9

Measures of health

Self-assessed general health status (SAH) is measured as either very good, good, fair, poor or very poor. Unlike the BHPS, respondents are not asked to compare themselves with others of the same age. In France a six-category scale was used but this is recoded to the five-category scale in the ECHP-UDB. Responses are also available for the question "Do you have any chronic physical or mental health problem, illness or disability? (yes/no)" and if so "Are you hampered in your daily activities by this physical or mental health problem, illness or disability? (no; yes, to some extent; yes, severely)". We use two dummy variables to indicate either some limitation or severe limitation.

Socioeconomic status

The ECHP income measure is disposable household income per equivalent adult, using the modified OECD equivalence scale (giving a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 4 in the household). Total household income includes all net monetary income received by the household members during the reference year. Education is measure by the highest level of general or higher education completed, i.e. third level education (ISCED 5-7), second stage of secondary level education (ISCED 3-4) or less than second stage of secondary education (ISCED 0-2)). Marital status distinguishes between married/living in consensual union, separated/divorced, widowed and unmarried. Activity status includes employed, self-employed, student, unemployed, retired, doing housework and 'other economically inactive'. Region of residence uses the EU's NUTS 1 level (Nomenclature of Statistical Territorial Units) except for countries where such information was withheld for confidentiality reasons (The Netherlands, Germany) or because the country is too small (Denmark, Luxembourg).

5.3 DESCRIPTIVE ANALYSIS OF NON-RESPONSE RATES

5.3.1 BHPS

Table 5.1 shows how the sample size and composition evolves across the waves of the BHPS for respondents who provided information on SAH. The table, which gives figures men and women separately, shows the number of observations that are available at each wave and the corresponding number of drop-outs and re-joiners between waves. These are expressed as wave-on-wave survival and drop-out rates. The survival rate is the percentage of original sample members remaining at wave *t*. The drop-out rate is the percentage of the number of drop-outs between waves t-1 and t to the number of observations at t-1. The raw drop-out rate excludes re-joiners, while the net drop-out rate includes them. Drop-out rates are highest between waves 1 and 2, with the rate tending

to decline over time. The table also disaggregates the raw drop-out rates according to individuals' SAH at wave t-1. This shows that drop-out rates are inversely related to past health and, in particular, non-response is highest among those who were in very poor health prior to dropping-out. This pattern of *health-related non-response* persists throughout the panel and is stronger for men than women.

Table 5.2 shows that the overall drop-out rate across all 11 waves of the panel varies with socioeconomic characteristics measured at wave 1. The average rate of drop-out over 11 waves is 39%. As expected, non-response increases with individuals' age at the start of the panel, ranging from 36% for those aged under 30 to 73% for those aged over 70. Some of this age-related non-response is likely to be associated with health, through

Table 5.2 SAH-related drop-out rates (%) over 11 waves by gender, age, income, educational and marital status, BHPS

	All	Ex at t1	Good at t1	Fair at t1	Poor at t1	VPoor at t1
All data	39	36	37	44	48	64
Gender:						
Men	42	38	39	49	53	67
Women	37	33	36	40	45	63
Age group:						
<30	36	40	34	37	22	22
31-50	32	30	31	35	32	47
51-70	39	33	37	41	53	64
>70	73	60	69	77	87	95
Income quintile:						
1	58	55	57	59	58	74
2	41	36	39	44	47	62
3	37	35	36	41	45	62
4	36	36	34	39	44	45
5	32	32	32	29	38	71
Education:						
Degree	26	26	27	29	24	0
Hnd/a level	30	31	28	32	28	67
O level / cse	34	36	32	36	32	49
No qualifications	48	42	46	52	55	65
Marital status:						
Widow	62	47	60	64	81	83
Single	42	45	40	44	44	42
Divorced/separated	41	38	37	45	47	60
Married/couple	35	32	34	40	41	64

deaths, serious illness and moves to institutional care. Non-response is greater among those with lower income and with less formal education: the poorest quintile have an overall drop-out rate of 58%, compared to 32% among the richest quintile; those with no qualifications have an overall drop-out rate of 48% compared to 26% among those with a degree. The table also shows that health-related non-response interacts with individuals' socioeconomic characteristics (some caution is required as some of the cell sizes are very small). So, for example, drop-out rates are very high among elderly individuals (aged >70) who start the survey in poor (87%) or very poor health (95%).

Tables 5.1 and 5.2 provide a description of simple bivariate relationships between drop-out rates and socioeconomic characteristics. To extend this to a multivariate analysis Table 5.3 presents probit models for response/non-response at each wave of the panel, from wave 2 to wave 11, using the full sample of men who are observed at wave 1 (the results for women are similar and are available from the authors on request). The dependent variables for these models equal 1 if the individual responds at the wave in question and 0 otherwise and are always defined relative to the full sample at wave 1 (where a response is defined as providing a usable observation for the ordered probit regression models). The probability of response is modelled as a function of the wave 1 values of all of the regressors that are included in our empirical model of SAH, along with additional wave 1 variables for region (NORTH-WEST, NORTH-EAST, SOUTH-EAST, SOUTH-WEST, MIDLANDS, SCOTLAND, WALES), activity status (SELF-EMPLOYED; UNEMPLOYED; RETIRED; family care and maternity leave, FAMILY-CARE; government training, students and other, EMP-OTHER) and occupational group (UNCLASSIFIED; MANUAL-TECHNICAL; skilled non-manual, SKILL-NON-MAN; skilled manual and armed forces, SKILLED-MANUAL; PART-SKILLED; UNSKILLED, LONG-TERM-SICK) and other indicators of morbidity (HLPRB, HLDSBL, HLLT). These additional observable variables form the basis of the inverse probability weighting approach to correcting for non-response, which is described in more detail below.

The table shows the partial effects of the regressors on the probability of response at each wave, along with an indication of which of these are statistically significant at the 5% level and 1% levels. The partial effects are computed as marginal effects for continuous regressors and average effects for discrete regressors, evaluated at the sample means of the other regressors in the model. These results reveal statistically significant associations between non-response and levels of educational attainment for both men and women. Those with DEGREE, HND/A and O/CSE qualifications are more likely to remain in the sample and the magnitude of this effect increases over the waves. On average, a man with a degree has a 0.07 higher probability of responding at wave 2, relative to one without academic qualifications. By wave 11 they have a 0.169 higher probability of responding. For women the corresponding figures are 0.084 and

Table 5.3 Probit models for response/non-response by wave; Men, BHPS

	Wave								
N = 4543	2	3	4	5	6	7	8	10	11
Ln(Income)	.023*	.026*	.034**	.020	.021	.025	.027	.011	.016
Widow	-.034	-.010	-.032	.010	.009	-.030	-.022	-.044	-.042
Single	-.018	-.040	-.048*	-.051*	-.046*	-.053*	-.065**	-.074**	-.070**
Div/sep	-.014	-.007	-.031	-.005	.002	-.020	-.034	-.061	-.075
Non-white	-.102**	-.161**	-.158**	-.160**	-.152**	-.151**	-.160**	-.152**	-.141**
Degree	.070**	.132**	.161**	.157**	.143**	.154**	.169**	.163**	.169**
Hnd/a	.057**	.083**	.090**	.093**	.092**	.100**	.115**	.121**	.134**
O/cse	.027*	.042*	.055**	.054**	.058**	.064**	.073**	.066**	.075**
Hhsize	-.019**	-.017*	-.031**	-.019*	-.011	-.009	-.0001	.002	-.001
Nch04	.068**	.064**	.089**	.072**	.057**	.044*	.035	.027	.034*
Nch511	.032**	.018	.036*	.012	.011	.014	-.003	-.013	-.005
Nch1218	.041**	.043**	.069**	.051**	.043	.042*	.017	.012	.0008
Age	-.027	-.047	-.041	-.042	-.043	-.030	-.031	-.029	-.013
Age2	.102	.166	.144	.147	.130	.122	.120	.117	.052
Age3	-.146	-.229	-.193	-.188	-.178	-.166	-.156	-.146	-.041
Age4	.071	.106	.085	.076	.074	.068	.055	.043	-.051
North-west	.003	-.009	-.021	-.008	-.013	.002	.016	-.006	-.010
North-east	.029	.002	.011	.024	.030	.027	.029	.024	.030
South-east	.003	.015	.018	.035	.033	.027	.025	.042	.034
South-west	.062**	.074**	.049	.060*	.072*	.062	.072*	.065	.068
Midlands	.009	.025	.024	.038	.040	.035	.035	.026	.026
Scotland	-.028	-.018	-.017	-.010	-.042	-.046	-.077*	-.082	-.086*
Wales	.027	.014	.028	.021	.020	-.002	-.022	-.006	.009

Table 5.3 (continued) Probit models for response/non-response by wave; Men, BHPS

	WAVE								
N = 4543	2	3	4	5	6	7	8	10	11
Self-employed	.006	-.0001	.018	.001	.005	-.002	.002	-.036	-.027
Unemployed	-.017	-.035	-.047	-.092	-.063	-.110	-.100	-.090	-.096
Retired	.006	.025	-.029	-.008	.018	-.044	.004	-.014	-.034
Family-care	.005	-.026	-.0004	-.032	-.030	-.064	-.048	-.019	-.029
Emp-other	.019	.024	.036	.018	.032	-.010	.042	.013	.014
Unclassified	.052	.039	.036	.055	.049	.083	.072	.060	.066
Manual-technical	.042	.046	.018	.017	.039	.033	.033	.044	.033
Skill-non-man	.041	.053	.058	.039	.045	.053	.065	.091*	.077
Skilled-manual	.030	.024	.018	.013	.016	.023	.036	.052	.041
Part-skilled	.040	.038	.057	.023	.038	.039	.045	.046	.034
Unskilled	.058	.073	.068	-.0008	.014	.017	.043	-.007	.022
SAHEx	.006	.019	.010	.010	-.001	.003	.003	.013	.009
SAHFair	.003	-.007	-.027*	-.044*	-.060**	-.051*	-.060**	-.066**	-.053*
SAHPoor	.015	.022	.020	.010	-.011	-.022	-.032	-.072	-.076
SAHVPoor	-.032	-.033	-.043	-.059	-.112	-.120	-.137*	-.151*	-.129
HLPRB	.002	.026	.049**	.054**	.049**	.045**	.045**	.040	.037*
HLDSBL	-.014	-.002	-.008	-.027	-.011	.021	.002	-.024	-.023
HLLT	-.025	-.044	-.049	-.087**	-.088**	-.095**	-.097**	-.062*	-.071*
Log-likelihood	-1660.9	-2293.5	-2472.3	-2671.8	-2730.1	-2677.6	-2846.2	-2897.1	-2927.8

1. * denotes $p \leq 0.05$, ** denotes $p \leq 0.01$

2. All regressors represent wave 1 responses, there were no men in the long-term sick category in our sample at wave 1.

3. Results are presented as partial effects on the probability of responding at wave t, evaluated at the sample means of the regressors.

0.202. Non-whites are less likely to remain in the sample, and this effect increases in magnitude as time progresses. By wave 11 the probability of responding among non-white men is 0.141 lower and among women it is 0.175 lower. There is no clear evidence of statistically significant income-related non-response.

The pattern of health-related non-response shows that, for both men and women, very poor initial health (SAHVPOOR) is associated with lower response rates, as is functional limitations (HLLT). These associations grow in magnitude and attain statistical significance as the panel lengthens. Disability (HLDSBL) does not show a clear-cut pattern in the multivariate analysis and health problems (HLPRB) shows that those who report health problems are more likely to respond at all waves. This may be because the variable HLPRB encompasses some relatively minor ailments – such that the majority of the sample report having at least one of them – and, after controlling for other measures of health this variable may be capturing other forms of non-response such as geographic mobility among healthy young people and the ease of making contact with interviewees.

5.3.2 ECHP

Table 5.4 reports the overall drop-out rates across the available waves for each of the countries that participated in the ECHP, along with comparable samples from the

Table 5.4 Summary of SAH-related drop-out rates (%) over all available waves by country, ECHP-UDB

Country (waves)	All	VGood	Good	Fair	Poor	VPoor
Austria (2-8)	41	41	38	40	50	65
Belgium (1-8)	48	44	46	54	61	73
Denmark (1-8)	49	44	47	58	66	75
Finland (3-8)	47	43	45	50	55	65
France (1-8)	44	42	42	44	53	61
Germany (1-3)	13	15	11	13	23	36
Germany (gsoep 1-8)	33	34	30	31	37	50
Greece (1-8)	41	39	41	40	46	59
Ireland (1-8)	69	69	68	68	73	78
Italy (1-8)	40	37	39	40	49	59
Luxembourg (1-3)	12	13	12	10	15	37
Netherlands (1-8)	44	42	43	48	57	63
Portugal (1-8)	30	29	26	29	35	51
Spain (1-8)	49	47	48	49	53	65
UK (1-3)	45	43	44	48	52	56
UK (BHPS 1-8)	29	27	26	32	35	53

BHPS and German Socioeconomic Panel (GSOEP) that are included in the ECHP-UDB. The evidence reinforces earlier studies (Peracchi, 2002; Behr *et al*., 2002; Behr, 2004). In particular the UK and Ireland stand out as having above average rates of drop-out, with 45% drop-out after only three waves in the UK and 69% after eight waves in Ireland. The high non-response in the UK is largely attributable to the decision by the national data unit (NDU) to follow only households with complete sets of personal interviews, rather than adopting the standard ECHP following rules. For the other countries that participated for the full eight waves overall drop-out rates are broadly similar, ranging from 40% in Italy to 49% in Spain. Germany and Luxembourg only participated for waves 1-3 and have low drop-out rates of 13% and 12%. The drop-out rates over the comparable period of 29% for the BHPS and 33% for the GSOEP are lower than the ECHP, reflecting the fact that these samples were established prior to 1994. As in the BHPS there is evidence of health-related non-response in the ECHP. When the samples are split by initial levels of self-assessed health a consistent pattern emerges across all countries, with higher rates of non-response among those in poor or very poor initial health. The gradient is not always monotonic, in some countries the lowest drop-out rates are for those with good or, in the case of Luxembourg, fair health.

To provide a sense of how drop-out rates vary by socioeconomic characteristics, Table 5.5 shows the overall drop-out rates across the available waves and across the countries split by socioeconomic characteristics at the first wave. The table shows the drop-out rates across all available waves for the upper and lower categories of income and education. However the results should be treated with caution as the number of observations in some of the cells are quite small. Patterns of overall non-response by income and education are different across countries: with a positive income gradient in Greece, Ireland, Italy, Portugal and Spain and a positive education gradient in Belgium, Denmark, Finland, France Germany, Netherlands and the UK. Generally the pattern of health-related drop-outs are similar across income and education groups, taking account of the small cell sizes in some cases for very poor health.

Table 5.5 Summary of SAH-related drop-out rates(%) over all available waves by income quintile and educationalstatus by country, ECHP-UDB

	All	VGood	Good	Fair	Poor	VPoor
Austria (2-8)						
Income quintile 1	43	41	41	40	51	60
Income quintile 5	43	45	42	38	37	67
Primary education	45	48	43	44	33	50
Tertiary education	42	41	38	40	53	64
Belgium (1-8)						
Income quintile 1	57	54	54	60	66	74
Income quintile 5	43	41	40	51	66	71
Primary education	41	39	40	46	47	71
Tertiary education	53	45	50	57	68	76
Denmark (1-8)						
Income quintile 1	60	54	59	65	72	83
Income quintile 5	37	35	38	48	50	83
Primary education	37	35	36	48	55	69
Tertiary education	60	57	57	63	71	78
Finland (3-8)						
Income quintile 1	49	45	45	52	58	65
Income quintile 5	44	43	42	46	57	67
Primary education	41	40	41	44	45	100
Tertiary education	53	49	52	53	59	68
France (1-8)						
Income quintile 1	49	49	46	48	59	63
Income quintile 5	40	37	38	43	52	67
Primary education	37	34	35	39	45	65
Tertiary education	46	43	42	46	55	61
Germany (1-3)						
Income quintile 1	18	20	15	20	24	32
Income quintile 5	12	14	10	11	19	38
Primary education	11	15	9	10	14	35
Tertiary education	15	16	12	15	24	36
Greece (1-8)						
Income quintile 1	35	34	31	35	41	57
Income quintile 5	48	46	48	53	59	67
Primary education	49	47	52	57	56	100
Tertiary education	36	29	34	37	45	57

Table 5.5 (continued) Summary of SAH-related drop-out rates(%) over all available waves by income quintile and educational status by country, ECHP-UDB

	All	VGood	Good	Fair	Poor	VPoor
Ireland (1-8)						
Income quintile 1	66	65	66	66	71	72
Income quintile 5	71	70	71	79	57	100
Primary education	70	69	69	73	89	.
Tertiary education	66	65	64	68	72	80
Italy (1-8)						
Income quintile 1	37	33	33	37	44	57
Income quintile 5	45	46	45	42	55	65
Primary education	40	39	40	39	48	100
Tertiary education	40	36	38	40	48	58
Luxembourg (1-3)						
Income quintile 1	19	28	17	15	7	42
Income quintile 5	13	11	17	10	14	50
Primary education	12	7	17	11	33	0
Tertiary education	12	15	11	9	13	37
Netherlands (1-8)						
Income quintile 1	47	46	43	51	63	68
Income quintile 5	44	39	42	52	53	67
Primary education	40	35	39	50	38	60
Tertiary education	53	48	51	55	63	65
Portugal (1-8)						
Income quintile 1	31	24	23	29	38	54
Income quintile 5	37	47	36	35	41	56
Primary education	40	35	38	44	58	100
Tertiary education	28	25	23	28	35	50
Spain (1-8)						
Income quintile 1	48	43	48	49	50	62
Income quintile 5	49	51	48	46	61	68
Primary education	51	51	50	52	60	92
Tertiary education	49	45	47	49	53	65
Uk (1-3)						
Income quintile 1	49	46	49	50	53	51
Income quintile 5	42	41	42	44	56	78
Primary education	38	38	38	39	40	62
Tertiary education	48	45	47	50	55	55

5.4 MODELS AND ESTIMATION METHODS

5.4.1 The ordered probit model

To model the association between the current level of self-assessed health (SAH) and socioeconomic status (SES) we use pooled ordered probit specifications of a dynamic model (see e.g., Contoyannis *et al.,* 2004b). The latent variable specification of the model can be written as:

$$h^*_{it} = \beta' x_{it} + \varepsilon_{it} \qquad (i=1,...,N;\ t=2,...T_i) \qquad (5.1)$$

where i denotes individuals and t denotes the waves of the panel; h^*_{it} is a latent variable that underlies reported SAH; x_{it} is a set of regressors, that includes dummy variables for each category of SAH in the previous year (to capture dynamics), along with observed socioeconomic variables; and ε_{it} is a time and individual-specific error term, which is assumed to be normally distributed. The pseudo-ML estimator of the pooled ordered probit model is consistent even if the error terms are not serially independent and does not require that the regressors are strictly exogenous, so it can accommodate pre-determined variables (see e.g., Wooldridge, 2002b). This makes the estimator more robust in comparison to a random effects specification. A robust estimator of the covariance matrix is used to allow for clustering within individuals. As we do not have a natural scale for the latent variable the variance of the error term (ε) is restricted to equal one.

In our data the latent outcome h^*_{it} is not observed; instead, we observe an indicator of the category in which the latent indicator falls (h_{it}). The observation mechanism can be expressed as,

$$h_{it} = j \quad if \ \ \mu_{j-1} < h^*_{it} \leq \mu_j \ , \quad j = 1,.....,m \qquad (5.2)$$

where $\mu_0 = -\infty$, $\mu_j \leq \mu_{j+1}$, $\mu_m = \infty$. Given the assumption that the error term is normally distributed, the probability of observing the particular category of SAH reported by individual i at time t (h_{it}), conditional on the regressors is,

$$P_{itj} = P(h_{it} = j) = \Phi(\mu_j - \beta' x_{it}) - \Phi(\mu_{j-1} - \beta' x_{it}) \qquad (5.3)$$

where $\Phi(.)$ is the standard normal distribution function. This formulation makes it clear that it is not possible to separately identify an intercept in the linear index (β_0) and the cutpoints (μ), the model only identifies ($\mu_j - \beta_0$). To deal with this we have adopted a conventional normalisation by setting β_0=0 (an alternative is to set μ_1=0).

We do not impose an explicit error components specification in equation (5.1), but to understand the nature of the non-response problem, it will often be helpful to think in terms of time invariant unobservable heterogeneity (an “individual effect”)

and idiosyncratic random shocks that vary over time ("health shocks"). Non-response associated with individual effects implies that there are certain "types" of individual who are prone to drop out of the panel and whose health is permanently different from those who stay in. This kind of non-response can therefore be detected by comparing the outcomes that are observed prior to drop-out. Non-response associated with idiosyncratic shocks is more problematic. A transient health shock would be reflected in h^*_{it} , and hence in h_{it}, but not necessarily in past health. The fundamental identification problem arises if the transient shock leads to the individual dropping-out of the panel, as h_{it} is unobservable for those who have dropped-out.

5.4.2 Non-response bias

Testing

The descriptive analysis has shown evidence of systematic patterns of non-response by socioeconomic characteristics and previous levels of health, but it remains to be seen whether this will lead to *non-response bias* in our empirical models of SAH. To provide an initial test for non-response bias we use the simple variable addition test proposed by Verbeek and Nijman (1992, p.688). The test variable we use is a count of the number of waves that are observed for the individual (NUMBER OF WAVES). This is added to our pooled ordered probit model and estimated with the unbalanced sample. The t-ratio on the added variables provides a test for non-response bias. The intuition behind the test is that, if non-response is random, indicators of an individual's pattern of survey responses (R) should not be associated with the outcome of interest (h) after controlling for the observed covariates (x): in other words, it tests a conditional independence condition $E(h|x,R)=E(h|x)$. Additional evidence can be provided by Hausman-type tests that compare estimates from the balanced - for whom we have complete information at all waves - and unbalanced - for whom we have incomplete information for some individuals - samples. In the absence of non-response bias these estimates should be comparable, but non-response bias may affect the unbalanced and balanced samples differently leading to a contrast between the estimates. It should be noted that the variable addition tests and Hausman-type tests may have low power; they rely on the sample of observed outcomes for h_{it} and will not capture non-response associated with idiosyncratic shocks that are not reflected in observed past health (Nicoletti, 2002).

Estimation

To allow for non-response we adopt an inverse probability weighted (IPW) estimator and apply it to the pooled ordered probit model (Robins *et al.,* 1995; Fitzgerald *et al.,* 1998; Moffitt *et al.,* 1999; Wooldridge, 2002a, 2002b). This approach is grounded in the

notion of missing at random or ignorable non-response (Rubin, 1976; Little and Rubin, 1987). Using *R* as an indicator of response (R=1 if observed, 0 otherwise) and *h* and *x* as the outcome and covariates of interest: missing completely at random (MCAR) is defined by $P(R=1|h,x)=P(R=1)$ and missing at random (MAR) is defined by $P(R=1|h,x)=P(R=1|x)$. The latter implies that, after conditioning on observed covariates, the probability of non-response does not vary systematically with the outcome of interest. By Bayes rule, the MAR condition can be inverted to give $P(h|x,R=1)=P(h|x)$, which provides a rationale for the Verbeek and Nijman (1992) approach to testing.

Fitzgerald *et al.* (1998) extend the notion of ignorable non-response by introducing the concepts of selection on observables and selection on unobservables. This requires an additional set of observables, z, that are available in the data but not included in the regression model for *h*. Selection on observables is defined by Fitzgerald *et al.* by the conditional independence condition $P(R=1|h,x,z)=P(R=1|x,z)$. Selection on unobservables occurs if this conditional independence assumption does not hold. Selection on unobservables, also termed informative, non-random or non-ignorable non-response, is familiar in the econometrics literature where the dominant approach to non-response follows the sample selection model (Heckman, 1976; Hausman and Wise, 1979). This approach relies on the z being "instruments" that are good predictors of non-response and that satisfy the exclusion restriction $P(h|x,z)=P(h|x)$. This is quite different from the selection on observables approach that seeks *z*'s which are endogenous to *h*. The statistics literature has related methods for non-ignorable non-response, some of which use the EM algorithm for data imputation (see e.g., Diggle and Kenward, 1994; Fitzmaurice *et al.*, 1996; Molenberghs *et al.*, 1997). Also it is worth mentioning that linear fixed effects panel estimators are consistent, in the presence of selection on unobservables, so long as the non-ignorable non-response is due to time invariant unobservables (see e.g., Verbeek and Nijman, 1992).

The validity of the selection on observables approach hinges on whether the conditional independence assumption holds and non-response can be treated as ignorable, once *z* is controlled for. If the condition does hold, consistent estimates can be obtained by weighting the observed data by the inverse of the probability of response, conditional on the observed covariates (Robins *et al.*, 1995). This gives more weight to individuals who have a high probability of non-response, as they are under-represented in the observed sample.

Fitzgerald *et al.* (1998) make it clear that this approach will be applicable when interest centres on a structural model for $P(h|x)$ and that the *z*'s are deliberately excluded from the model, even though they are endogenous to the outcome of interest. They suggest lagged dependent variables as an obvious candidate for *z*. Rotnitzky and Robins (1997) offer a similar interpretation when they describe possible candidates for *z* being

intermediate variables in the causal pathway from *x* to *h*. This property implies that it would not be sensible to use solely "field variables" such as changes in interviewer as candidates for the additional observables (see e.g., Behr *et al.,* 2002). These kinds of variables may be good predictors of non-response but are unlikely to be associated with SAH. Horowitz and Manski (1998) show that if the observables (*z*) are statistically independent of *h*, conditional on (*x, R=1*), then the weighted estimates reduce to the unweighted ones. This would explain why no difference between weighed and unweighted estimates may be reported in empirical analyses that use inappropriate variables for *z*.

In our application we are interested in the distribution of self-assessed health conditional on socioeconomic status, rather than the distribution conditional on socioeconomic status and on other indicators of morbidity. We use past morbidity among our *z* variables. Of course, this approach will break-down if an individual suffers an unobserved health shock, that occurs after their previous interview, that leads them to drop out of the survey and that is not captured by conditioning on lagged measures of morbidity. In this case non-response would remain non-ignorable even after conditioning on *z*. It is possible to test the validity of the selection on observables approach. The first step is to test whether the *z*'s do predict non-response; this is done by testing their significance in the probit models for non-response at each wave of the panel (as in Table 5.3). The second is to do Hausman-type tests to compare the coefficients from the weighted and unweighted estimates. In addition the ordered probit models are compared in terms of the magnitudes of estimated average partial effects.

Implementation of the Fitzgerald *et al.* (1998) form of the ignorability condition implies that *x* is observable when *R=0*. In the case of the kind of unit non-response we are dealing with in the BHPS and ECHP, non-response means that there is missing data for the current period covariates (*x*) as well as self-assessed health (*h*). So we implement a stronger form of conditional independence $P(R=1|h,x,z)=P(R=1|z)$ as proposed by Wooldridge (2002a). To compute the IPW estimator we estimate (probit) equations for response ($R_{it}=1$) versus non-response ($R_{it}=0$) at each wave, t=2,...,T, conditional on a set of characteristics (z_{i1}) that are measured for all individuals at the first wave (as in Table 5.3). As described above, this relies on selection on observables and implies that non-response can be treated as ignorable non-response, conditional on z_{i1} (Fitzgerald *et al.*, 1998; Wooldridge, 2002b, p.588). Selection on observables requires that z_{i1} contains variables that predict non-response and that are correlated with the outcome of interest (SAH) but which are deliberately excluded from the model for health.

In practice z_{i1} includes the initial values of all of the regressors in the health equation. Also it includes initial values of SAH and of the other indicators of morbidity: for the BHPS, whether the individual reports a specific health problem (HLPRB), whether they

report that health limits their daily activities (HLLT) and whether they report a disability (HLDSBL); for the ECHP, whether the individual was mildly or severely hampered in their daily activities. In addition, z_{i1} includes initial values of the respondent's activity status, occupational socioeconomic group and region. The probits for response/non-response are estimated at each wave of the panel, from wave 2 to wave 11 in the case of the BHPS and waves 2 to 8 for the ECHP, using the full sample of individuals who are observed at wave 1.

The inverse of the fitted probabilities from these models, $1/1/\hat{p}_{it}$, are then used to weight observations in the IPW-ML estimation of the pooled ordered probit model using:

$$LogL = \sum_{i}^{n} \sum_{t}^{T} \left(R_{it} / \hat{p}_{it} \right) LogL_{it} \qquad (5.4)$$

Wooldridge (2002a) shows that, under the ignorability assumption:

$$P(R_{it} = 1 | h_{it}, x_{it}, z_{i1}) = P(R_{it} = 1 | z_{i1}), \ t=2,...,T \qquad (5.5)$$

the IPW-ML estimator is $\sqrt{n}$-consistent and asymptotically normal. Wooldridge (2002a) also shows that using the estimated $\hat{p}_{it}$ rather than the true p_{it} and ignoring the implied adjustment to the estimated standard errors leads to "conservative inference" so that the standard errors are larger than they would be with an adjustment for the use of fitted rather than true probabilities (see also Robins *et al.,* 1995).

The IPW-ML estimator can be adapted to allow the elements of *z* to be up-dated and change across time, for example adding *z* variables measured at *t-1* to predict response at *t*. This should improve the power of the probit models to predict non-response and hence make the ignorability assumption more plausible. In this case the probit model for non-response at wave *t* is estimated relative to the sample that is observed at wave *t-1.* This relies on non-response being an absorbing state and is therefore confined to "monotone attrition" where respondents never re-enter the panel. Also, because estimation at each wave is based on the selected sample observed at the previous wave, the construction of inverse probability weights has to be adapted. The predicted probability weights are constructed cumulatively using $\hat{p}_{it} = \hat{\pi}_{i2} \times \hat{\pi}_{i2} \ldots \times \hat{\pi}_{i2}$, where the $\hat{\pi}_{it}$ denote the fitted selection probabilities from each wave. In this version of the estimator the ignorability condition has to be extended to include future values of *h* and *x* (see Wooldridge, 2002b, p. 589). Once again Wooldridge shows that omitting a correction to the asymptotic variance estimator leads to conservative inference.

We have not pursued a selection on unobservables approach in this paper. This stems from the lack of credible exclusion restrictions that would define variables that predict health-related non-response but are not associated with SAH. Also, the use of fixed effects estimators is not possible for probit and ordered probit models, due to the

incidental parameters problem (although we have experimented with models that use Mundlak (1978) type specifications to deal with correlated effects and this had little impact on our findings concerning non-response (see also Contoyannis *et al.,* 2004b). With the public use versions of the BHPS and ECHP-UDB we do not have any scope for using methods based on refreshment samples (see e.g., Dolton, 2004). The IPW approach is attractive as it is easy to apply in the context of nonlinear models, such as the ordered probit model, and only requires a re-weighting of the data. In contrast to the published longitudinal weights that are supplied with the BHPS and ECHP, our IPW weights are model-specific and specifically designed for the outcome of interest (SAH) and the associated problem of health-related non-response. Past values of SAH, along with other indicators of morbidity, provide promising candidates for the *z*-variables; although the validity of the approach depends on the credibility of the ignorability assumption. For comparison, we present estimates based on the published BHPS and ECHP weights alongside estimates based on our own weights.

5.5 ESTIMATION RESULTS

The results for the various model specifications outlined above are reported in this section. For the detailed analysis of the BHPS, models for men and women are presented separately throughout. For the more parsimonious analysis of the ECHP the samples are pooled.

5.5.1 Tests for non-response

Table 5.6 presents the Verbeek and Nijman (1992) variable addition tests for non-response bias in the pooled ordered probit model for SAH in the BHPS and ECHP. This is based on adding the NUMBER OF WAVES to the model. The first set of results are for the benchmark pooled ordered probits with covariates *x*. The second includes the additional observables (*z*) that are used to compute the inverse probability weights. The latter can be regarded as a test for the ignorability assumption behind the IPW estimator. With the exception of Austria and Luxembourg, all of the test statistics show evidence of non-response bias. Adding these test variables to the model is not intended to "correct" the estimates for non-response, but it is informative to compare the estimates with the baseline model that does not include the test variables. It is striking that, for key variables such as income and education, the differences between the estimated coefficients are small (these results are available on request).

5.5.2 Estimates of ordered probits for SAH

Table 5.7 presents the coefficient estimates for the dynamic pooled ordered probits for SAH estimated with the BHPS data for men (the results forwomen are available from

Table 5.6 Verbeek and Nijman tests for non-response, BHPS and ECHP-UDB

Country (waves)	Model based on regressors *x* (t ratio)	Model based on regressors *x* (t ratio)
BHPS (IPW-1)		
Men:	6.42	6.27
Women:	3.53	3.24
BHPS (IPW-2)		
Men:	4.76	4.95
Women:	2.03	2.10
Echp (ipw-2)		
Austria (2-8)	0.39	0.27
Belgium (1-8)	5.45	4.16
Denmark (1-8)	4.02	3.15
Finland (3-8)	5.28	4.74
France (1-8)	6.03	4.95
Germany (1-3)	3.15	2.79
Greece (1-8)	3.69	2.80
Ireland (1-8)	6.81	6.27
Italy (1-8)	6.54	5.82
Luxembourg (1-3)	-0.64	-0.44
Netherlands (1-8)	3.22	2.03
Portugal (1-8)	7.25	5.89
Spain (1-8)	6.07	4.88
UK (1-3)	2.04	1.94

the authors on request). The models were estimated on the balanced, the unbalanced sample and the available observations for the sample of drop-outs. The estimates for the pooled ordered probit models allow for clustering within individuals in the errors by using a robust estimator of the covariance matrix. In addition we estimated the model using the published longitudinal weights supplied with the BHPS, along with our own inverse probability weights (IPW) to adjust for non-response. Both variants of the IPW-ML estimator are presented: IPW-1 uses the full sample and wave 1 regressors to predict non-response, with IPW-2 the sample is restricted by excluding observations that exhibit non-monotone attrition and previous period regressors are used. The unbalanced sample is selected so that the same observations are used for the unweighted and all of the unweighted estimators, to allow a direct comparison of the estimators.

The LR tests reject the poolability of the unweighted models for the balanced and drop-outs samples when they are compared to the combined results for the unbalanced panel. The coefficients on lagged SAH show a clear gradient in the magnitude of the coefficients, running from good to very poor SAH (excellent is the omitted category),

across all of the models. The results in Table 5.7 show differences in the sign and size of the coefficients on the age variables and on WIDOW between the three samples and between the weighted and unweighted estimates for the unbalanced sample, reflecting age-related non-response. The sign and size of the coefficients on the education variables are similar across all samples and comparing the weighted and unweighted estimates. The coefficients for ln(INCOME) are also similar across all of the specifications, but with a larger coefficient for the balanced sample than the sample of drop-outs and with the unbalanced sample bracketed between them. Pairwise comparisions of the contrasts between weighted and unweighted estimates of the coefficients on ln(INCOME) and the education variables shows that the differences are small in magnitude, relative to the size of the coefficients, with the largest differences when the IPW-2 weights are used. Pairwise Hausman-type t-tests suggest that the differences between coefficients from the unweighted, BHPS-weighted and IPW-1 weighted estimates are not significantly different. But they do reject the null that there is no statistically significant difference between the unweighted estimates and the IPW-2 estimates. Although this may be due, in part, to the smaller sample size when the sample is restricted to monotone attrition.

Average partial effects

The scaling of the ordered probit coefficients is arbitrary. To provide an indication of the magnitude of the associations between SAH and the regressors we present average partial effects (APEs). For continuous regressors, such as income, these are obtained by taking the derivative of the ordered probit probabilities with respect to the variable in question. For discrete regressors, such as the educational qualifications, they are obtained by taking differences. In general, average partial effects are averaged over the population distribution of heterogeneity and computed using the population averaged parameters (see e.g., Wooldridge, 2002b). In the pooled ordered probit models the total error variance is normalised to 1 and the estimated bs are population averaged parameters by default, so the APEs are given by the standard formula for partial effects.

In the ordered probit model it is possible to compute APEs for each of the five categories of self-assessed health. For parsimony, Table 5.8 summarises the APEs of lagged SAH, income and educational attainment on the probability of reporting excellent health in the BHPS data. In this case the sign of the APE has a clear qualitative interpretation, with a positive sign implying a positive association with health and vice versa. A partial effect is computed for each observation in the sample, evaluated at the observed values of the regressors. The table presents the sample mean of the partial effects – the APE – along with the sample standard deviation, in parentheses, to give a sense of the variability

Table 5.7 Dynamic ordered probit models for SAH – Men, BHPS

	-1	-2	-3	-4	-5	-6
	Balanced sample	Drop-outs sample	Unbalanced sample	BHPS longitudinal weights	Inverse probability weights	
					IPW-1	IPW-2
	NT = 18,616	NT = 6,593	NT = 25,209	NT = 25,209	NT = 25,209	NT=21,630
SAHGood(t-1)	-.981 (.029)	-.939 (.050)	-.970 (.025)	-.954 (.026)	-.966 (.025)	-.953 (.030)
SAHFair(t-1)	- 1.867 (.039)	- 1.794 (.064)	- 1.845 (.033)	- 1.820 (.035)	- 1.841 (.034)	- 1.808 (.041)
SAHPoor(t-1)	- 2.757 (.062)	- 2.597 (.081)	- 2.723 (.049)	- 2.701 (.051)	- 2.725 (.050)	- 2.677 (.065)
SAHVPoor(t-1)	- 3.356 (.126)	- 3.309 (.124)	- 3.383 (.089)	- 3.385 (.086)	- 3.414 (.087)	- 3.305 (.114)
Ln(Income)	.146 (.016)	.103 (.023)	.130 (.013)	.125 (.015)	.129 (.015)	.149 (.016)
Widow	.004 (.055)	.048 (.070)	.015 (.044)	-.012 (.044)	.011 (.047)	.012 (.057)
Single	-.079 (.033)	.004 (.049)	-.067 (.027)	-.066 (.028)	-.075 (.028)	-.083 (.033)
Div/sep	.046 (.048)	-.063 (.063)	-.003 (.039)	-.004 (.042)	-.009 (.041)	.014 (.048)
Non-white	-.174 (.048)	-.161 (.072)	-.165 (.040)	-.181 (.041)	-.160 (.041)	-.118 (.057)
Degree	.180 (.035)	.166 (.053)	.177 (.029)	.182 (.030)	.180 (.030)	.170 (.034)
HND/A	.144 (.028)	.156 (.045)	.153 (.024)	.163 (.025)	.159 (.024)	.130 (.027)
O/CSE	.110 (.029)	.145 (.039)	.121 (.023)	.124 (.024)	.120 (.024)	.103 (.027)
HHSize	.030 (.013)	-.018 (.018)	.013 (.010)	.004 (.011)	.015 (.011)	.019 (.013)
NCH04	.001 (.025)	.019 (.043)	.007 (.022)	.012 (.024)	.013 (.022)	.004 (.025)
NCH511	-.009 (.020)	.046 (.034)	.008 (.017)	.008 (.018)	.0005 (.017)	.008 (.022)
NCH1218	.026 (.023)	.035 (.041)	.030 (.020)	.038 (.021)	.022 (.020)	.041 (.025)
Age	-.013 (.045)	-.024 (.054)	-.030 (.034)	-.022 (.037)	-.041 (.039)	-.014 (.042)
Age2	.028 (.144)	.100 (.170)	.081 (.107)	.054 (.117)	.120 (.125)	.025 (.134)
Age3	-.035 (.193)	-.180 (.224)	-.103 (.142)	-.069 (.156)	-.159 (.168)	-.024 (.179)
Age4	.020 (.093)	.102 (.104)	.046 (.067)	.031 (.074)	.074 (.081)	.007 (.086)
Cut 1	- 2.789 (.531)	- 2.956 (.634)	- 3.009 (.400)	- 2.980 (.435)	- 3.136 (.459)	- 2.664 (.495)
Cut 2	- 1.728 (.531)	- 2.040 (.633)	- 2.019 (.400)	- 1.984 (.434)	- 2.148 (.458)	- 1.667 (.491)
Cut 3	-.561 (.529)	-.952 (.633)	-.883 (.400)	-.849 (.434)	- 1.004 (.457)	-.548 (.490)
Cut 4	1.117 (.529)	.641 (.633)	.770 (.399)	.789 (.433)	.643 (.457)	1.092 (.490)
Log Likelihood	- 18568.9	- 7139.7	- 25782.0	- 26058.2	- 26041.9	- 22450.2
LR test for pooling			146.8 (0.000)			
Hausman test:						
Ln(INCOME)				-.668	-.134	2037
Degree				.651	.391	-.394
HND/A				1429	1034	-1859
O/CSE				.438	-.146	-1273

1. SEs are reported in parentheses, 2. Cut 1-4 are estimated cut points or thresholds, 3. Coeff. for the year dummies are not reported, 4. Descriptive summary of BHPS longitudinal weights: Mean = 1.056, SD = 0.351, Min = 0.190, Max = 2.5, 5. Descriptive summary of IPW–1 with health variables weights: Mean = 1.440, SD = 0.388, Min = 1.01, Max = 15.73, 6. Descriptive summary of IPW–2 with health variables weights: Mean = 1.908, SD = 1.173, Min = 1.01, Max = 33.89, 7. The LR test for pooling compares the unrestricted estimates (balanced+drop-outs samples) with the restricted estimates (unbalanced sample). The statistic is chi-squared, 8. The Hausman test reports the t-test for pairwise comparisons of the contrast between the weighted estimates of the coeff. (models 4-6) with those from the unweighted estimate (model 3).

of the partial effects across observations. These are presented for all versions of the model. Comparing the balanced sample, drop-outs sample and unbalanced samples gives very similar results, suggesting that non-response does not lead to differences in the estimated APEs. This is reinforced by the fact that the estimates with and without weights are very similar in magnitude. The largest differences are between the unweighted and the IPW-2 estimates in the sample of women for educational qualifications.

Table 5.9 summarises the APEs on the probability of reporting very good health, the highest category in the ECHP, for the lagged SAH variables, estimated with the ECHP-UDB data. While Table 5.10 presents the APEs for ln(INCOME) and education. The tables compare the estimates for the unweighted ordered probit and the weighted (ECHP published weights and IPW-2 weights) ordered probit estimated on the unbalanced sample. The estimates for lagged SAH, in Table 5.9, are very stable across all three estimators in all of the countries. Table 5.10 shows that, in all of the countries, there is a positive association between both income and education and SAH in the unweighted estimates and those based on the published. The average partial effects of income are lowest in Portugal, Italy and Spain and highest in Denmark, Ireland and Luxembourg. The average partial effect of completing tertiary education is lowest in Portugal and France and highest in Ireland and Denmark. As with the BHPS, the quantitative differences between the unweighted and weighted estimates of the average partial effects are small for most countries when the published ECHP weights are used. However there is more difference when the IPW-2 weights are used in place of the published weights and in some cases the partial effects change sign. This occurs in the countries where less than eight waves are available (Austria, Germany, Luxembourg, UK). In all of these cases the underlying coefficients for income and education are not statistically significant.

5.6 DISCUSSION

This analysis shows that there is clear evidence of health-related non-response in both the BHPS and ECHP. In general, individuals in poor initial health are more likely to drop out, although for younger groups non-response is associated with good health. Furthermore, variable addition tests provide evidence of non-response bias in the models of SAH. Nevertheless a comparison of estimates based on the balanced samples, the unbalanced samples and corrected for non-response using inverse probability weights shows that, in many cases, substantive differences in the magnitudes of the average partial effects of lagged health, income and education are small. The largest differences in the BHPS results are for the comparison of the weighted and IPW-2 weighted estimates of the APEs for education among women. For the ECHP the estimates of dynamics are unaffected by weighting but the IPW-2 estimates for income and

Table 5.8 Average partial effects (APE) on the probability of reporting excellent SAH, BHPS

	-1 Balanced sample	-2 Drop-outs sample	-3 Unbalanced sample	-4 BHPS longitudinal weights	-5 Inverse probability weights	-6
					IPW-1	IPW-2
Men						
SAHGood(t-1)	-.277 (.141)	-.234 (.144)	-.266 (.143)	-.262 (.140)	-.265 (.142)	-.264 (.139)
SAHFair(t-1)	-.337 (.178)	-.297 (.178)	-.327 (.179)	-.325 (.176)	-.326 (.178)	-.326 (.176)
SAHPoor(t-1)	-.295 (.205)	-.260 (.197)	-.286 (.203)	-.286 (.201)	-.286 (.203)	-.289 (.202)
SAHVPoor(t-1)	-.276 (.208)	-.238 (.200)	-.267 (.207)	-.267 (.205)	-.266 (.206)	-.270 (.206)
Ln(Income)	.038 (.018)	.024 (.014)	.033 (.016)	.032 (.016)	.033 (.016)	.038 (.019)
Degree	.049 (.021)	.040 (.022)	.046 (.022)	.048 (.022)	.047 (.022)	.045 (.021)
HND/A	.038 (.017)	.037 (.021)	.040 (.019)	.042 (.020)	.041 (.020)	.034 (.016)
O/CSE	.029 (.013)	.034 (.019)	.031 (.015)	.032 (.015)	.031 (.015)	.027 (.013)
Women						
SAHGood(t-1)	-.252 (.148)	-.197 (.133)	-.239 (.146)	-.240 (.146)	-.240 (.146)	-.237 (.141)
SAHFair(t-1)	-.296 (.177)	-.250 (.157)	-.286 (.172)	-.284 (.173)	-.287 (.173)	-.283 (.168)
SAHPoor(t-1)	-.246 (.191)	-.216 (.173)	-.239 (.187)	-.237 (.187)	-.240 (.187)	-.242 (.185)
SAHVPoor(t-1)	-223 (.191)	-.195 (.174)	-.216 (.187)	-.215 (.187)	-.217 (.188)	-.220 (.186)
Ln(Income)	.028 (.015)	.010 (.006)	.023 (.013)	.022 (.013)	.021 (.012)	.017 (.009)
Degree	.057 (.029)	.059 (.034)	.057 (.030)	.062 (.032)	.061 (.032)	.071 (.036)
HND/A	.033 (.018)	.055 (.033)	.040 (.021)	.039 (.021)	.040 (.022)	.053 (.028)
O/CSE	.040 (.021)	.027 (.017)	.037 (.021)	.039 (.022)	.040 (.022)	.049 (.026)

1. The partial effects are computed for each individual using their observed values of the regressors. The table presents the sample mean of the partial effects – the APE – along with the sample standard deviations in parentheses.

Table 5.9 Average partial effects on the probability of reporting very good SAH, ECHP-UDB

Country (waves)	Unbalanced sample	ECHP published weights	Inverse probability weights IPW-2
Austria (w2-8, n*t 26368)			
Very poor health	-0.172 (0.190)	-0.171 (0.190)	-0.170 (0.183)
Poor health	-0.187 (0.189)	-0.185 (0.190)	-0.184 (0.183)
Fair health	-0.270 (0.171)	-0.269 (0.173)	-0.266 (0.166)
Good health	-0.208 (0.155)	-0.213 (0.158)	-0.202 (0.149)
Belgium (w1-8, n*t 31699)			
Very poor health	-0.209 (0.206)	-0.212 (0.207)	-0.227 (0.209)
Poor health	-0.225 (0.206)	-0.228 (0.208)	-0.244 (0.210)
Fair health	-0.301 (0.192)	-0.304 (0.193)	-0.323 (0.193)
Good health	-0.268 (0.164)	-0.270 (0.165)	-0.272 (0.162)
Denmark (w1-8, n*t 26848)			
Very poor health	-0.442 (0.264)	-0.440 (0.262)	-0.455 (0.269)
Poor health	-0.453 (0.248)	-0.450 (0.245)	-0.462 (0.248)
Fair health	-0.466 (0.176)	-0.460 (0.173)	-0.460 (0.172)
Good health	-0.255 (0.117)	-0.249 (0.113)	-0.257 (0.113)
Finland (w3-8, n*t 34439)			
Very poor health	-0.309 (0.270)	-0.308 (0.267)	-0.351 (0.282)
Poor health	-0.325 (0.259)	-0.324 (0.256)	-0.366 (0.266)
Fair health	-0.364 (0.210)	-0.364 (0.207)	-0.396 (0.202)
Good health	-0.260 (0.154)	-0.255 (0.151)	-0.254 (0.142)
France (w1-8, n*t 66988)			
Very poor health	-0.117 (0.116)	-0.116 (0.116)	-0.111 (0.108)
Poor health	-0.116 (0.116)	-0.115 (0.115)	-0.110 (0.107)
Fair health	-0.183 (0.111)	-0.181 (0.110)	-0.168 (0.100)
Good health	-0.124 (0.099)	-0.124 (0.099)	-0.107 (0.085)
Germany (w1-3, n*t 16403)			
Very poor health	-0.127 (0.142)	-0.128 (0.143)	-0.124 (0.134)
Poor health	-0.137 (0.143)	-0.139 (0.144)	-0.131 (0.132)
Fair health	-0.188 (0.137)	-0.191 (0.139)	-0.165 (0.121)
Good health	-0.179 (0.125)	-0.180 (0.126)	-0.121 (0.090)
Greece (w1-8, n*t 63826)			
Very poor health	-0.483 (0.312)	-0.477 (0.306)	-0.505 (0.336)
Poor health	-0.471 (0.264)	-0.463 (0.257)	-0.503 (0.292)
Fair health	-0.417 (0.176)	-0.407 (0.170)	-0.483 (0.213)
Good health	-0.225 (0.112)	-0.219 (0.108)	-0.292 (0.154)

1. The partial effects are computed for each individual using their observed values of the regressors. The table presents the sample mean of the partial effects – the APE – along with the sample standard deviations in parentheses.

Table 5.9 (continued) Average partial effects on the probability of reporting very good SAH, ECHP-UDB

Country (waves)	Unbalanced sample	ECHP published weights	Inverse probability weights IPW-2
Ireland (w1-8, n*t 37699)			
Very poor health	-0.444 (0.251)	-0.446 (0.251)	-0.452 (0.253)
Poor health	-0.447 (0.239)	-0.448 (0.237)	-0.455 (0.240)
Fair health	-0.456 (0.155)	-0.454 (0.152)	-0.457 (0.152)
Good health	-0.251 (0.100)	-0.251 (0.098)	-0.252 (0.098)
Italy (w1-8, n*t 96509)			
Very poor health	-0.173 (0.180)	-0.174 (0.178)	-0.183 (0.200)
Poor health	-0.190 (0.176)	-0.191 (0.174)	-0.203 (0.197)
Fair health	-0.236 (0.153)	-0.238 (0.151)	-0.264 (0.177)
Good health	-0.173 (0.128)	-0.171 (0.126)	-0.205 (0.154)
Luxembourg (w1-3, n*t 3503)			
Very poor health	-0.214 (0.198)	-0.217 (0.201)	-0.243 (0.212)
Poor health	-0.226 (0.194)	-0.229 (0.197)	-0.252 (0.204)
Fair health	-0.282 (0.160)	-0.286 (0.162)	-0.272 (0.152)
Good health	-0.191 (0.126)	-0.195 (0.128)	-0.183 (0.109)
Netherlands (w1-8, n*t 55673)			
Very poor health	-0.177 (0.164)	-0.178 (0.166)	-0.182 (0.159)
Poor health	-0.189 (0.166)	-0.190 (0.168)	-0.193 (0.159)
Fair health	-0.255 (0.157)	-0.257 (0.159)	-0.256 (0.145)
Good health	-0.232 (0.135)	-0.235 (0.137)	-0.205 (0.119)
Portugal (w1-8, n*t 69236)			
Very poor health	-0.030 (0.054)	-0.033 (0.058)	-0.034 (0.073)
Poor health	-0.042 (0.059)	-0.047 (0.065)	-0.056 (0.085)
Fair health	-0.065 (0.075)	-0.074 (0.083)	-0.088 (0.107)
Good health	-0.076 (0.091)	-0.086 (0.102)	-0.112 (0.134)
Spain (w1-8, n*t 85111)			
Very poor health	-0.173 (0.148)	-0.174 (0.147)	-0.177 (0.148)
Poor health	-0.184 (0.137)	-0.184 (0.136)	-0.187 (0.135)
Fair health	-0.183 (0.110)	-0.183 (0.110)	-0.189 (0.109)
Good health	-0.086 (0.059)	-0.083 (0.057)	-0.071 (0.048)
Uk (w1-3, n*t 12587)			
Very poor health	-0.317 (0.239)	-0.316 (0.239)	-0.321 (0.224)
Poor health	-0.337 (0.233)	-0.335 (0.233)	-0.334 (0.213)
Fair health	-0.397 (0.187)	-0.396 (0.188)	-0.368 (0.161)
Good health	-0.257 (0.146)	-0.257 (0.147)	-0.225 (0.118)

1. The partial effects are computed for each individual using their observed values of the regressors. The table presents the sample mean of the partial effects – the APE – along with the sample standard deviations in parentheses.

education are substantially different than the unweighted estimates in the countries where less than eight waves are available (Austria, Germany, Luxembourg, UK). So, while health-related non-response clearly exists, on the whole it does not appear to distort the magnitudes of the estimated dynamics of SAH and the relationship between socioeconomic status and self-assessed health. Similar findings have been reported concerning the limited influence of non-response bias in models of income dynamics and various labour market outcomes (see e.g., Hausman and Wise, 1979; Becketti *et al.,* 1988; Lillard and Panis, 1998; Zabel, 1998; Ziliak and Kniesner, 1998; Jimenez-Martin and Peracchi, 2002; Behr, 2004) and on measures of social exclusion such as poverty rates and income inequality indices (Watson, 2003; Rendtel *et al.,* 2004). To understand our findings, recall that the descriptive analysis for the BHPS shows little evidence of income-related non-response. There is evidence of strong education-related and health-related non-response, but the latter is concentrated among those in poor initial health who are relatively few in number. There is no clear interaction between health-related non-response and levels of income or education. The finding that non-response has a limited impact on the estimates of health dynamics and, to a lesser extent, estimates of the association between socioeconomic status, measured by income and education, and self-assessed health holds for the BHPS and for many of the countries within the ECHP.

Adams, P., Hurd, M.D, McFadden, D., Merrill, A. and Ribeiro, T. (2003) Healthy, wealthy and wise? Tests for direct causal paths between health and socioeconomic status. Journal of Econometrics, 112, 3-56.

Becketti, S., Gould, W., Lillard, L. and Welch, F. (1988) The Panel Study of Income Dynamics after fourteen years: an evaluation. Journal of Labor Economics, 6, 472-492.

Behr, A., Bellgardt, E. and Rendtel, U. (2002) A cross-national comparison of panel attrition in the ECHP and its impact on the estimation of wage equations. Mimeo.

Behr, A. (2004) Comparing estimation strategies in the presence of panel attrition. Empirical results based on the ECHP. Mimeo, Wilhelms-Universitäüt, Münster.

Benzeval, M. and Judge, K. (2001) Income and health: the time dimension. Social Science and Medicine, 52, 1371-1390.

Benzeval, M., Taylor, J. and Judge, K. (2000) Evidence on the relationship between low income and poor health: Is the Government doing enough? Fiscal Studies, 21, 375-399.

Buckley, N.J., Denton, F.T., Robb, L. and Spencer, B.G. (2004) The transition from good to poor health: an econometric study of the older population. Journal of Health Economics, 23, 1013-1034.

Burström, B. and Fredlund, P. (2001) Self rated health: is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? Journal of Epidemiology and Community Health, 55, 836-840

Contoyannis, P. and Jones, A. (2004) Socioeconomic status, health and lifestyle. Journal of Health Economics, 23, 965-995.

Contoyannis, P., A.M. Jones, and N. Rice (2004a) Simulation-based inference in dynamic panel probit models: an application to health. Empirical Economics, 29, 49-77.

Contoyannis, P., Jones, A.M. and Rice, N. (2004b) The dynamics of health in the British Household Panel Survey. Journal of Applied Econometrics, 19, 473-503.

Deaton, A..S. and Paxson, C.H. (1998) Ageing and inequality in income and health. American Economic Review, Papers and Proceedings 88, 248-253.

Diggle, P. and Kenward, M.G. (1994) Informative drop-out in longitudinal data analysis. Applied Statistics, 43, 49-93.

Dolton, P. (2004) Reducing attrition bias using targeted refreshment sampling and matching. University of Newcastle upon Tyne., mimeo.

Ettner, S. (1996) New evidence on the relationship between income and health. Journal of Health Economics, 15, 67-85

Fitzgerald, J., Gottshalk, P. and Moffitt, R. (1998) An analysis of sample attrition in panel data. The Michigan Panel Study on Income Dynamics. Journal of Human Resources, 33, 251-299.

Fitzmaurice, G.M., Heath, A.F and Clifford, P. (1996) Logistic regression models for binary panel data with attrition. Journal of the Royal Statistical Society Series A(Statistics in Society), 159, 249-263.

Frijters, P., Haisken-DeNew, J.P. and Shields, M.A. (2003) Estimating the causal effect of income on health: evidence from post reunification East Germany, Centre for Economic Policy Discussion Paper No. 465, Australian National University.

Hausman, J. and Wise, D. (1979) Attrition bias in experimental and panel data: the Gary Income Maintenance Experiment. Econometrica, 47, 455-474.

Heckman, J.J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. The Annals of Economic and Social Measurement, 5, 475-492.

Hernández-Quevedo, C., Jones, A.M. and Rice, N., (2004) Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey, ECuity III Project Working Paper #19.

Horowitz, J.L. and Manski, C.F. (1998) Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations. Journal of Econometrics, 84, 37-58.

Idler, E. L. and Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. Journal of Health and Social Behavior, 38, 21-37.

Idler, E.L. and Kasl, S.V. (1995) Self-ratings of health: do they also predict change in functional ability? Journal of Gerontology, 50B, S344-S353.

Jimenez-Martin, S. and Peracchi, F. (2002) Sample attrition and labor force dynamics: evidence from the Spanish labor force survey. Spanish Economic Review, 4, 79-102.

Kenkel, D. (1995). Should you eat breakfast? Estimates form health production functions. Health Economics, 4, 15-29.

Lillard, L.A. and Panis, C.W.A. (1998) Panel attrition from the Panel Study of Income Dynamics: household income, marital status and mortality. Journal of Human Resources, 33, 437-457.

Little, J.A. and Rubin, D.B. (1987) Statistical analysis with missing data. John Wiley and Sons, New York.

Meer, J., Miller, D.L. and Rosen, H.S. (2003) Exploring the health-wealth nexus. Journal of Health Economics, 22, 713-730.

Moffitt, R., Fitzgerald, J. and Gottschalk, P. (1999) Sample attrition in panel data: the role of selection observables. Annales d'Economie et de Statistique, 55-56, 129-152.

Molenberghs, G., Kenward, M.G. and Lesaffre, E. (1997) The analysis of longitudinal ordinal data with non-random drop-out. Biometrika, 84, 33-44.

Mundlak, Y. (1978) On the pooling of time series and cross-section data. Econometrica, 46, 69-85.

Nicoletti, C. (2002) Non-response in dynamic panel data models. Working papers of the Institute for Social and Economic Research, paper 2002-31, Colchester: University of Essex.

Nicoletti, C. and Peracchi, F. (2005) A cross-country comparison of survey nonparticipation in the ECHP. Journal of the Royal Statistical Society Series A forthcoming..

Peracchi, F. (2002) The European Community Household Panel: a review. Empirical Economics, 27, 63-90.

Rendtel, U., Behr, A., Bellgardt, E. and Sisto, J. (2004) Does panel attrition disturb comparative analysis with the European Community Household Panel (ECHP)? Mimeo.

Riphahn, R. (1999) Income and employment effects of health shocks. A test case for the German welfare state. Journal of Population Economics, 12, 363-389.

Robins, J., Rotnitzky, A. and Zhao, L.P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90, 106-121.

Rotnitzky, A. and Robins, J. (1994) Analysis of semi-parametric regression models with non-ignorable non-response. Statistics in Medicine, 16, 81-102.

Rubin, D.B. (1976) Inference and missing data. Biometrika, 63, 581-592.

Salas, C. (2002) On the empirical association between poor health and low socioeconomic status at old age. Health Economics, 11, 207-220.

Smith, J.P. (1999) Healthy bodies and thick wallets: the dual relationship between health and economic status. Journal of Economic Perspectives 13, 145-166.

Taylor, M., Brice, J., Buck, N. and Prentice, E. (1998) British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. University of Essex, Colchester.

Van Doorslaer, E. and Gerdtham, U-G. (2003) Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. Social Science and Medicine, 57, 1621-1629.

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

Van Doorslaer, E., Jones, A.M. and Koolman, X. (2004). Explaining income-related inequalities in doctor utilisation in Europe: a decomposition approach. Health Economics, 13, 629-647.

Van Doorslaer, E., et al. (2000) Equity in the delivery of health care in Europe and the US. Journal of Health Economics, 19, 553-583.

Van Doorslaer, E, Wagstaff, A., Bleichrodt, H. et al, (1997) Income-related inequalities in health: some international comparisons, Journal of Health Economics, 16, 93-112.

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

Verbeek, M. and Nijman, T.E. (1992) Testing for selectivity bias in panel data models. International Economic Review, 33, 681-703.

Verbeek, M. (2000) A guide to modern econometrics. Wiley, Chichester.

Watson, D. (2003) Sample attrition between waves 1 and 5 in the European Community Household Panel. European Sociological Review, 19, 361-378.

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

Wooldridge, J. (2002a) Inverse probability weighted M-estimators for sample stratification, attrition and stratification. Portuguese Economic Journal, 1, 117-139.

Wooldridge, J. (2002b) Econometric Analysis of cross section and panel data. MIT Press.

Zabel, J.E. (1998) An analysis of attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an application to a model of labor market behavior. Journal of Human Resources, 33, 479-506.

Ziliak, J.P. and Kniesner, T.J. (1998) The importance of sample attrition in life cycle labor supply estimation. Journal of Human Resources, 33, 507-530.

6 CHAPTER

Health care inequalities in Europe decomposed

ABSTRACT

This paper presents new international comparative evidence on the factors driving inequalities in the use of GP and specialist services in 12 EU member states. The data are taken from the 1996 wave of the *European Community Household Panel* (ECHP). We examine two types of utilisation (the probability of a visit and the conditional number of positive visits) for two types of medical care: general practitioner and medical specialist visits using probit, truncated Negbin and generalized Negbin models. We find little or no evidence of income-related inequity in the probability of a GP visit in these countries. Conditional upon at least one visit, there is even evidence of a somewhat pro-poor distribution. By contrast, substantial pro-rich inequity emerges in virtually every country with respect to the probability of contacting a medical specialist. Despite their lower needs for such care, wealthier and higher educated individuals appear to be much more likely to see a specialist than the less well-off. This phenomenon is universal in Europe, but stronger in countries where either private insurance cover or private practice options are offered to purchase quicker and/or preferential access. Pro-rich inequity in subsequent visits adds to this access inequity but appears more related to regional disparities in utilisation than other factors. Despite decades of universal and fairly comprehensive coverage in European countries, utilisation patterns suggest that rich and poor are not treated equally.

6.1 INTRODUCTION

It is well known that, despite many years of near universal coverage for physician services, income-related inequalities in the use of such services continue to persist in many European countries. There is abundant evidence that in many countries - European and non-European alike - both the probability of seeing a doctor and the number of contacts, given at least one contact, are not identically distributed across income groups after correcting for differences in the need for such care at different income levels. But there are also important differences between countries in the degree to which this occurs. Previous cross-country comparative work has concentrated on the measurement and testing of horizontal inequity in the use of physician services by assessing to what extent any observed differentials in use across income groups cannot be accounted for by need differences (Van Doorslaer et al., 1993). The premise of this research was that those in equal need ought to be treated equally, irrespective of income position and that violations of this principle constitute empirical evidence of horizontal inequity (Van Doorslaer et al., 1992; Wagstaff and Van doorslaer, 2000; Van Doorslaer et al, 2000).

More recently, attention has shifted from the measurement to the explanation of the differences in the degree of horizontal inequity observed in different countries. Using ECHP data, Van Doorslaer *et al*. (2002) not only generated comparable estimates of horizontal inequity, they also explored the role of differences in private health insurance status and region of residence in the generation of these findings. As in earlier work (Van Doorslaer et al., 2000), they found relatively little evidence of income-related inequity in the GP visits but substantial evidence of inequity favouring the rich in visits to a medical specialist: after controlling for need differences, higher income individuals report significantly more specialist visits than lower income individuals. Moreover, they found that – while insurance and location of residence do contribute to these findings – these two determinants do not "explain away" the inequity results.

This paper goes beyond the earlier work. First, it explicitly incorporates the two-stage decision process in physician utilisation. It examines inequity in the probability of a visit and the conditional (positive) number of visits separately by adopting two-part models and comparing these to a one-part model. This allows for an analysis of total inequity, as well as first and second part inequity. Secondly, it adopts a new (indirect) need standardisation approach by using the *partial* contributions of the need indicators as estimated in the decomposition procedure. Third, it allows for a decomposition 'by factors' of inequality (Wagstaff et al., 2003). The paper starts with an outline of the measurement and decomposition methodology in section 6.2. Section 6.3 provides a description of the data and estimation methods and section 6.4 presents the main results. We conclude with a discussion of the implications of the findings in section 6.5.

6.2 EXPLAINING INEQUITY IN HEALTH CARE UTILISATION

6.2.1 Measuring and decomposing inequality in use

The method we use in this paper to explain inequality in health care utilisation is conceptually identical to the method used in Van Doorslaer and Koolman (2002) to explain health inequality. We use a concentration index (C_M) as our measure of relative income-related inequality in use of health care. For weighted data, C_M can be computed conveniently using the (weighted) covariance between y_i and the fractional rank (based on weights) as (Lerman and. Yitzhaki, 1989):

$$C_M = \frac{2}{N\bar{y}} \sum_{i=1}^{N} w_i \left(y_i - \bar{y} \right) \left(R_i - \tfrac{1}{2} \right) = \frac{2}{\bar{y}} \operatorname{cov}_w (y_i, R_i) \qquad (6.1)$$

where y is our measure of medical care, cov_w denotes the weighted covariance. N is the sample size, $\bar{y}$ is the (weighted) mean health care use, w_i is the sampling weight of each individual i (with the sum of w_i equal to N). R_i is the relative fractional rank (based on weights) of the ith individual which indicates the weighted cumulative proportion of the population up to the midpoint of each individual weight.

A straightforward way of decomposing the measured degree of inequality into the contributions of explanatory factors was proposed by Wagstaff et al. (2003) in the context of a linear additive explanatory model such as:

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i \qquad (6.2)$$

where the x variables are a set of regressors associated with health care demand and ε is a disturbance term. One could think of this equation as a reduced form of a demand for health care equation where all the x variables are exogenous determinants. Given the relationship between y_i and x_{ki} in equation (6.2), the concentration index can be written as (Wagstaff et al., 2003):

$$C_M = \sum_k (\beta_k \bar{x}_k / \bar{y}) C_k + GC_\varepsilon / \bar{y} \qquad (6.3)$$

where $\bar{y}$ is the mean of y, $\bar{x}_k$ is the mean of x_k, C_k is the concentration index for x_k (defined analogously to C_M) and GC_ε is the generalized concentration index for ε_i. Equation (6.3) shows that C_M can be thought of as being made up of two components. The first is the deterministic component, equal to a weighted sum of the concentration indices of the k regressors, where the weight or "share" for, say, x_k, is simply the elasticity of y with respect to x_k. The second is a residual component, captured by the last term. This reflects the inequality in health care that cannot be explained by systematic variation across income groups in the x_k. Thus equation (6.3) shows that, by coupling regression

analysis with distributional data, we can partition total inequality into inequalities associated with each of the regressors, x_k.

The decomposition also makes clear how each regressor's separate contribution to total income-related inequality in health care demand can be decomposed into two meaningful parts: (i) its impact on demand, as measured by the demand elasticity, and (ii) its degree of unequal distribution across income, as measured by the (income) concentration index (C_k). This decomposition method therefore not only allows us to separate the contributions of the various regressors, but also to identify the importance of each of these two components within each factor's total contribution. This property makes it a powerful tool for unpacking the mechanisms contributing to a country's degree of inequality in use of health care.

One problem in this context is that demand for health care may not be very well modelled using linear estimation techniques such as OLS. Typically, models are intrinsically non-linear, either because of the probability or count data nature of the utilisation variables or because of the two-part structure of the demand decision process (Jones, 2000). In section 6.2.3 below we indicate how we have dealt with the non-linearity of the estimated models.

6.2.2 Measuring horizontal inequity in health care utilisation

Many OECD countries have explicitly included equity in the use of to health care as one of the main objectives in their health policy documents (Van Doorslaer et al., 1993; Hurst 2002). In most European countries, an egalitarian viewpoint of social justice seems to have been an important source of inspiration for these positions with respect to health care access. Usually, the horizontal version of the egalitarian principle is interpreted to require that people in equal need of care are treated equally, irrespective of characteristics such as income, place of residence, race, etc. In line with most of the previous work in this area (cf Wagstaff and Van Doorslaer (2000) for a review), the present study uses this principle of *horizontal inequity* (*HI*) as the yardstick for the international comparisons. While the concentration index of medical care use (C_M) measures the degree of inequality in the use of medical care by income, it does not yet measure the degree of inequity. For any inequality to be interpretable as inequity, legitimate or need-determined inequality has to be taken into account.

There are two broad ways of standardising distributions for need differences: the direct and the indirect method. The direct method proceeds by computing a concentration index for the medical care use that would emerge if each individual (or income group) had the same need characteristics as the population as a whole. Wagstaff *et al.* (1991) have used this procedure to compute what they call HI_{WVP} indices, which are essentially directly standardised concentration indices. Wagstaff and Van Doorslaer (2000) have

advocated the technique of indirect standardisation for the measurement of so-called HI_{WV} indices on the grounds that it is computationally easier and does not rely on grouped data. A measure of the need for medical care is obtained for each individual as the predicted use of a regression on need indicators. This means that in order to statistically equalize needs for the groups or individuals to be compared, one is effectively using the average relationship between need and treatment for the population as a whole as the vertical equity norm and horizontal inequity is measured by systematic deviations from this norm by income level.

Wagstaff and van Doorslaer (2000) proposed to measure *HI* by the difference between the inequality in actual and needed use of medical care:

$$HI_{WV} = C_M - C_N \qquad (6.4)$$

where C_M and C_N denote the concentration index corresponding to actual and needed use of medical care, respectively. C_N is computed using predicted values $\hat{y}_i$, which can be estimated for each individual *i* as the expected amount of medical care he or she would have received if he or she had been treated as others with the same need characteristics were, on average, treated by the system. Typically, these are obtained from regressing actual y_i on a set of need indicators like health status and morbidity measures and demographics. The average relationship between need indicators and utilization, as embodied in the regression coefficients, is the implied norm for assessing equity in this health care system. A positive (negative) value of *HIWV* indicates horizontal inequity favoring the better-off (worse-off). A zero index value indicates no horizontal inequity, i.e. that medical care and need are proportionally distributed across the income distribution. It is worth emphasizing that coinciding concentration curves for need and actual use provide a sufficient but not a necessary condition for no inequity. These indices were used to measure, test and compare horizontal inequity across countries in van Doorslaer *et al.* (2000).

One further step in the direction of explaining horizontal inequity was made in Van Doorslaer et al. (2002) by including other, non-need determinants in the (indirect) need standardisation process. In their search for an explanation of cross-country differences in the *HIWV* indices, they found, for instance, that inclusion of factors like health insurance and regional fixed effects in the standardisation did reduce the degree of pro-rich inequity in specialist use, but seldom to an extent that made it insignificant. They interpreted this as evidence that health insurance and regional variation do play a role in explaining the occurrence and degree of horizontal inequity.

The issue of the role of explanatory models in the measurement of inequity deserves some further attention. Some authors have drawn attention to the potential biases involved in these standardisation procedures. First, the problem of determining which

systematic variations in medical care use by income are "needed" and therefore, in a sense, justifiable, and which are not, bears some resemblance to the problem of determining legitimate compensation in the risk adjustment literature. Schokkaert and Van de Voorde (2000) have argued that while there is a difference between the positive exercise of *explaining* medical care expenditure (or use) and the normative issue of justifying medical expenditure (or use) differences, the results of the former exercise have relevance for the second. Drawing on the theory of fair compensation, they show that failure to include 'responsibility variables' (which *do not* need to be compensated for in the capitation formula) in the equation used for estimating the effect of 'compensation variables' (which *do* need to be compensated for) may give rise to omitted variable bias in the determination of the 'appropriate' capitations (or fair compensations). Their proposed remedy to this problem is to include the 'omitted variables' in the estimation equation but to 'neutralize' their impact by setting these variables equal to their means in the need-prediction equation. They claim that the argument that even this more fully specified model may suffer from omitted variable bias due to the unavailability of certain variables cannot be used as an excuse for not including what is available. Schokkaert and Van de Voorde point out that the procedure to neutralize the responsibility variables does not hold if the model is not linearly additive.

A similar argument to Schokkaert and Van de Voorde (2000) was made and taken further by Gravelle (2001) in the context of the measurement of income-related inequality of health or health care. He uses an 'augmented partial concentration index' which is defined as the (directly) standardised concentration index, but controlling for income and other non-standardising variables in the process. In effect, he distinguishes between three types of x_k variables in equation (6.2): income itself (x^r), need standardising variables (a vector x^n) and other, possibly policy-relevant variables (a vector x^p):

$$y_i = \alpha + \beta_r x_i^r + \sum\nolimits_n \beta_n x_i^n + \sum\nolimits_p \beta_p x_i^p + \varepsilon_i \qquad (6.5)$$

The equivalent of equation (6.3) for this specification then becomes:

$$\hat{C} = (\beta_r \bar{x}_r / \bar{y})\hat{C}_r + \sum\nolimits_n (\beta_n \bar{x}_n / \bar{y})\hat{C}_n + \sum\nolimits_p (\beta_p \bar{x}_p / \bar{y})\hat{C}_p + GC_\varepsilon / \bar{y} \qquad (6.6)$$

where the first term denotes the (partial) contribution of income inequality ($\hat{C}_r$ equals the Gini coefficient of income inequality if income is entered linearly), the second the contribution of need variables, the third the contribution of other, potentially policy-relevant variables and the last term is, as before, the generalised concentration index of å. Gravelle (2001) labels the first term the partial concentration index and the sum of the first and third term the 'augmented partial concentration index'.

In the context of a linear model, equation (6.6) therefore provides a neat way to decompose the total measured inequality in medical care use into four sources: (a) the

contribution of income, defined as the product of the income elasticity of medical care use and the concentration index of income; (b) the contribution of the need variables, (c) the contribution of other variables, potentially amenable to policy intervention, and (d) a residual term which basically captures the degree to which the residual is correlated with income rank. Assuming that equation (6.5) leads to a better estimate of the (partial) need contribution, then a model without the x^r and x^p variables, equation (6.6) provides an alternative estimate of horizontal inequity as the C_M minus the second term, or equivalently as the sum of (a), (c) and (d).

As long as the model for y is linear, as in equation (6.5), then the Schokkaert and Van de Voorde (2000) approach of estimating the full regression first and then neutralizing the non-need variables by setting them equal to their mean (or, in fact, any constant value) and the decomposition approach lead to the same measure of horizontal inequity (see Appendix). The decomposition approach has the additional advantage of greater transparency in the presentation of results. Estimating a regression model for use of health care does not require *a priori* agreement on what constitute 'justifiable' and 'unjustifiable' causes of inequality in health care use by income. Some may, for instance, prefer to exclude variables like gender or age from the x^n vector and to include them in the x^p vector, on the grounds that, after having controlled for other health differences, age and gender in and of themselves do not constitute legitimate reasons for differential medical care consumption. Similarly, the question arises whether the residual contribution - term (d) in equation (6.8) - needs to be attributed to justifiable or unjustifiable sources of inequality. In our approach, we have decided to classify all of it as unjustifiable variation. At the other extreme, it could be argued that the residuals mainly capture unmeasured need and hence that the residual contribution should be subtracted from *HI*. The decomposition method and, in particular the graphical analysis, make the implications of these different assumptions transparent.

6.2.3 Nonlinear regression models

One important problem with measuring horizontal inequity and applying the decomposition analysis in the present context is that they will not be linear because the dependent variable in health care demand models is modelled as a nonlinear function of the *x* variables. Our empirical models of health care use are based on logistic and truncated and generalized negative binomial regression models, which are intrinsically nonlinear. The general functional form *G* of such a nonlinear model can be written as:

$$E(y_i|x_i) = G(\sum\nolimits_k \beta_k x_i^k) \qquad (6.7)$$

To compute horizontal inequity in the context of a nonlinear model, again we have used a two-step approach. In the first step we predict need-expected utilisation based on the

actual values of the x_n variables, but these predictions are contingent on the level of the non-need variables (x^r and x^p) that is selected. By analogy with the linear case, we have chosen to set the non-need variables equal to their sample means. So:

$$\hat{y}_i = E(y \mid x_i^n, \bar{x}^r, \bar{x}^p) = G(\sum\nolimits_n \hat{\beta}_n x_i^n + \sum\nolimits_r \hat{\beta}_r \bar{x}_i^r + \sum\nolimits_p \hat{\beta}_p \bar{x}_i^p) \quad (6.8)$$

As before, in the second step the HI index is then obtained by subtracting the concentration index of $\hat{y}$ from the concentration index of y. A complication compared to the linear case is that the HI index for the nonlinear model is contingent on the values used for the non-need variables and therefore their effect is not completely neutralised.

It was noted above that, in the context of linear models, the two-step approach to neutralizing the non-need variables and the decomposition approach give the same measure of horizontal inequity. This does not hold for a nonlinear model, as the linear decomposition does not directly apply to equation (6.7). However it is possible to approximate the decomposition analysis. To do this, we have opted to use the 'partial effects' representation for the decomposition. This has the advantage of being a linear additive model of actual utilisation, but it is only an approximation. A linear approximation of this function is given by:

$$y_i = \sum\nolimits_k \beta_k^m x_i^k + u_i \quad (6.9)$$

where the β_k^m are the partial effects of each x and u_i is the implied error term which includes approximation errors. For the dummy variables local average partial effects are computed using a procedure equal to computing treatment effects evaluated for the treated (Wooldridge, 2002). This means that β_k^m is measured by computing the average effect for each individual with characteristic k and then taking the sample mean over this sub-set of individuals. So, for instance, the average effect of unemployment is calculated as the mean of β_k^m for those who are unemployed. This captures the fact that the unemployed differ from the population as a whole in terms of other characteristics such as age, education, etc. The partial effect for the continuous variable (income) was computed by evaluating the change in $\hat{y}_i$ for each individual around the observed income and other characteristics. To obtain our estimate we added and subtracted a small number to the observed value, we computed the difference and divided it by twice the fraction.

While equation (6.9) is an approximation of the non-linear relationship estimated by the logit or the truncated or generalised Negbin models, it does allow us to restore the mechanics of the decomposition framework by replacing the β_k parameters in equation by the β_k^m parameters. This modified version of equation (6.3) provides the basis for our decompositions of the first and the second part of two-part models presented in

section 6.5. They provide an estimate of the contributions of individual factors to the overall level of inequality in use of health care, but because of the linear approximation error, the HI estimate obtained through the decomposition equation (6.10) will, in general, not be identical to the HI estimate obtained through the two-step approach using equation (6.6).

One way of checking the reliability of the linearization method is to apply the two-step standardization method described in equation (6.8) to each of the groups of variables (need, non-need, income) to obtain estimates of their contributions to inequality. Like the linearization method, this still induces an approximation error, but it tends to be smaller since it still captures the curvature of the nonlinear function, G(.), and it does not suppress multiplicative interactions within the groups of variables, only between the variables of interest and the 'standardising' variables. As such, it represents a somewhat closer approximation of the partial contributions. However, unlike the linear approximation, it does not restore the full mechanics of the decomposition into the contributions of elasticities and of inequalities of the regressors.

6.2.4 Statistical inference

In addition to measuring inequality and inequity, we aim to test for cross-country differences. Standard errors for the *C* and *HI* indices were computed using the convenient regression procedure for weighted data, and corrected for heteroscedasticity and clustering on a household level. For the contribution terms we have opted to use a "bootstrap" method (Efron and Tibshirani, 1993; Deaton, 1997) to assess sampling variability and to obtain standard errors for the estimates of the contributions in the decomposition analysis. A standard bootstrap procedure hinges on the assumption that the observed distribution is a random sample of the underlying population distribution, and that individuals within the sample are independent. This assumption does not hold for the complex multi-stage sampling designs used to gather the ECHP data. Therefore we have implemented the bootstrap using the following procedure. First, for the countries for which data were sampled in two stages (i.e. BE, UK, IE, IT, GR, ES, PT), we have drawn a random sub-sample (with replacement) of the primary sampling units (PSU) of a size equal to the original sample size. This step was not used for Germany, the Netherlands and Austria, where PSU information was not made available, or for Denmark and Luxembourg, where PSUs were not used. Second, we have drawn a random sub-sample (with replacement) of households within each of the sampled PSUs, and included all members of these households. Third, for each draw, we have normalised the sampling weights to a mean of one, and have run the entire (weighted) procedure to obtain the factor contributions, including the regressions, marginal effects, fractional rank construction and covariance computations. Fourth, repeating this whole

process, we have generated 100 resample data sets each providing us with estimates of the contributions. Sixth, using these datasets we have computed the standard deviations as an estimate of the standard error of each factor's contribution and for the *HI* index.

6.3 DATA AND ESTIMATION METHODS

6.3.1 ECHP Data

The data are taken from the third wave (held in 1996) of the *European Community Household Panel* (ECHP) conducted by Eurostat, the European Statistical Office. The ECHP is a survey based on a standardised questionnaire that involves annual interviewing of a representative panel of households and individuals of 16 years and older in each EU member state (Eurostat, 1999). It covers a wide range of topics including demographics, income, social transfers, health, housing, education, employment, etc. We use data for the following twelve member states of the EU: Austria, Belgium, Denmark, Germany, Greece, Ireland, Italy, Luxemburg, Netherlands, Portugal, Spain and the United Kingdom. The three missing member states are France (missing utilisation questions), Finland (missing income data) and Sweden (not taking part in ECHP).

The ECHP income measure (our ranking variable) is disposable (i.e. after-tax) household income per equivalent adult, using the modified OECD equivalence scale (giving a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 4 in the household). Total household income includes all net monetary income received by the household members during the reference year (which is 1995 for the 1996 wave). It includes income from work (employment and self-employment), private non-labour income (from investments and property and private transfers to the household), pensions and other direct social transfers received. No account has been taken of indirect social transfers (e.g. reimbursement of medical expenses), receipts in kind and imputed rent from owner-occupied accommodation.

Measurement of utilisation of general practitioner (GP) and medical specialist services in the ECHP is based on the question "During the past 12 months, about how many times have you consulted a GP/medical specialist?" We use one-year lagged health measures from wave 2 (1995) based on two questions: (a) responses to a question on self-assessed general health status as either very good, good, fair, bad or very bad; and (b) responses to "Do you have any chronic physical or mental health problem, illness or disability? (yes/no)" and if so "Are you hampered in your daily activities by this physical or mental health problem, illness or disability? (no; yes, to some extent; yes, severely)". We use two dummies to indicate either some limitation or severe limitation.

We also included a number of other variables which have been shown in the literature to affect health care utilisation of health care but are not direct indicators of health status or need for care. Education and marital status are assumed to affect then efficiency of health production and the propensity to seek care, while both activity status and region of residence are more likely to affect the time price of health care use. We included the following variables: (i) the highest level of general or higher education completed, i.e. third level education (ISCED 5-7), second stage of secondary level education (ISCED 3) or less than second stage of secondary education (ISCED 0-2)); (ii) Marital status, distinguishing between married/living in consensual union, separated/ divorced, widowed and unmarried; (iii) Activity status includes employed, self-employed, student, unemployed, retired, doing housework and 'other economically inactive'. Region of residence uses the EU's NUTS 1 level (Nomenclature of Statistical Territorial Units) except for countries where such information was withheld for privacy reasons (The Netherlands, Germany) or because the country is too small (Denmark, Luxembourg). Regional identifiers are presented in the companion paper (Van Doorslaer and Koolman, 2002). Although most country's sample sizes are between 7000 and 11,000 adults, some are larger (Spain, Italy) and some are smaller (Denmark and Luxembourg). Cross-sectional sample weights at the individual level were applied in all analyses.

6.3.2 Cross-country differences

All 12 countries included in this analysis had, by 1996, achieved close to universal coverage of their population for the majority of physician services, but some important between-country differences remain with respect to potentially equity-relevant features of their financing and delivery systems. Van Doorslaer *et al.* (2002) have summarized some of the salient system characteristics which may have an impact on any differential utilisation of the general practitioners or medical specialists by income level. In some countries, there are different groups of insured with varying degrees of coverage or rules of reimbursement at different levels of income. This is the case for rather small numbers of high-income earners with private coverage in Denmark and Germany, but it concerns sizeable portions of the population in Ireland and the Netherlands. Some countries' public insurance have substantial co-payments for GP and specialist consultations (e.g. up to 30-40% of fees in Belgium and Luxembourg; €20 for higher income patients in Ireland), some charge small co-payments (e.g. Portugal) while in many other countries (like Denmark, Germany, Greece, Spain and the UK) visits to public sector doctors are free at the point of delivery [cf tables in Mossialos (2002) and Robinson (2002) for details]. In some countries, notably Denmark, Ireland, Italy, The Netherlands, Portugal, Spain and the UK, the primary care physician acts as a "gatekeeper" referring to secondary care provided by medical specialists, whereas in other countries, there is

direct access to all physicians. Some countries pay their general practitioners mainly by capitation (Denmark, Italy, Netherlands, UK) or salary (Greece, Portugal, Spain) whereas others rely mainly on fee-for-service payment. We will have to keep these system differences in mind when interpreting the cross-country comparative results.

6.3.3 Estimation methods

Health care utilisation data like physician visits are known to have a very skewed distribution with the majority of survey respondents reporting zero or few visits and only a very small proportion reporting frequent use. In such cases, integer count data regression is appropriate and a variety of models have been proposed and used (Cameron and Trivedi, 1998). The negative binomial model, which allows for overdispersion, has often been shown to be an adequate choice in studies of health care utilisation. Many studies have also emphasised the principal-agent relationship between doctor and patient and stressed the distinction between patient initiated decisions, such as the first contact with a GP, and decisions that are influenced by the doctor, such as repeat visits, prescriptions, and referrals (Pohlmeier and Ulrich, 1995). The consequence, in statistical terms, is a hurdle or two-part model (TPM) which allows the participation decision and the positive count, to be generated by separate probability processes. The two-part model has often been estimated using a logit for the first stage and a truncated negbin model for the second stage (Grootendorst, 1995; Gerdtham, 1997; Hakkinen et al., 1996; Jones and O'Donnell, 2002).

We have chosen to adopt a TPM model combining a logit and a truncated negbin for both the GP and specialist equations because we are interested to see whether the effect of income differs between first and subsequent contacts. In addition, we have estimated equations for the total number of visits using the generalised negbin model (cf Jones, 2000). The generalisation consists of modelling the excess zeros as unobservable heterogeneity; allowing the heterogeneity parameter (alpha) to be a function of the *x*'s rather than being constant. Like Jimenez *et al.* (2002), we have exploited the availability of previous waves of the ECHP to use lagged values of the health variables in order to reduce the risk of endogeneity in the health status variables. Because of their rejection of cross-country homogeneity, we have chosen not to pool the data across countries. For all countries and surveys, cross-sectional sample weights were used in all computations in order to make the results more representative of the countries' populations. Robust standard errors were obtained by applying White-Huber-sandwich estimator that corrects for heteroskedasticity of unknown form. This estimator was adjusted to also correct cluster sampling. We clustered on households, as this is the smallest cluster unit and information on the primary sampling units was not available for all countries.

6.4 RESULTS

While the focus of this paper is on the differences in relative inequality in utilisation by income level *within* European countries, it is clear that there is tremendous variation also in the average levels of physician utilisation *across* these countries. In Van Doorslaer et al. (2002), it is shown that the mean annual number of visits to a GP varies from a low of 2.19 in Greece to a high of 5.39 in Austria, and mean visits to a specialist from a low of 0.62 in Ireland to a high of 3.29 in Germany. Some countries, notably Germany and Austria, have above-European average rates of utilization for both GP and specialist visits. Countries with below-average utilisation rates for both types of visits include Ireland, Netherlands, Denmark, UK, Portugal, Spain and Greece. Belgium and Italy have above-average GP visit rates only and Luxemburg is the only country with above-average specialist visit rates only. These inter-country differences in mean utilisation levels are probably closely related to GP and specialist availability and remuneration across countries. They have to be interpreted carefully given the cross-country differences in the definition of GP visits (office, home, health centre) and specialist visits (private practice, in hospital, etc).

6.4.1 Indices for total inequality and horizontal inequity

GP care utilisation

The results summarized in Table 6.1 can be compared with those obtained in Van Doorslaer et al. (2002) for the total number of visits, but here the findings are decomposed *by parts* of the decision process. Statistically significant contributions are indicated in bold. Virtually all concentration indices for the probability of a visit, the conditional and the total number of visits are negative. This means that generally lower income groups are *both* more likely to seek care from a GP than higher income groups, *and* they do so more frequently. But the unequal distribution of GP care to a large extent appears to be in line with the similarly unequal distribution of the need for such care. After controlling for the unequal need distributions, the resulting horizontal inequity indices are insignificant in seven countries, pro-poor in Spain and Denmark, and significantly pro-rich in Belgium, UK and the Netherlands. However, in all countries, the horizontal inequity (*HI*) index for the visit probability is fairly small, i.e. within the range [-0.01; 0.02]. There appears to be only a very small degree of income-related horizontal inequity in the GP contact probability.

The picture is somewhat different for the *conditional (positive) number* of visits. Table 6.1 shows that income-related inequity is more negative (i.e. favouring the lower income groups) and significant in eight of the twelve countries in the second part of the demand model. Both the C_M and the *HI* indices here are generally more negative than for the

Table 6.1 Inequality and inequity in GP visits, ECHP, 1996

GP visits	Probability of a visit		Conditional # of visits		Total # of visits	
	Inequality (C_M)	Inequity (*HI*)	Inequality (C_M)	Inequity (*HI*)	Inequality (C_M)	Inequity (*HI*)
Ireland	**-0.0187**	0.0035	**-0.1136**	**-0.0657**	**-0.1323**	**-0.0696**
Belgium	0.0037	**0.0121**	**-0.1183**	**-0.0564**	**-0.1145**	**-0.0508**
Spain	**-0.0294**	**-0.0167**	**-0.0612**	**-0.0371**	**-0.0906**	**-0.0492**
Luxembourg	-0.0076	0.0002	**-0.0841**	**-0.0428**	**-0.0918**	**-0.0406**
Italy	-0.0055	-0.0002	**-0.0594**	**-0.0322**	**-0.0649**	**-0.0349**
Greece	**-0.0413**	-0.0041	**-0.0845**	**-0.0212**	**-0.1258**	**-0.0308**
Germany	**-0.0124**	**-0.0082**	**-0.0513**	-0.0173	**-0.0636**	**-0.0268**
UK	-0.0076	**0.0109**	**-0.0930**	**-0.0301**	**-0.1006**	**-0.0240**
Netherlands	-0.0019	**0.0103**	**-0.0517**	**-0.0201**	**-0.0535**	-0.0113
Denmark	**-0.0200**	0.0061	**-0.0631**	-0.0085	**-0.0831**	-0.0008
Portugal	**-0.0143**	0.0099	**-0.0549**	-0.0038	**-0.0692**	0.0051
Austria	**-0.0082**	-0.0018	**-0.0417**	0.0114	**-0.0499**	0.0146

Notes: Countries ranked by inequity index for total visits (last column). Inequity indices computed using a logit model for the probability, a truncated negbin model for the conditional number and a generalised negbin for the total number of visits. Statistically significant *HI* indices in bold (P<0.05).

Table 6.2 Inequality and inequity in specialist visits, ECHP 1996

	Probability		Cond Number		Total	
	Inequality (C_M)	Inequity (*HI*)	Inequality (C_M)	Inequity (*HI*)	Inequality (C_M)	Inequity (*HI*)
Luxembourg	0.0195	**0.0346**	**-0.0899**	**-0.0594**	**-0.0704**	-0.0282
Belgium	0.0125	**0.0344**	-0.0394	-0.0008	-0.0269	0.0255
Netherlands	-0.0041	**0.0307**	-0.0137	0.0197	-0.0178	**0.0413**
Greece	-0.0175	**0.0355**	-0.0242	0.0216	**-0.0418**	**0.0492**
Germany	**0.0130**	**0.0243**	0.0029	**0.0269**	0.0158	**0.0517**
UK	0.0163	**0.0723**	**-0.0397**	-0.0062	-0.0234	**0.0524**
Italy	**0.0416**	**0.0617**	-0.0237	-0.0035	0.0179	**0.0537**
Spain	**0.0439**	**0.0658**	-0.0171	0.0121	0.0267	**0.0714**
Austria	0.0108	**0.0214**	0.0237	**0.0554**	0.0345	**0.0740**
Denmark	-0.0074	0.0223	0.0297	**0.0581**	0.0223	**0.0844**
Ireland	**0.0621**	**0.1168**	0.0149	0.0299	**0.0770**	**0.1388**
Portugal	**0.0774**	**0.1103**	0.0197	**0.0549**	**0.0971**	**0.1604**

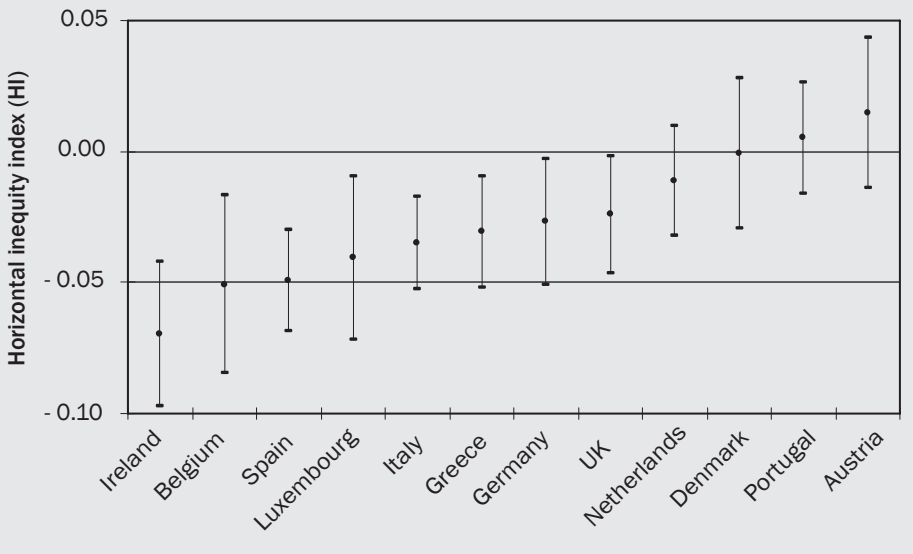
Note: Countries ranked by total inequity index (last column). Contributions computed using a logit model for the probability, a truncated negbin model for the conditional number and a generalised negbin for the total number of visits. Statistically significant *HI* indices in bold (P<0.05).

probability of a visit. As a result, we also find substantial inequity in *total* GP visits, which are concentrated among the poorer segments in most countries, with significant inequity favouring the lower income groups in eight of the countries. The index values with the corresponding 95% confidence intervals for total number of GP visits are shown graphically in Figure 6.1. All the significant values are in the range [-0.02; -0.05] except for Ireland, for which it equals -0.07. In only four countries — Austria, Denmark, Portugal and Netherlands — the hypothesis of no inequity cannot be rejected. There does not appear to be one system characteristic that explains this finding since these four countries all have very diverse characteristics. There appears to be a general tendency — irrespective of the system characteristics — for lower income groups to have more frequent GP visits in European countries. In the two countries with most negative index values, this tendency is exacerbated by pro-poor discrimination. In Ireland, only the 30% on the lowest incomes are medical cardholders and entitled to free GP services, while others must pay out-of-pocket. In Belgium the elderly and chronically ill on low incomes pay much reduced co-payments.

Specialist care utilisation

The distribution of specialist care utilisation by income, summarized in Table 6.2, looks dramatically different from the use of GP services. In all but three countries (the Netherlands, Denmark and Greece), higher income groups are *more* likely to report at least one visit to a specialist, while the need for such care is invariably higher among the

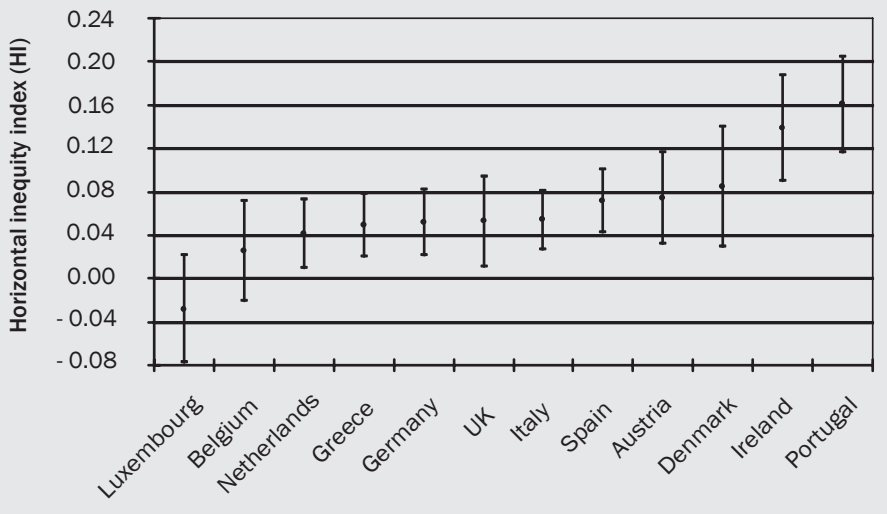
Figure 6.1 Inequity indices for number of GP visits (with 95% confidence intervals)



lower income groups. It is therefore not a surprise that, after controlling for these need differences, we find substantial degrees of horizontal inequity favouring the rich for this probability, which are statistically significant in all countries except Denmark. In seven of the countries, this is compounded by similar pro-rich inequity in the conditional number of specialist visits. Overall, we see a high and significant degree of pro-rich inequity in total specialist visits in all countries except Luxembourg and Belgium. Luxembourg is a somewhat special case because of its small size (and sample), the lack of academic hospitals, the high degree of cross-border care delivery and the unclear distinction between a specialist and a general practitioner. Belgium's more equal distribution may be due to its positive discrimination in favour of certain lower income groups through lower rates of co-payment.

Figure 6.2 illustrates the between-country differences and confidence intervals. Most countries show inequity indices between 0.04 and 0.08 which are not significantly different from one another. Portugal and Ireland, in particular, show significantly horizontal inequity index values than many of the other countries, and only the indices for Luxembourg and Belgium are not significantly different from zero. The reason for the strong pro-rich pattern in Ireland seems fairly obviously related to the dual insurance system by income level in this country, where low income groups with a medical card (30%) are entitled to free GP services but higher income groups have to pay out-of-pocket and increasingly buy private insurance to cover outpatient (specialist service) charges. The situation is different in Portugal where much of the pro-rich distribution

Figure 6.2 Inequity indices for total number of specialist visits (with 95% confidence intervals)



appears related to the high share paid for out-of-pocket (or through private insurance) for private consultations and the low access to specialist services in poorly endowed regions.

6.4.2 Decomposition analysis

In the companion paper on health inequalities (Van Doorslaer and Koolman, 2002), we have demonstrated in more detail how the estimated elasticities and concentration indices of the regressors translate into contributions to inequality. In the extended version of this paper, we have illustrated the 'mechanics' of the decomposition method for the example of specialist visit probability in Spain. The illustration shows how, for instance, positive use elasticities combined with negative concentration indices lead to negative inequality contributions for most of the need variables (e.g. those indicating older age groups and lower health states). As such, these variables contribute to pro-poor inequality. On the other hand, variables like income, education and living in one of the richer Spanish regions (Madrid, the East and the North-east) show both positive elasticities and positive concentration indices and therefore contribute to pro-rich inequality in the use distribution.

For reasons of space, we have included inequality contributions by type of utilisation and by country for just one type of utilization (the probability of a specialist visit) in Table 6.3. Here we concentrate on the broad picture by looking at the inequality decomposition into the contribution of four sources: (i) income itself, (ii) need variables, including health status at the beginning of the reference period and age and gender dummies, (iii) other demand determinants like education, labour force or marital status and region, and (iv) the residual term. As explained in section 6.2.1, each of these determinants will contribute to the total income-related inequality in use to the extent that (a) it has a significant demand elasticity, and (b) it is unequally distributed by income.

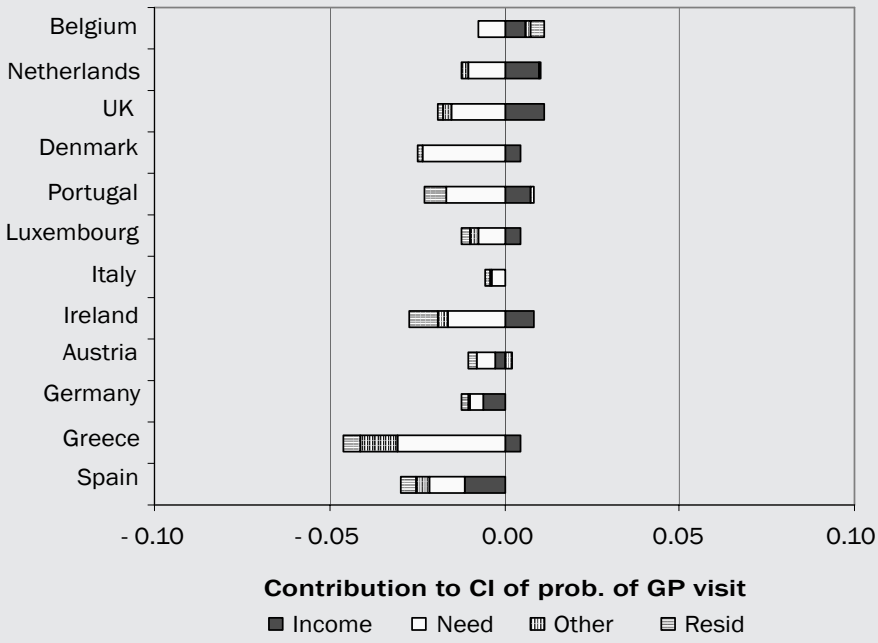
As explained in section 6.2.3, in the nonlinear model setting, the degree of horizontal inequity can only be decomposed approximately using an equation like equation (6.10). The generalized concentration index of the error term then includes both an estimation error and an approximation error. This is an inevitable price to be paid to restore linearity. As a result, the HI indices estimated using the two-step approach explained in equation (6.6) with the need predictions generated by equation (6.8) will not, in general, be identical to estimates generated with the linear approximation of the decomposition. The approximate contribution estimates nonetheless provide some useful insight into the direction and magnitude of the various source contributions. A comparison of the contributions estimated with the linearization and the standardisation methods (as described in section 6.2.3) shows that in the great majority of cases the results are very similar. Where they do differ, the size of the error term tends to be smaller in

the standardization method estimates and this is often accompanied by a smaller contribution of the need variables. In other words, if the approximation causes bias in the partial effect, it tends to somewhat underestimate the need contributions. The linear approximation clearly appears to perform much better for the probability of use than for the conditional positive use. And for conditional positive use, it performs better for specialist visits than for GP visits, probably because of its inability to properly capture the infrequent number of very high counts in GP visits.

GP care utilisation

Figure 6.3 presents the contributions of the four sources of inequality as distinguished in equation (6.6). Inequality in the *probability* of a GP visit in each of the 12 countries is (approximately) decomposed into the partial contributions of (a) (the log of) household income, (b) need indicators (self-reported morbidity, age and sex), (c) other non-need variables (education, marital and activity status and region) and (d) a residual term. As explained in section 6.2.3, the latter term includes both a prediction error and an error generated by the linear approximation used to obtain the marginal effects. It is to

Figure 6.3 Decomposition of inequality in GP visit probability



Note: Decomposition based on linear approximation using marginal effects from a logit regression. Countries ranked by degree of horizontal inequity

be noted that aggregating the contributions of several (dummy) variables means that positive and negative contributions may cancel out in the aggregate so that a small contribution may 'hide' the summation of larger positive and negative contributions. The income variable was transformed logarithmically in keeping with previous practice in the literature.

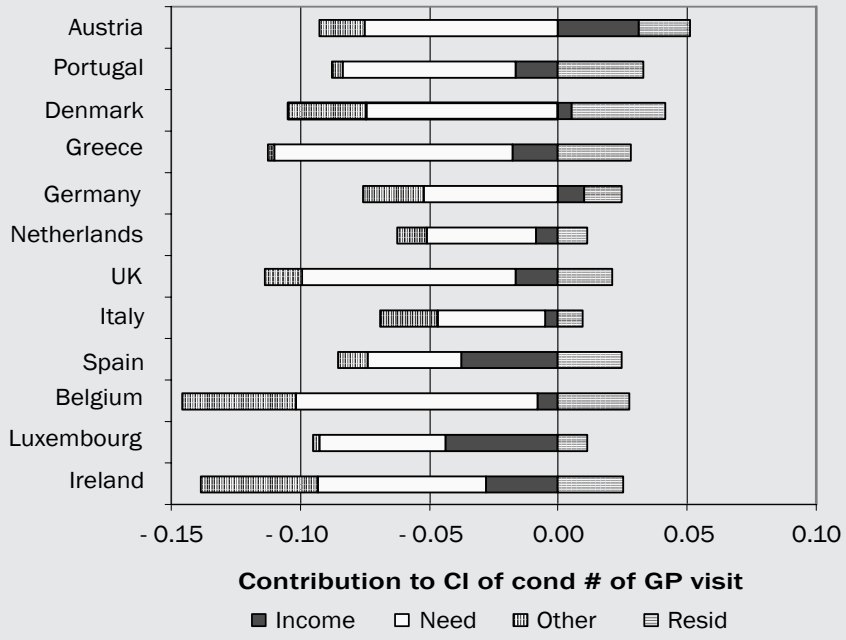
One way of reading the chart is as follows. In a country where the probability of a GP visit were equally distributed across income, the sum of the bars would be zero. In a country with a perfectly equitable distribution of GP visits across income, the sum of the bars would be equal to the need bar, which indicates the distribution of need by income. As soon as discrepancies emerge between the actual and the need-expected distribution, the other bars appear. They indicate what share of the discrepancy between need and use is due to either income itself, or to other variables included in the equation, or to variables not included.

We can see that inequality in the GP use probability is fairly small and pro-poor, and mainly accounted for by the contribution of need factors in all countries. This means that the distribution is pro-poor because the need distribution is pro-poor. The partial contribution of income is generally positive but rather modest. All other variables show negative but small contributions. Where inequality is substantial, as in Greece, it is mainly a consequence of the unequal distribution of *education* by income: the higher educated tend to be richer but, *ceteris paribus*, less likely to use GP services. The influence of education may capture differences in communication skills or simply taste differences.

Figure 6.4 shows greater inequality in the distribution of the conditional (positive) number of GP visits. It is more pro-poor, but again this is mainly due to the greater needs of the poor. The often negative partial contribution of income indicates pro-poor treatment patterns (except in Austria, Germany and Denmark). The full decomposition results show that important 'other' non-need variables contributing to the pro-poor distribution are education (in all countries except Denmark, Netherlands, Luxembourg and Austria), non-active status like retired, unemployed, housewives, inactive status in Belgium, Ireland, and Italy (in all countries except Netherlands, Luxembourg and Greece) and region (in the Mediterranean countries). To the extent that some of these categories may reflect a greater need for care (like e.g. inactive status may include recipients of disability pensions) and have a negative contribution, we may, in effect, be underestimating the need contribution and therefore overestimating the degree of pro-poor inequity.

The residual contribution in this chart is comparatively larger than in the other charts and always positive. We have decomposed the "linear" residual, that appears in the bar charts, into the genuine residual from the nonlinear model and the approximation error

Figure 6.4 Decomposition of inequality in conditional number of GP visits



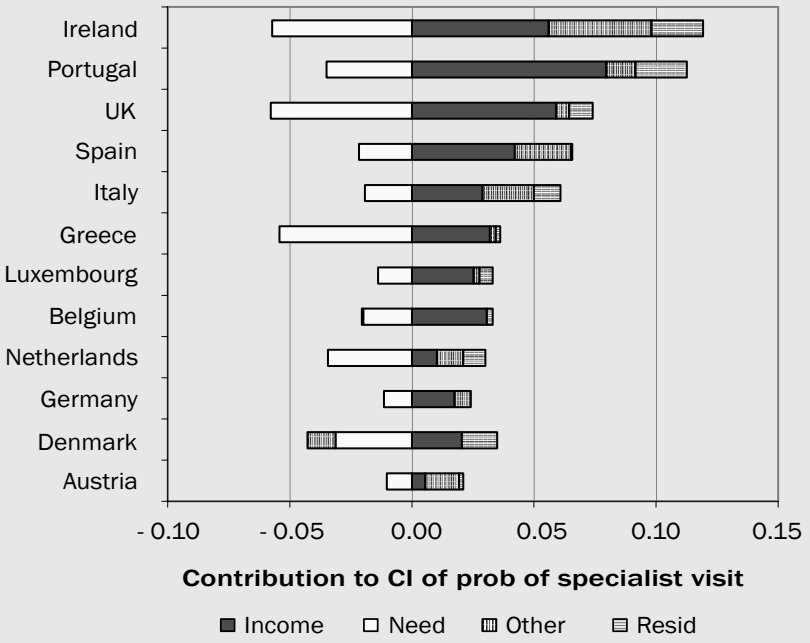
Note: Decomposition based on linear approximation using marginal effects from a truncated negbin regression. Countries ranked by degree of horizontal inequity

and found that the tendency to observe positive bars is predominantly attributable to the approximation error.

Specialist care utilisation

From the (aggregate) decomposition in figure 6.5 we can see that pro-rich inequity is mainly the result of a strong partial contribution of income in most countries (lowest in Austria and the Netherlands), which is exacerbated by the contribution of other variables in some countries, notably in Ireland, Spain, Italy and Portugal. Table 6.3 shows that the effect of these 'other variables' is primarily due to the very pro-rich contribution of higher education. While we did not include a variable indicating coverage by private health insurance in these reduced form equations, it is likely that such private cover will contribute significantly to the pro-rich distribution of specialist visit probabilities. It may not be a coincidence that the highest pro-rich inequity indices are found for precisely the five countries for which such 'duplicate private coverage' is most prevalent (i.e. Ireland, Portugal, UK, Spain and Italy). In other words, much of the income and 'other' variable

Figure 6.5 Decomposition of inequality in specialist visit probability



Note: Decomposition based on marginal effects from logit regression. Countries ranked by degree of horizontal inequity

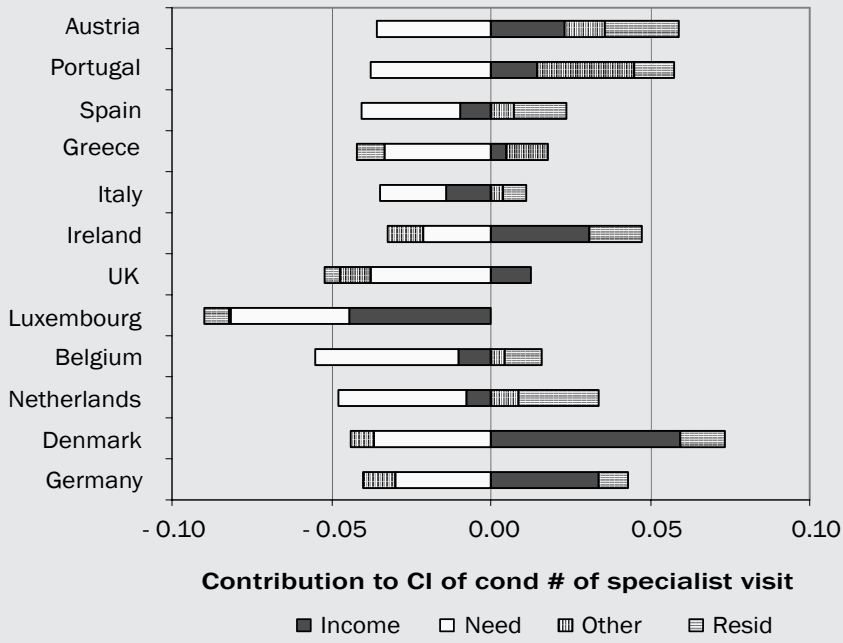
contributions may, in fact, reflect the role of the unequal distribution of private insurance coverage (cf. Van Doorslaer et al. (2002).

Figure 6.6, on the other hand, shows that the contribution of income is less important for inequity in subsequent specialist visits. It is only significantly positive in Denmark and Germany, and significantly negative in Luxembourg. In a few countries, e.g. Portugal, Spain and Greece, other variables contribute more to the pro-rich distribution of these visits. The full decomposition results reveal that in this case it is the regional disparities which play an important role. In the three southern countries, a sizeable share of the pro-rich inequity is due to the much higher use of specialist visits in the richer capital regions of Madrid, Lisbon and Athens. This finding highlights the usefulness of the decomposition approach to trace the sources of inequality patterns in medical care use.

6.5 CONCLUSION AND DISCUSSION

This paper provides new evidence on the sources of differences between European countries in the degree to which health care use is unequally distributed by income.

Figure 6.6 Decomposition of inequality in conditional number of specialist visits



Note: Decomposition based on linear approximation using marginal effects from a truncated negbin regression. Countries ranked by degree of horizontal inequity

While it builds on previous international comparative work, it also offers a number of advances, both in terms of new data analysed and in terms of new methods used. First, it exploits new and comparable data on the use of general practitioner and specialist services in 12 EU member states collected in the 1996 wave of the *European Community Household Panel*. Secondly, it employs new methods for decomposing the total observed inequality in utilisation by 'sources'. While such methods have been deployed previously and successfully for the decomposition of inequalities in health, they have hitherto not been used to examine the sources of inequality in utilisation. The main reason for this is that the decomposition method was developed for linear models, while it is well known that medical care use is typically and most appropriately modelled using inherently non-linear models. We show that a linear approximation of these models using a 'marginal effects' representation of the decomposition is one way of dealing with this non-linearity problem. As a result, we can decompose (an approximation of) the inequality in *actual* use, not in the latent index representing the propensity to use medical care. Thirdly, we also perform a decomposition 'by parts' of the decision process by doing this separately for the probability of a visit and for the

Table 6.3 Contributions to inequality in the specialist visit probability

	Germany	Denmark	Netherl	Belgium	Luxemb	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
C (actual)	0.0130	-0.0074	-0.0041	0.0125	0.0195	0.0163	0.0621	0.0416	-0.0175	0.0439	0.0774	0.0108
C (pred)	0.0131	-0.0220	-0.0130	0.0099	0.0137	0.0066	0.0408	0.0306	-0.0198	0.0434	0.0567	0.0089
GC (resid)	-0.0001	0.0145	0.0089	0.0026	0.0058	0.0097	0.0212	0.0110	0.0023	0.0005	0.0207	0.0019
HI	0.0243	0.0223	0.0307	0.0344	0.0346	0.0723	0.1168	0.0617	0.0355	0.0658	0.1103	0.0214
CI contrib. of:												
ln(inc)	**0.0173**	0.0206	0.0102	**0.0305**	**0.0250**	**0.0591**	**0.0561**	**0.0288**	**0.0321**	**0.0422**	**0.0793**	0.0054
m30-44	-0.0008	-0.0004	-0.0003	**-0.0039**	-0.0001	0.0015	-0.0017	-0.0006	0.0007	-0.0003	-0.0002	0.0004
m45-59	0.0005	0.0017	-0.0007	-0.0008	0.0003	**0.0054**	-0.0007	0.0002	**0.0019**	0.0000	0.0005	**0.0034**
m60-69	0.0000	-0.0006	0.0005	0.0002	-0.0001	-0.0011	0.0002	**0.0005**	**-0.0017**	-0.0001	-0.0007	0.0000
m70+	0.0000	-0.0029	**-0.0017**	0.0003	0.0002	**-0.0072**	-0.0008	-0.0001	**-0.0084**	**-0.0007**	-0.0017	**-0.0019**
f16-29	**-0.0057**	-0.0022	**-0.0019**	-0.0003	-0.0010	**-0.0024**	0.0010	**-0.0043**	-0.0004	**-0.0018**	**0.0023**	-0.0004
f30-44	-0.0002	0.0034	-0.0006	0.0000	-0.0004	0.0009	0.0000	0.0011	**0.0076**	**0.0028**	**0.0028**	-0.0024
f45-59	**0.0042**	0.0021	0.0018	0.0008	-0.0005	**0.0071**	0.0000	**0.0019**	**0.0022**	-0.0002	0.0012	**0.0056**
f60-69	**-0.0022**	-0.0026	-0.0008	0.0002	-0.0013	-0.0027	-0.0010	**-0.0012**	**-0.0050**	-0.0008	**-0.0025**	**-0.0031**
f70+	-0.0014	0.0024	**-0.0046**	-0.0010	-0.0005	**-0.0088**	-0.0068	**-0.0021**	**-0.0084**	**-0.0010**	0.0010	**-0.0053**
H good	**0.0014**	0.0007	**0.0022**	**0.0045**	0.0001	**0.0048**	**-0.0026**	**0.0027**	**0.0033**	**0.0018**	**0.0093**	0.0003
H fair	**-0.0030**	**-0.0119**	**-0.0101**	**-0.0079**	**-0.0032**	**-0.0194**	**-0.0200**	**-0.0025**	**-0.0170**	**-0.0039**	**-0.0053**	**-0.0029**
H poor	**-0.0016**	**-0.0078**	**-0.0060**	**-0.0047**	-0.0033	**-0.0139**	**-0.0055**	**-0.0085**	**-0.0151**	**-0.0094**	**-0.0271**	**-0.0025**
H v poor	**-0.0011**	**-0.0027**	-0.0012	**-0.0015**	-0.0003	**-0.0028**	**-0.0024**	**-0.0022**	**-0.0058**	**-0.0019**	**-0.0063**	-0.0005
Some lim	-0.0005	**-0.0036**	**-0.0039**	**-0.0025**	**-0.0027**	**-0.0106**	**-0.0139**	**-0.0016**	**-0.0029**	**-0.0036**	**-0.0025**	-0.0002
Severe lim	**-0.0009**	**-0.0066**	**-0.0069**	**-0.0035**	-0.0009	**-0.0086**	-0.0032	**-0.0029**	**-0.0049**	**-0.0027**	**-0.0058**	-0.0008

Decomposition based on linear approximation using the average marginal effects from a logit regression. Significant HI indices and contributions in bold ($p<0.05$).

Table 6.3 (continued) Contributions to inequality in the specialist visit probability

	Germany	Denmark	Netherl	Belgium	Luxemb	UK	Ireland	Italy	Greece	Spain	Portugal	Austria
Second educ	-0.0001	-0.0003	**-0.0009**	0.0001	**0.0042**	**0.0015**	**0.0102**	**0.0088**	0.0015	**0.0025**	**0.0087**	**0.0060**
Higher educ	**0.0070**	**0.0120**	**0.0093**	**0.0092**	0.0069	**0.0127**	**0.0201**	**0.0043**	0.0010	**0.0095**	**0.0175**	**0.0051**
Self-employed	**-0.0011**	-0.0007	**-0.0011**	0.0000	0.0003	**-0.0042**	-0.0007	0.0002	-0.0005	0.0003	**-0.0031**	**0.0008**
Student	-0.0004	**-0.0059**	0.0006	0.0001	-0.0007	-0.0010	0.0000	0.0000	0.0002	0.0000	**0.0012**	-0.0001
Unemployed	-0.0002	-0.0010	-0.0006	-0.0010	-0.0019	-0.0018	0.0046	0.0014	**0.0014**	0.0002	-0.0001	-0.0001
Retired	-0.0014	-0.0063	0.0006	**-0.0047**	**-0.0050**	**-0.0110**	-0.0019	0.0002	**-0.0037**	0.0000	**-0.0080**	0.0002
Houswife	**0.0013**	-0.0007	-0.0002	-0.0013	0.0003	0.0013	0.0025	-0.0005	0.0003	**-0.0027**	-0.0015	0.0006
Oth inactive	-0.0001	-0.0001	0.0015	-0.0007	-0.0005	0.0000	**-0.0065**	**-0.0006**	**-0.0007**	**-0.0017**	**-0.0027**	0.0004
Sep/divorced	0.0004	-0.0012	0.0000	**-0.0017**	0.0001	0.0007	**0.0031**	0.0003	0.0003	0.0001	0.0002	0.0003
Widowed	0.0008	**-0.0063**	-0.0008	**0.0034**	0.0001	0.0036	0.0039	**0.0007**	0.0003	0.0004	0.0005	**0.0015**
Not married	**0.0007**	-0.0011	**0.0027**	-0.0006	-0.0014	-0.0003	0.0012	0.0006	**-0.0019**	**-0.0006**	**-0.0019**	-0.0002
region 2				**-0.0034**		-0.0009	**0.0057**	-0.0004	**-0.0017**	0.0007	0.0008	-0.0006
region 3				0.0001		0.0000		0.0004	**0.0053**	**0.0102**	-0.0015	0.0000
region 4						0.0005		0.0005	0.0004	0.0001	**0.0023**	
region 5						0.0021		-0.0001		**0.0051**	**0.0010**	
region 6						0.0001		**0.0009**		-0.0019	0.0002	
region 7						0.0004		0.0003		0.0005	**-0.0012**	
region 8						0.0002		-0.0009				
region 9						0.0007		0.0013				
region 10						0.0005		**0.0031**				
region 11						0.0002		0.0007				

Decomposition based on linear approximation using the average marginal effects from a logit regression. Significant HI indices and contributions in bold

conditional positive number of visits. As such, we are better able to distinguish between factors driving inequality in initial visits and in subsequent visits. Finally, we illustrate how statistical inference can be based on standard error estimates of the inequality contributions generated with bootstrapping methods.

The results provide a number of new insights. First, we find that in *all* European countries, both the need for GP services and the use of such care are more concentrated among the poorer population segments. But in many cases the actual distribution is even more pro-poor than the need distribution. Violations of the principle of "equal treatment for equal need" by income are very modest: rich and poor face very similar probabilities of seeing a GP when need differences have been adjusted for. Some pro-poor inequity emerges for the conditional number of visits, but it is relatively small. To the extent that the decision for subsequent or repeated visits is more likely to be influenced by the doctor than by the patient, this pro-poor discrimination may be doctor-driven.

Secondly, the findings are dramatically different in the case of specialist visits. While needs are often greater among the poor, specialist use is often higher among the rich or, at best, distributed fairly equally. Consequently, after controlling for the greater needs of the poor, substantial degrees of horizontal inequity favouring the rich emerge in *all* countries. Everywhere in Europe, the use of specialist visits is higher (than expected on the basis of need) for the rich and lower for the poor, but the degree to which occurs differs substantially between countries: the pro-rich pattern is strongest in Ireland and Portugal, and weakest in the Benelux countries. But also the 'decomposition by parts' provides a different picture for specialist visits: the probability of an (initial) visit is much more important than the (conditional) number of (subsequent) visits in generating the observed patterns of income-related horizontal inequities. In most countries, by far the greater share of overall inequity in specialist use stems from the unequal distribution of an initial contact. This would suggest that inequity here is rather more patient-initiated than doctor-driven, although in countries with gatekeeping roles for GPs it may be GP-initiated. Notable exceptions to this rule are Austria and Denmark, where most of the inequity stems from the conditional number of positive visits and may therefore be related to specialist self-referral patterns.

Third, the paper also sheds light on the relative contributions of the factors driving the cross-country differences in inequalities. For GP care utilisation, the most important variables contributing to a more pro-poor distribution are not income itself but rather other indicators of social disadvantage, such as low education, retirement, and non-participation in the labour force. In so far as regional disparities can be captured with our data, they appear relatively unimportant here. This may either be interpreted as some sort of positive discrimination by GPs of lower socio-economic categories but

an alternative and equally plausible explanation is measurement error in the need variables. It is not impossible that self-reporting of morbidity is systematically different among these categories. If these groups were to under-report morbidity compared to some objective measure of health then, for a given level of self-reported morbidity, their needs may actually be greater than those of other, more advantaged groups. Unfortunately, this hypothesis cannot be tested in the absence of a more objective measure of need. In the case of specialist visits, the contribution of income to the pro-rich distribution is much clearer, especially for the probability of seeing a specialist. Particularly in those countries where higher income can buy quicker or preferential access to a medical specialist, this contribution seems to be larger. It can be because those with higher incomes buy supplemental private insurance, as in Ireland, Spain and the UK, or because they are more likely to use the private sector, as in Portugal and Italy. It is less obvious why income also contributes substantially to a pro-rich distribution of specialist visits in a country like Denmark, where both private insurance and private practice (for specialist services) are nearly non-existent. In that country, the horizontal inequity definitely arises within the public system. Among the other non-need variables included in the analysis, education and region stand out as other important contributing factors. In almost all countries, the higher educated (which tend to be richer) also tend to be (much) more inclined to contact a specialist than the lower educated. Whether such medical consumption behaviour is 'more appropriate' is impossible to answer from this analysis, but it does mean that rich and poor do not get the same kind of treatment, given need. If it is the case that, given the same need, specialist visits represent higher quality treatment than GP visits, then the better-off are getting more out of their health care systems than the less well-off.

We conclude by reminding the reader of the limitations of our analysis. First of all, it only refers to differences in quantities of use, not qualities. We cannot but assume that "a visit is a visit" since we have no means of controlling for differences in the quality of doctor visits within or between countries. Adjusting for quality differences (e.g. by distinguishing public from private visits) might make the differentials larger or smaller. A similar remark applies to the appropriateness of care use. We had to assume that the average relationship observed in a country between reported morbidity and use is the norm for "appropriateness of care" and register systematic relative deviations from this norm. In practice, it is almost certain that there are differences between countries in the extent to which such as a norm is indeed "appropriate". Finally, while the ECHP data offer some fascinating new options for cross-European comparisons by coupling rich information on socio-economic characteristics with information on health and health care use, it is still constrained in its coverage. In particular, the limited information on the

type and degree of insurance coverage and the type of health care use precludes a more detailed analysis of the public-private sector interactions in medical care utilisation.

But keeping these limitations in mind, we find that in European countries, despite decades of universal and fairly comprehensive coverage, utilisation patterns suggest that rich and poor are not treated equally. At equal levels of need, the access to and use of specialist services is greater for higher income groups. Only in some countries, like Ireland, Spain or Belgium, this seems to be somewhat compensated by pro-poor patterns in the use of GP care. Unless this finding is a consequence of a deliberate policy to offer such groups private access options over and above their public entitlements, we cannot but conclude that — despite a long tradition of public intervention in health care — there is still some way to go before equals are treated equally in Europe.

Proof of equivalence of one-step and two-step estimator of horizontal inequity index in linear models.

Using a method of indirect standardisation, Wagstaff and van Doorslaer (2000) define an index of horizontal inequity for utilisation (y) as in equation (6.6) as

$$HI_{WV} = C_M - C_N = C(y) - C(\hat{y}) = C(y - \hat{y}) \qquad (6.A.1)$$

where $\hat{y}$ is need-predicted utilisation from regressing y on a vector of need variables x^n. Following Schokkaert and van de Voorde (2000), one could propose that, in order to use the partial effects of the need variables in the standardisation process, the estimated regression should also include a vector of non-need variables x^p, but their effect should be 'neutralized' when generating the need predictions, e.g. by setting these equal to their mean values ($\bar{x}^p$). For example, if

$$y = \beta' x^n + \alpha' x^p + \varepsilon \qquad (6.A.2)$$

and

$$\hat{y}_{PN} = \beta' x^n + \alpha' \bar{x}^p \qquad (6.A.3)$$

then a horizontal index defined on the basis of these partial need effects is

$$HI_{PN} = C(y) - C(\hat{y}_{PN}) \qquad (6.A.4)$$

Substituting (6.A.2) and (6.A.3.) in (6.A.4) gives

$$HI_{PN} = C(\beta' x^n + \alpha' x^p + \varepsilon) - C(\beta' x^n + \alpha' \bar{x}^p) \qquad (6.A.5)$$

Using the covariance definition of a CI (equation (6.1)) and additive separability of covariances

$$HI_{PN} = C(\beta' x^n) + C(\alpha' x^p) + C(\varepsilon) - C(\beta' x^n) - C(\alpha' \bar{x}^p) \qquad (6.A.6)$$

Note that $C(\alpha' \bar{x}^p) = 0$ because the covariance of a constant is zero, and this is true whatever fixed values for x^p are used. So,

$$HI_{PN} = C(\alpha' x^p) + C(\varepsilon) \qquad (6.A.7)$$

This expression is equivalent to the HI definition based on the decomposition methods in equation (6.8) (if income x^r is included in the non-need vector x^p). Note that the equivalence of both approaches relies on the linearity of equation (6.A.2) but not on the choice of the fixed values for x^p.

Cameron, A.C. and P.K. Trivedi (1998). Regression analysis of count data. Econometric Society monographs ; no. 30. Cambridge, UK ; New York, NY, USA: Cambridge University Press. xvii, 411.

Deaton, A. (1997). The analysis of household surveys: a microeconometric approach to development policy. Baltimore: Johns Hopkins University Press.

Deb, P. and P.K. Trivedi (1997). Demand for medical care by the elderly: A finite mixture approach. Journal of Applied Econometrics, 12(3): 313-336.

Efron, B. and R.J. Tibshirani (1993). An introduction to the bootstrap. London: Chapman & Hall.

Eurostat (1999). European Community Household Panel (ECHP): selected indicators from the 1995 wave. European Commission, Eurostat, Luxembourg,

Gerdtham, U.-G. (1997). Equity in health care utilization: further tests based on hurdle models and Swedish micro data. Health Economics, 6(3): 303-19.

Gravelle, H. (2001). Measuring income related inequality in health: standardisation and the partial concentration index. Health Economics, 12(10): 803-19.

Grootendorst, P.V. (1995). A comparison of alternative models of prescription drug utilization. Health Economics, 4(3): 183-198.

Hakkinen, U., G. Rosenqvist, and S. Aro (1996). Economic depression and the use of physician services in Finland. Health Economics, 5(5): 421-434.

Hurst, J.W., Performance measurement and improvement in OECD health systems: overview of issues and challenges, in Measuring Up: Improving Health Systems Performance in OECD Countries, OECD, Editor. 2002, OECD: Paris. p. 35-54.

Jimenez-Martin, S., J.M. Labeaga, and M. Martinez-Granado (2002). Latent class versus two-part models in the demand for physician services across the European Union. Health Economics, 11(4): 301-321

Jones, A.M. and O. O'Donnell (2002). Econometric analysis of health data. Chichester, UK: John Wiley & Sons.

Jones, A.M., Health Econometrics, in Handbook of Health Economics, J.P. Newhouse, Editor. 2000, Elsevier North-Holland: Amsterdam. p. 265-346.

Lerman, R.I. and S. Yitzhaki (1989). Improving the Accuracy of Estimates of Gini Coefficients. Journal of Econometrics, 42(1): 43-47.

Mossialos, E.A.S.T., Voluntary health insurance in the European Union, in Funding health care: options for Europe, E. Mossialos, A Dixon, J Figueras and J Kutzin, Editor. 2002, Open University Press: Buckingham. p. 128-60.

Pohlmeier, W. and V. Ulrich (1995). An econometric model of the two-part decision-making process in the demand for health care. Journal of Human Resources, 30(2): 339-361.

Robinson, R., User charges for health care, in Funding health care: options for Europe, E. Mossialos, A Dixon, J Figueras and J Kutzin, Editor. 2002, Open University Press: Buckingham. p. 161-183.

Schokkaert, E. and C. van de Voorde (2000). Risk selection and the specification of the risk adjustment formula. Centre for Economic

Studies Discussion Paper, 2000/011, University of Leuven, Leuven,

Van Doorslaer, E., A. Wagstaff, and F. Rutten, eds. Equity in the finance and delivery of health care: An international perspective. 1993, Oxford University Press: Oxford.

Van Doorslaer, E., A. Wagstaff, and et al. (1992). Equity in the Delivery of Health Care: Some International Comparisons. Journal of Health Economics, 11(4): 389-411.

Van Doorslaer, E., et al. (2000). Equity in the delivery of health care in Europe and the US. Journal of Health Economics, 19(5): 553-583.

Van Doorslaer, E., X. Koolman, and F. Puffer, Equity in the use of physician visits in OECD countries: has equal treatment for equal need been achieved?, in Measuring Up: Improving Health Systems Performance in OECD Countries, OECD, Editor. 2002, OECD: Paris. p. 225-248.

Van Doorslaer, E., and X. Koolman (2002). Explaining the differences in income-related health inequalities across European countries. ECuity II Project Working Paper, #6, Erasmus University, Rotterdam,

Wagstaff, A. and E. Van Doorslaer, Equity in health care finance and delivery, in Handbook of Health Economics, J.P. Newhouse, Editor. 2000, Elsevier North-Holland: Amsterdam. p. 1803-1862.

Wagstaff, A., E. Van Doorslaer, and P. Paci (1991). On the measurement of horizontal inequity in the delivery of health care. Journal of Health Economics, 10(2): 169-205.

Wagstaff, A., E. Van Doorslaer, and N. Watanabe (2003). On decomposing the causes of health sector inequalities, with an application to malnutrition inequalities in Vietnam. Journal of Econometrics, 112(1): 219-27.

Wagstaff, A. and E. van Doorslaer (2000). Measuring and testing for inequity in the delivery of health care. Journal of Human Resources, 35(4): 716-733.

Wooldridge, J.M. (2002). Econometric analysis of cross-section and panel data. Cambridge, Massachusetts: MIT Press.

CHAPTER 7

Impact of private health insurance on specialist consults

ABSTRACT

The aim of this paper is to use the European Community Household Panel (ECHP) to estimate the impact of having private health insurance coverage on the use of specialist visits in four European countries that have systems which allow supplementary coverage: Ireland, Italy, Portugal and the UK. The central questions are whether access to supplementary private insurance encourages greater utilisation and whether it contributes to horizontal inequity in the use of specialists. Empirical analysis of this issue is complicated by the fact that the decision to take out voluntary health insurance is an individual choice that is likely to be influenced by risk selection. We compare different estimators to correct for this cause of endogeneity. The empirical results show that the probability of having private insurance increases with income and with better reported health. Private insurance has a positive association with the probability of specialist visits in all countries although the magnitude of the effect is sensitive to the choice of method. These findings imply that private insurance contributes to 'pro-rich' horizontal inequity in the use of specialist visits.

7.1 INTRODUCTION

Recently, a number of studies have reported that in many, if not all, European and OECD countries the probability of consulting a medical specialist is positively related to income, after controlling for differences in need for such services. Van Doorslaer, Koolman and Jones (2004) found a significant and substantial degree of pro-rich inequity in the probability of seeing a specialist for all European countries but did not explicitly examine the role and contribution of private insurance options in these countries. Van Doorslaer, Koolman and Puffer (2003) showed that the degree of such pro-rich inequity was reduced when access to private insurance was controlled for, indicating that private insurance has a pro-rich contribution to inequity. This was confirmed in a study by Van Doorslaer, Masseria *et al* (2004) who estimated significant pro-rich contributions of private insurance coverage when decomposing the degree of inequity in the probability of a specialist visit in France, Ireland, Switzerland and the UK.

From a policy point of view, however, it is of critical importance to know whether, and to what extent, the contribution of insurance is due to selection effects or to a direct utilisation effect or both. If the observed effect of insurance were entirely due to self-selection of those more likely to use specialists, then private insurance coverage merely acts as a marker for such propensity and reducing private insurance options will not reduce the pro-rich distribution of care. If, on the other hand, the association is mostly due to the effect of access to insurance on utilisation, then the expansion or reduction of private insurance options *will* have an impact on the degree to which care is distributed by income. Therefore, a central question becomes whether access to private insurance encourages greater utilisation or not. In general, insurance coverage may influence utilisation for a variety of reasons:

- A *moral hazard effect*; that the level of utilisation is greater when insurance reduces the out-of-pocket price for health care (Pauly, 1968, Manning et al., 1987, Coulson et al., 1995, Chiappori et al., 1998).
- A *risk reduction effect*; that the desired level of utilisation is greater under the financial certainty created by insurance than under uncertainty (Meza, 1983, Vera-Hernández, 1999).
- An *income transfer effect*; that insurance creates an ex post transfer of income from the healthy to the ill and this may increase utilisation through an income effect on the demand for medical care (Pauly, 1968, Nyman, 1999a, Nyman and Maude-Griffin, 2001).
- Another dimension of the income effect is the *access effect*; that insurance may extend an individual's opportunity set by giving access to health care that would not otherwise be affordable to them. Nyman (1999b) has argued that the pooling effect of insurance provides access to expensive

> medical technologies that would not be affordable out-of-pocket. In the context of supplementary private insurance in European systems, private insurance may provide access to a 'quality' of care that is not provided by the public system, for example offering reduced waiting times for elective surgery (see e.g., Jofre-Bonet, 2000).

All of these factors may encourage greater utilisation and they will be referred to collectively as the *insurance effect* on utilisation. We are not able to distinguish between the four factors in our empirical analysis of the insurance effect, but the relevance of each factor will depend on the specific health care system under investigation. For example, in a system where private insurance provides supplementary cover alongside a universal public system - that offers a basic package of services that are free at the point of use - the access effect may be the prime reason for increased utilisation among those with private cover (see e.g., Shmueli, 2001). Indeed, a study by Harmon and Nolan (2001) for Ireland finds that the most important reasons quoted by respondents for buying private health insurance were "being sure of getting into hospital quicker when you needed treatment" and "being sure of getting consultant care". So it is likely that the access effect will dominate in our analysis of supplementary insurance; the insurance will mainly reflect a demand for amenity and quality of care that can be purchased through private insurance. It is important to emphasise that our aim is to measure the effect of having, versus not having, private insurance. We do not have information on the choice of different insurance plans or on different levels of coverage, through deductibles and copayments. The potential for adverse selection of risks may be more relevant for analysis of plan choice and levels of coverage (see e.g. Gardiol et al., 2005).

Our aim is to estimate the impact of having private health insurance on the use of specialist visits in four European countries that have systems which allow supplementary coverage and that have suitable data for our empirical methods available in the European Community Household Panel (ECHP): Ireland, Italy, Portugal and the UK.[a] This insurance can take the form of providing cover for services that are not available in the public system, including coverage for copayments charged in the public system. Mossialos and Thomson (2002) refer to this as 'complementary' insurance. Alternatively private insurance may supplement the public system by providing access to different services, such as increased amenities and reduced waiting time. Mossialos and Thomson (2002) refer to this as 'supplementary' insurance. As supplementary private insurance covers services that are also funded through the public system, this is sometimes known as 'double coverage' (see e.g., Vera-Hernández, 1999). Table 7.1 provides a summary of the kind of private insurance available and the estimates of its

Table 7.1 Coverage of complementary and supplementary private health insurance (sources: Mossialos and Thomson (2002) and Colombo & Tapay (2004))

			% pop covered	
	Complementary	Supplementary	Mossialos & Thomson (2002)	Colombo & Tapay (2004)
Ireland	Copayments for outpatient care	Care in private hospitals, consultants, private beds, some outpatient costs	42.0	43.8
Italy	Hospital convalescence, outpatient costs	Upgraded hospital accommodation, free choice of doctor, diagnostic services/specialist visits	5.0	15.6
Portugal		Cash benefits for hospital care, total coverage for other treatments, free choice of doctor or hospital	10.0	14.8
UK	Dental care, alternative therapies	Upgraded hospital accommodation, cash benefits, private beds, care in private hospitals	11.5	10.0

prevalence in each country. The table makes it clear that possession of private insurance may provide easier access to specialists in all of the countries under investigation. The demand for private insurance is likely to reflect factors that influence the demand for enhanced quality of care, which may include more immediate and responsive treatment, rather than a price effect of different levels of coverage.

Empirical analysis of international differences in the impact of private insurance on use of specialists is complicated by the fact that the decision to take out voluntary health insurance is an individual choice that is likely to be influenced by unobservable individual characteristics, such as the individual's level of 'risk'. This may give rise to *adverse selection* where those with higher risk are more likely to take out private health insurance and to make more use of health care (e.g. Van de Ven, 1987, Coulson et al., 1995, Ettner, 1997, Chiappori et al., 1998). It may also create an incentive for *risk selection* (or cream skimming) where private insurers attempt to attract good risks to their policies and to avoid bad risks (e.g. Coulson et al., 1995, Ettner, 1997). Shmueli (2001) provides empirical evidence of the offsetting effects of adverse selection and risk selection by insurers on the ownership of supplementary insurance in Israel: sicker individuals are more likely to apply for insurance but also more likely to be rejected, such that there are no health effects in the reduced form for ownership of supplementary insurance.

In the United States, the RAND Health Insurance Experiment took an experimental approach to deal with the problem of selection in plan choice, with participants

randomised to different insurance plans (Manning et al., 1987). Chiappori et al. (1998) were able to exploit a quasi- or natural experiment in which French insurees faced an exogenous change in levels of copayment. With the ECHP we have to rely on non-experimental data and follow the general approaches for dealing with endogeneity of earlier studies (see e.g., Cameron *et al.*, 1988; Coulson *et al.*, 1995; Holly *et al.*, 1998; Vera-Hernández, 1999; Schellhorn, 2001; and Buchmueller *et al.,* 2004).

The aim is to separate the *insurance effect* from the *selection effect*. The problem of distinguishing adverse selection from moral hazard in observational data is well known and presents a severe challenge, particularly with cross section data (see e.g., Chiappori, 2000; Geoffard, 2006). Our approach is to compare a variety of empirical methods for estimating casual effects in the presence of selection bias and see whether they provide robust estimates under different identifying assumptions. Our empirical strategy exploits the longitudinal data available in the ECHP and estimates the partial effect of ownership of private insurance on the probability of using specialist care by a variety of methods: a standard probit model, propensity score matching (PSM) and a recursive structural model for binary measures of health insurance and specialist visits estimated using full-information maximum likelihood (FIML).

7.2 ECONOMETRIC STRATEGIES

Our empirical methods are based on binary variables for whether an individual (*i*) has private health insurance during year t (y_{it}^1) and whether they have visited a specialist at least once during the last year (y_{it}^2). To gain statistical power we pool the waves and compute robust standard errors allowing for clustering within individuals to take account of the repeated measurements on each individual. We compare estimators based on two different identification strategies: 'selection on observables' and 'selection on unobservables'.

7.2.1 Simple probit model

Our baseline estimate is given by the partial effect of private health insurance in a simple probit model for at least one visit to the specialist. As well as insurance coverage, the model for any specialist visit conditions on a set of observable individual characteristics (x) including a constant equal to one which are described in more detail in Section 7.3 below. So,

$$P(y_{it}^2 = 1 \mid y_{it}^1, x_{it}) = \Phi(\gamma y_{it}^1 + \beta' x_{it}) \qquad (7.1)$$

where $\Phi(.)$ is the standard normal distribution function. The average partial effect (APE) of insurance on specialist visits (the *'insurance effect'*) is computed by taking the

sample mean of the partial effect (PE) for each individual observation. The partial effect is,

$$PE_i = \Delta P\,(y_{it}^2 = 1 \mid y_{it}^1, x_{it}) / \Delta y_{it}^1 = \Phi(\gamma + \beta' x_{it}) - \Phi(\beta' x_{it}) \quad (7.2)$$

Then the average partial effect (APE) is given by the mean across the sample observations,

$$APE = \frac{1}{nT} \sum_i \sum_t \Delta P\,(y_{it}^2 = 1 \mid y_{it}^1, x_{it}) / \Delta y_{it}^1$$
$$= \frac{1}{nT} \sum_i \sum_t \{\Phi(\gamma + \beta' x_{it}) - (\beta' x_{it})\} \quad (7.3)$$

As well as reporting the average effect, the availability of individual-specific partial effects allows us to explore heterogeneity in the effect across individuals, for example, by displaying a histogram of the effects.

7.2.2 Propensity score matching (PSM)

Matching provides a more general approach to deal with selection on observables. It addresses the problem that in the observed data confounding factors (matching variables) are non-randomly distributed over the treated and control individuals. Rosenbaum and Rubin (1983) showed that, rather than matching on an entire set of observable characteristics, the dimensions of the problem could be reduced by matching on the basis of their probability of receiving treatment, $P\left(y_{it}^1 = 1 \mid x_{it}, z_{it}\right)$, known as the *propensity score*. In practice the estimators do not rely on exact matching and instead weight observations by their proximity, in terms of their propensity score.

We construct the propensity scores using a probit model for private and use predicted probability of treatment,

$$P_{it} = \Phi(\alpha' x_{it} + \varphi' x_{it}) \quad (7.4)$$

We match treated individuals with non-treated individuals inversely weighted for the distance in terms of propensities, $\{\hat{P}_{it} - \hat{P}_{jt}\}$. More precisely, weights are constructed using kernel smoothed distance weighting. The Epanechnikov kernel is used as it is computationally convenient and efficient. We ensure that all cases are supported by controls. The quality of the matching can be assessed by computing the reduction of the pseudo R-squareds of the insurance regression before and after matching (Table 7.A.2). To evaluate the extent to which matching on propensity scores balances the distribution of the *x*'s between the insured and the uninsured group, we computed the bias reduction due to matching for each of the *x*'s (Table 7.A.3).

Following the matching we use the relative weights of the treated and controls to compute the insurance effect as the weighted mean difference in the probability of at

least one specialist visit between the two groups. It should be noted that an important requirement is that th«e participation model, used to construct the propensity score, should only include variables that are unaffected by participation, or the anticipation of participation (Heckman, LaLonde and Smith, 1999). This suggests that matching variables should be either time invariant characteristics or variables that are measured before participation in the treatment and that are not affected by anticipation of participation.

7.2.3 The FIML estimator

The approaches described above rely on the notion of selection on observables. In contrast, selection on unobservables, also termed informative, non-random or non-ignorable selection, is familiar in the econometrics literature where the dominant approaches follow the instrumental variables and the sample selection or control function approaches (Heckman, 1976). These approaches typically rely on there being "instruments" (w) that are good predictors of the endogenous treatment and that satisfy an exclusion restriction: that they do not have a direct effect on the outcome of interest.

Here we adopt a structural approach with full information maximum likelihood estimation, based on FIML estimates of a recursive bivariate probit model. The first issue in specifying a structural model for insurance and specialist visits is how to specify a coherent econometric model that allows for the potential endogeneity of insurance. A similar question is addressed by Windmeijer and Santos Silva (1997), who use data from the 1991 British Health and Lifestyle Survey to investigate nonlinear simultaneous equations models for GP visits in which self-assessed health is treated as an endogenous binary regressor. They adopt Blundell and Smith's (1993) framework, and compare type I and type II specifications. In the type II model, recorded health status is assumed to influence GP visits. In the type I model it is the latent health index that influences the number of visits. The coherency conditions for the type II model imply that the model is only logically consistent when it is specified as a recursive system. In other words, the type II specification can only be coherent when the endogeneity of self-assessed health stems from unobservable heterogeneity bias rather than classical simultaneous equations bias.

In our application a type II specification makes more sense than a type I specification: we want to identify the impact of *actually having* private insurance on specialist visits rather than the impact of the propensity to have insurance. For this reason we adopt a recursive model in which insurance coverage during the year is assumed to influence the probability of a specialist visit during the subsequent year. This exploits the longitudinal data available in the ECHP. The chronology of events means that the use of specialist

visits cannot have a direct feedback effect on the decision to take out insurance in the previous year, thus ruling out simultaneity bias. Insurance may still be an endogenous regressor due to unobservable heterogeneity, such as an individual's level of risk or risk aversion, that has a direct influence on both their decision to take out insurance and their use of health care in the subsequent wave. This unobservable heterogeneity can be captured by using a bivariate probit specification.

The bivariate probit model applies to a pair of binary dependent variables and allows for correlation between the corresponding error terms. In our application, the use of specialist visits is modelled as a recursive bivariate probit model (see e.g., Maddala, 1983 p.123; Holly et al., 1998; Waters, 1999; Greene, 2000 p.852; Buchmueller et al., 2004). The model consists of two latent variable equations for insurance and specialist visits:

$$y^{*1}_{it} = \alpha' x_{it} + \eta' w_{it} + \varepsilon^{1}_{it} \quad (7.5)$$

$$y^{*2}_{it} = \gamma y^{1}_{it} + \beta' x_{it} + \varepsilon^{2}_{it} \quad (7.6)$$

where, $(\varepsilon^1,\varepsilon^2) \sim N(0, \Omega)$ and $y^j = 1$, if $y^{*j} > 0$ and $y^j = 0$ otherwise.

Our identification strategy relies on the fact that we are modelling sequential decisions: utilisation is a function of prior uptake of insurance which depends on lagged values of the regressors. Estimation of the model by FIML, taking account of the joint distribution of ε^1 and ε^2 deals with the endogeneity of y^1 (see Holly et al., 1998, Greene, 2000 p.852). The log-likelihood for the model is,

$$\log L = \sum_{i=1}^{n} \sum_{t=1}^{T} \Phi \{ d^{1}_{it} (\alpha' x_{it} + \eta' w_{it}), d^{2}_{it} (\gamma y^{1}_{it} + \beta' x_{it}), d^{1}_{it} d^{2}_{it} \rho \} \quad (7.7)$$

where $\Phi[.]$ is the bivariate normal CDF, $d^j = 2y^j - 1$ and ρ is the coefficient of correlation between ε^1 and ε^2. The asymptotic t-ratio for the estimate of ρ provides a test for exogeneity. The partial effects of insurance in this model can be computed from the marginal distribution for specialist visits, using the same formula as the univariate probit, but with the parameter estimates from the bivariate probit model.

Wilde (2000) shows that, given the full rank of the regressor matrix, it is only necessary to have varying exogenous regressors to avoid identification problems in this recursive bivariate probit model and exclusion restrictions are not required. This identification by functional form relies on the assumption of normality and it is common practice to impose additional exclusion restrictions to improve identification. In our model the regressors in the insurance equation are measured at the previous wave, while those in the utilisation equation are measured at the current wave: so identification relies on the timing of events. In addition, we add lagged information for each household on whether at least one individual's employer provides free or subsidized health care or medical

insurance to the model for insurance coverage. According to Mossialos and Thomson (2002), group contracts now account for almost all voluntary health insurance policies in Portugal, well over two-thirds in Ireland and Italy and more than 50% of voluntary health insurance subscriptions in the UK. Previous studies have included the price of insurance as a predictor of choice of insurance (see e.g., Cameron and Trivedi, 1991, p.18). Price cannot be measured directly in our data but institutional differences in access to private insurance, in particular whether the respondent's employer offers access to insurance, provide a proxy for the cost of acquiring insurance.[b]

7.3 DATA AND VARIABLE DEFINITIONS

The data used in this paper are taken from the *European Community Household Panel User Database* (ECHP-UDB). The ECHP was designed and coordinated by Eurostat, the European Statistical Office and is a standardised multi-purpose annual longitudinal survey carried out at the level of the European Union (see e.g., Peracchi, 2002). The survey is based on a standardised questionnaire that involves annual interviewing of a representative panel of households and individuals of 16 years and older in each of the participating EU member states. It covers a wide range of topics including demographics, income, social transfers, health, housing, education and employment. The first wave was conducted in 1994, Austria was added to the second wave in 1995, and Finland to the third in 1996. The analysis presented in this paper is based on information for Ireland, Italy, Portugal, and the UK for the first four waves, 1994-97, when insurance data are available.

Our binary indicator of private insurance coverage is based on question PH013 which asks whether the respondent is covered by private medical insurance. Our measure equals 0 if they are not insured and 1 if they covered in their own name or through another family member at both the beginning and the end of the year in question.[c] Our binary indicator of specialist visits is based on question PH009 which asks for the number of times the person has consulted a specialist over the past 12 months.[d]

We have followed previous studies of the demand for health insurance (e.g., Propper, 1989; Cameron and Trivedi, 1991; Rhine and Ng, 1998; Shmueli, 2001) and health care (e.g., Cameron et al., 1988; Pohlmeier and Ulrich, 1995; Hakkinen et al., 1996; Gerdtham, 1997; Vera-Hernández, 1999; Harmon and Nolan, 2001; Schellhorn, 2001; Buchmueller *et al.,* 2004; Rodríguez and Stoyanova, 2004) in selecting a set of explanatory variables from the information available in the ECHP. The variables that are common to both the insurance and the specialist visits equation include the logarithm of equivalised household income, indicators of the respondent's education, gender, age, activity, marital and health status. Health status is measured by self-assessed health, on a 5-point categorical scale from very bad to very good, and by whether and to what

degree the respondent is hampered in their daily activities by a physical or mental health problem, illness or disability.

Our income measure is the log of disposable household income per equivalent adult, using the modified OECD equivalence scale. This scale gives a weight of 1.0 to the first adult, 0.5 to the second and each subsequent person aged 14 and over, and 0.3 to each child aged under 4 in the household. Total household income includes all the net monetary income received by the household members during the reference year. It includes income from work (employment and self-employment), private income (from investments and property and private transfers to the household), pensions and other direct social transfers received. In the income measure available in the ECHP-UDB, no account has been taken of indirect social transfers (e.g. reimbursement of medical expenses), receipts in kind and imputed rent from owner-occupied accommodation. Education is measured by the highest level of general or higher education completed and is available at three levels: recognised third level education (ISCED 5-7), second stage of secondary level of education (ISCED 3) and less than second stage of secondary education (ISCED 0-2). The ECHP coding is based on the pre-1997 ISCED.

Table 7.2 presents sample sizes and descriptive statistics for some of the key variables. The 'eligible sample' size includes all eligible cases. The 'analysis sample' is the selection of the eligible cases without missing values for any of the variables used in any of the models. The means for current insurance coverage, and coverage throughout the past year - show that the proportion of people with private insurance coverage is fairly stable over the waves. The difference between current coverage and coverage over the whole year shows that - in Italy and Portugal– substantial proportions of respondents are reported as switching during the year. The highest penetration of private insurance is in Ireland and the lowest is in Italy. The prevalence of at least one specialist visit during the year is fairly similar on average, with Ireland lower (21% on average) than the other countries (31-37%). The final column shows that the prevalence of employer provided

Table 7.2 Descriptive statistics for key variables

	Sample proportion with:					
	Eligible Sample	Analysis Sample	Specialist visit last year	Private cover now	Private cover now and last year	Employer provided health benefits
Ireland	14355	7714	0.21	0.48	0.44	0.23
Italy	34330	19016	0.37	0.06	0.03	0.20
Portugal	23327	13445	0.34	0.10	0.04	0.19
UK	6940	3332	0.31	0.22	0.17	0.36

health benefits to at least one household member, is higher — often much higher — than the prevalence of actual individual PI cover reported in all countries except Ireland.

7.4 RESULTS

Before considering the regression results, some diagnostic tests are reported in the Appendix. Table 7.A.1 shows the pseudo R-squareds for the model we use to estimate the propensity scores before and after matching (the weighted model). Even though the pseudo R-squareds reduce dramatically when the cases are IP-weighted, the LR-based chi-squared test shows that the combination of variables remains significant in all countries but the UK. This imbalance can be studied in more detail by looking at the bias (unbalance) between the treated and the non-treated in the distribution of the covariates (definitions given in Table 7.A.2). Table 7.A.3 shows that the percentage bias reduction is often substantial, but varies considerably between the covariates and is frequently not close to 100%. As a result, residual confounding may remain an issue.

The results for the *insurance effect* on utilization are presented as the average partial effect (APE) of private insurance coverage on the probability of a specialist visit, in Table 7.3. We also present the histograms of individual PE's in Figure 7.1 to provide an indication of the individual heterogeneity in these effects and the shifts in their distributions across different estimators. We compare simple binary probit models for the probability of a specialist visit with propensity score matching weighted estimates and FIML estimates of the bivariate probit model.

7.4.1 Insurance choice

All countries show a positive income effect on the probability of having private insurance and these estimates are highly statistically significant (see Tables A5 and A6).[e] This finding has implications for equity in the use of health care. In the health economics literature horizontal inequity in the use of health care has typically been measured by concentration indices (see e.g., Wagstaff and van Doorslaer, 2000, van Doorslaer, Koolman and Jones, 2004). Wagstaff *et al.* (2002) present methods for decomposing concentration indices to give the contribution of explanatory factors, such as health insurance. The contribution of a factor depends on the product of the elasticity of health care with respect to the factor and the concentration index for the factor itself. In turn, the sign of this concentration index depends on the covariance between the factor and an individual's relative rank in the distribution of income. So, a positive elasticity of specialist visits with respect to insurance, coupled with a positive covariance between insurance and income rank, would imply that insurance contributes to 'pro-rich' inequity in the use of specialist visits.

Previous studies have tended to find only limited evidence of a relationship between observed health status and choice of supplemental private insurance and not to find support for the existence of adverse selection (e.g. Cameron et al., 1988, Cameron and Trivedi, 1991, Ettner, 1997, Hurd and McGarry, 1997, Vera-Hernández, 1999). This is borne out in our results. There are statistically significant effects of self-assessed health in Ireland and of being hampered by health problems for Italy, Portugal and Spain. But the striking thing about the estimated effects is that we find that, in general, those in poorer health are *less* likely to have insurance. This contradicts the notion of adverse selection effects, with respect to health indicators that are observed in the survey, which would suggest the opposite finding. Of course the existence of adverse selection would rely on these indicators of health being known to the respondent but not their insurer. The result may reflect the fact that selection of good risks by insurers (probably through employment) more than offsets adverse selection (Shmueli, 2001). The finding is also consistent with the notion that we are modelling the demand for amenity and quality of care provided by access to supplementary insurance, if this demand is higher among better-off and healthier individuals. Employer-offered health benefits, has a statistically significant positive effect on private insurance for all countries. When it was included in the specialist equation, it was never significant (not shown).

7.4.2 Specialist visits

The full results for the specialist visits equations are given in Tables 7.A.4 and 7.A.5 in the Appendix. Table 7.3 provides a summary or the estimated average partial effect (APE) of insurance on the probability of using a specialist. The first column of Table 7.3 shows the (pooled) binary probit estimates of the partial effect of insurance on the probability of a specialist visit, which do not allow for the endogeneity of insurance. For ease of comparison, bootstrapped standard errors of the APEs were generated for all methods using identical bootstrap procedures. These procedures incorporate the entire estimation procedure, i.e. both the private insurance equation and the specialist

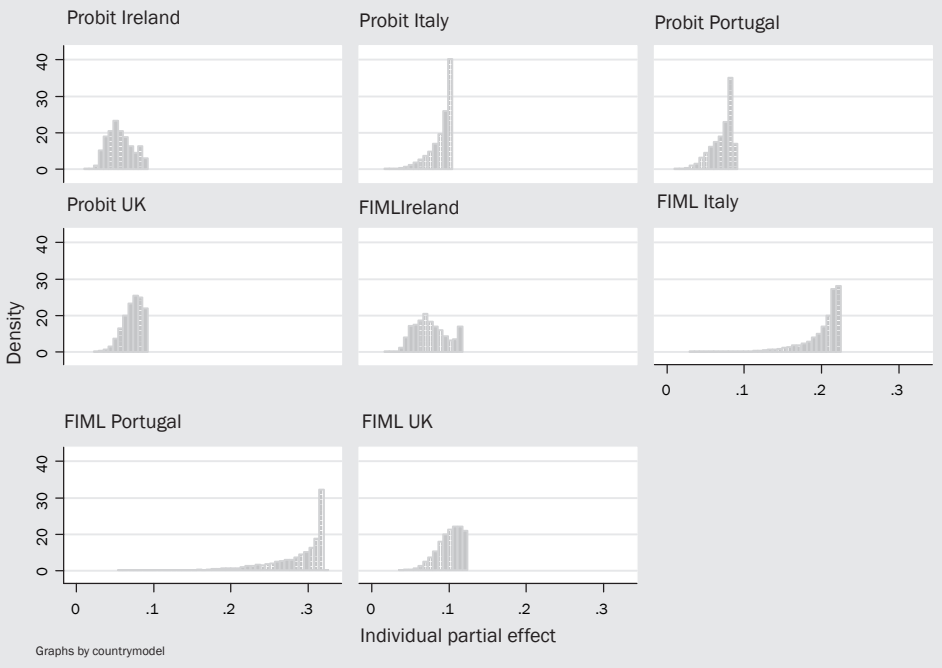
Table 7.3 Estimates of the average partial effect (APE) of insurance on the probability of visiting a specialist *(t-values in parentheses based on bootstrapped standard errors)*

	Probit		PSM		FIML	
Ireland	0.056	(5.17)	0.068	(5.71)	0.076	(2.51)
Italy	0.090	(4.06)	0.124	(4.38)	0.199	(1.80)
Portugal	0.072	(3.26)	0.041	(1.67)	0.283	(2.92)
UK	0.073	(3.38)	0.058	(2.18)	0.100	(1.54)

visit equation when applicable. The bootstrapping itself entails 50 random draws with replacement (resamples) from the sample, where each resample has a size equal to the sample. The APE was computed for each of the resamples, and the distribution of the 50 APEs was used to compute standard deviation. These standard deviations were then used to construct t-values. All countries show a significantly positive partial association between the ownership of private insurance and the probability of a specialist visit. The probability of visiting a specialist among those who report to have had private insurance cover both at the beginning and the end of last year is, on average, about 0.07 higher than among those without cover, with estimates ranging from 0.05 in Ireland to 0.09 in Italy.

When comparing propensity score matching (PSM) to the default probit results, we see that Ireland and Italy show stronger insurance effects and Portugal and the UK smaller effects. The full results for FIML are given in Table 7.A.5. The estimates of rho show that the null of exogeneity is only rejected for Portugal. However, for all countries, the bivariate probit model provides a higher estimate of the insurance effect than the probit. The increase in the insurance effect is more important in Italy and Portugal than in the UK and Ireland. The estimates for Portugal in particular suggest that, because of positive selection, the unadjusted insurance effect *under*estimates the utilisation

Figure 7.1 Histograms of individual partial effects (PE's) of insurance on the probability of visiting a specialist for probit and FIML estimates



effect occurring as a result of private insurance coverage. This positive selection (on unobservables) could be the result of occupational choice or – what is more likely – be a consequence of cream-skimming by insurers through the employment-tied group insurance purchases. While, on average, the insured are lower risk, once they are insured they do seem to exploit their additional coverage through increased use of specialist care.

7.5 CONCLUSION

In this paper we have used the European Community Household Panel (ECHP) to estimate the impact of private health insurance coverage on the use of specialist visits in four European countries that allow for supplementary private insurance coverage: Ireland, Italy, Portugal and the UK. The evidence suggests that the probability of having private insurance increases with income and, to some extent, with better health. For all countries there is a positive insurance effect on the use of specialists. The fact that the probability of having private insurance increases with income, coupled with the fact that having private insurance increases the probability of seeing a specialist means that private insurance contributes to 'pro-rich' inequality in the use of specialists.

For policy purposes, however, it is of relevance to know to what extent this insurance effect is a result of selection, and to what extent it represents a genuine utilisation effect of additional cover. If it is entirely driven by selection, changing the availability of private insurance coverage will not alter the degree to which the use of specialists is related to income. If, on the other hand, the insurance effect is mostly a direct effect on utilisation, then expansion or reduction of access to private insurance (through tax or other incentives) will have an effect on the distribution of specialist care by income. Our findings suggest that expansion or reduction of private insurance will, through its effects on utilisation, have an important effect on the degree to which specialist care gets distributed by income. The results also appear consistent with the observation that private insurance is often obtained as a (group) fringe benefit in certain employment contracts. While, for insurers, this may result in cream-skimming, our analysis does show that, once insured, the beneficiaries are more likely to consult a medical specialist than they would have done in the absence of such coverage. As a result, our findings suggest that the presence of supplementary private coverage has consequences for the degree of horizontal inequity in the use specialist visits. Private insurance is not simply a marker of a higher propensity to consume specialist care but induces additional use over and above what would be used in the absence of such cover.

Notes

a Empirical analysis (e.g. Spain).

b Vera-Hernández (1999) uses measures of social class and occupation as instruments for insurance choice. In his analysis of Swiss data Schellhorn (2001) uses measures of the availability of supplementary insurance cover and of differences in premium levels between cantons as instruments. Buchmueller et al. (2004) use an indicator for public employees along with measures of labour market and occupational status.

c There was an important change in the wording of this question between wave 1 in 1994 – Are you medically insured, either in your own name or through another family member? – and wave 2 in 1995 – Are you (also) covered by private medical insurance, whether in your own name or through another family member?

d Information is available in the ECHP on the total number of visits to a specialist during the year. Our binary measure can be thought of as the first part of a 'two-part' model of utilisation. Van Doorslaer et al. (2004) estimate reduced form versions of a two-part model for specialist visits, that do not include a measure of private insurance. They find that most of the pro-rich inequity in the use of specialists is attributable to the probability of contact rather than the number of visits.

e The full results for all of the equations and each country are presented in the Appendix. This paragraph refers in particular to the insurance equation in the FIML estimates.

Diagnostic and regression results by country

Table 7.A.1 Summary PSM Balancing test

	Pseudo R2	Chis2	p>chi2
Ireland	0.251	2658.8	0.000
	0.007	75.9	0.000
Italy	0.138	645.5	0.000
	0.031	805.3	0.000
Portugal	0.123	503.8	0.000
	0.027	497.9	0.000
UK	0.127	386.6	0.000
	0.002	8.3	0.998

Table 7.A.2 Definitions of variables

Variable name	Definition
specialist	1 if visited a specialist, 0 otherwise
private insurance	1 if privately insured, 0 otherwise
age2635	1 if age between 26 and 36, 0 otherwise
age3645	1 if age between 36 and 46, 0 otherwise
age4655	1 if age between 46 and 56, 0 otherwise
age5665	1 if age between 56 and 66, 0 otherwise
age66plus	1 if age older than 66, 0 otherwise
male	1 if male, 0 if female
sahgood	1 if self assessed health is good, 0 otherwise
sahfair	1 if self assessed health is fair, 0 otherwise
sahbad	1 if self assessed health is bad, 0 otherwise
sahvbad	1 if self assessed health is very bad, 0 otherwise
hampsome	1 if hampered to some extend, 0 otherwise
hampsev	1 if severely hampered, 0 otherwise
selfemploy	1 if self employed, 0 otherwise
student	1 if student, 0 otherwise
unemployed	1 if unemployed or inactive, 0 otherwise
retired	1 if retired, 0 otherwise
housework	1 if housekeeping, 0 otherwise
separated	1 if separated, 0 otherwise
divorced	1 if divorced, 0 otherwise
widowed	1 if widowed, 0 otherwise
never married	1 if never married, 0 otherwise
ln(income)	log modified OECD equivalised household income
secondary educ	1 if second stage of secondary education, 0 otherwise
tertiary educ	1 if tertiary education, 0 otherwise
employer benefit	1 if employer provided health care benefits to household member, 0 otherwise
{variablename}L	value of variable of preceding year

Table 7.A.3 PSM Balancing test

		Ireland		Italy		Portugal		UK	
		% bias	% bias reduction	% bias	% bias reduction	% bias	% bias reduction	% bias	% bias reduction
age2635	Unmatched	-7		3.4		7.8		-8.4	
	Matched	-2.1	70.4	3	11.7	3.6	53.6	0.1	98.8
age3645	Unmatched	20.6		29.8		25.4		12.3	
	Matched	2.6	87.3	10	66.6	7.5	70.6	-2.2	82
age4655	Unmatched	21.4		1.1		-0.9		12.4	
	Matched	0.9	95.9	1.8	-65.9	-1.4	-44.8	2.3	81.9
age5665	Unmatched	4.2		0.2		-16.7		5.5	
	Matched	3.6	13.5	-1.4	-627.1	-5.9	64.8	1.5	73
age66plus	Unmatched	-13.4		-8.5		-30.7		-8.5	
	Matched	4.4	67.2	-2.7	68	-11.9	61.3	1.3	85
male	Unmatched	-6.3		28.9		21		7.9	
	Matched	-10.2	-63	10.8	62.6	8.3	60.8	-0.2	97.5
sahgood	Unmatched	-3.3		8.4		14.8		3.7	
	Matched	0.3	92.3	1.9	77.6	5.4	63.4	-0.7	80.7
sahfair	Unmatched	-13.7		-8		1		-16.8	
	Matched	-0.5	96.7	-2.7	65.9	0.5	51.5	-0.1	99.7
sahbad	Unmatched	-14.1		-13.7		-23.7		-15.5	
	Matched	-1.2	91.4	-5.6	59.2	-8.8	62.9	0.1	99.1
sahvbad	Unmatched	-5.1		-2.7		-7.1		NA	
	Matched	1.7	66.1	-0.9	67.6	-2.4	65.6	NA	NA
hampsome	Unmatched	-11		-5		-6.8		-6.1	
	Matched	0.4	96	-2.1	58.7	-1.3	80.3	1.1	82.4
hampsev	Unmatched	-10.2		-4.8		-15.6		-14.8	
	Matched	-0.7	93	-2	58	-6.3	59.6	-0.3	97.7

Table 7.A.3 (continued) PSM Balancing test

		Ireland		Italy		Portugal		UK	
		% bias	% bias reduction	% bias	% bias reduction	% bias	% bias reduction	% bias	% bias reduction
selfemployed	Unmatched	6		53.3		-12.4		-3	
	Matched	-8.9	-48.4	14.2	73.3	-4	68.2	0.2	93
student	Unmatched	0.6		-24.3		2.7		-17.1	
	Matched	-4.8	-764.3	-9.3	61.7	2.9	-8.9	-1.1	93.5
unemployed	Unmatched	-35.3		-30.1		-27.2		-19.3	
	Matched	-1	97.2	-12.9	57.1	-10.5	61.2	-0.2	98.7
retired	Unmatched	-1.1		-8.7		-28.8		-6.9	
	Matched	4.8	-352.2	-1.7	80.9	-11.2	61.1	-1.6	77.5
housework	Unmatched	-11.1		-33.3		-31.1		-2.6	
	Matched	9.7	12.8	-12.6	62.3	-12.8	59	4	-53.3
separated	Unmatched	-6.7		5.2		2.3		-12.1	
	Matched	2.1	68.2	1.2	77.6	2.3	1.9	-0.5	95.7
divorced	Unmatched	-5.2		6		5.6		-11.7	
	Matched	0.3	93.5	0.8	86.5	-0.2	96.5	-1.5	86.8
widowed	Unmatched	-14.3		-9		-26.7		-17.1	
	Matched	-0.2	98.8	-3.2	64.5	-10.4	61.1	-0.9	95
never married	Unmatched	-41.2		-32.1		-14.3		-21	
	Matched	-10.3	74.9	-13	59.6	-4.7	67.4	-6.3	70
ln(income)	Unmatched	83.6		79.2		76.3		70.8	
	Matched	-0.4	99.5	30.1	62	23.5	69.1	0.8	98.9
secondary educ	Unmatched	21.5		30.9		19.7		2.8	
	Matched	-1.7	92.2	14.4	53.4	10.6	46.4	0.6	76.8
tertiary educ	Unmatched	56.7		33		55.6		39.1	
	Matched	0.9	98.4	8.5	74.2	4.8	91.5	0.9	97.6

Table 7.A.4 Specialist contact probit regression coefficients

	Ireland		Italy		Portugal		UK	
	coeff	t-value	coeff	t-value	coeff	t-value	coeff	t-value
private insurance	0.218	5.40	0.257	4.31	0.217	3.39	0.222	3.43
age2635	-0.051	-0.75	-0.087	-2.25	0.025	0.56	-0.204	-2.02
age3645	-0.187	-2.44	-0.151	-3.38	0.046	0.94	-0.286	-2.65
age4655	-0.157	-1.98	-0.137	-2.98	0.019	0.37	-0.270	-2.39
age5665	-0.241	-2.72	-0.147	-2.76	-0.092	-1.60	-0.155	-1.23
age66plus	-0.441	-3.68	-0.253	-3.44	-0.167	-2.30	-0.220	-1.15
male	-0.245	-6.01	-0.456	-20.77	-0.470	-18.39	-0.237	-4.66
sahgood	0.327	8.50	0.308	11.17	0.382	4.80	0.239	4.30
sahfair	0.944	15.70	0.596	18.44	0.746	9.04	0.662	9.13
sahbad	1,267	9.58	1,286	20.45	1,012	10.93	0.872	5.35
sahvbad	1,309	5.19	1,251	6.65	1,231	9.98	NA	NA
hampsome	0.616	9.77	0.508	9.93	0.417	9.51	0.523	6.88
hampsev	0.352	2.95	0.466	5.12	0.467	7.57	1,110	6.61
selfemployed	-0.021	-0.32	-0.054	-1.37	-0.028	-0.66	-0.165	-1.61
student	-0.092	-0.84	0.082	1.79	0.022	0.36	0.171	0.89
unemployed	0.017	0.22	-0.032	-0.78	0.054	1.20	0.061	0.47
retired	0.313	3.17	0.097	2.16	0.161	2.93	-0.016	-0.12
housework	-0.052	-0.94	-0.002	-0.05	-0.016	-0.36	0.005	0.06
separated	-0.136	-0.97	-0.140	-1.54	-0.101	-0.95	-0.449	-1.86
divorced	-0.642	-1.01	0.083	0.71	-0.037	-0.42	0.072	0.75
widowed	-0.310	-2.98	-0.133	-2.11	-0.228	-4.13	-0.448	-2.57
never married	-0.220	-3.91	-0.318	-9.84	-0.217	-5.62	-0.200	-2.55
ln(income)	0.133	4.22	0.176	10.76	0.288	14.62	0.189	4.36
secondary educ	0.034	0.84	0.173	7.82	0.238	6.04	0.053	0.94
tertiary educ	0.117	2.09	0.213	5.46	0.409	6.83	0.156	2.36
_cons	-2,053	-7.91	-1,992	-13.03	-4.598	-16.93	-2,207	-6.02

Table 7.A.5 FIML regression coefficients

	Ireland		Italy		Portugal		UK	
	coeff	t-value	coeff	t-value	coeff	t-value	coeff	t-value
specialist								
private insurance	0.294	2.18	0.561	1.76	0.826	2.97	0.303	1.68
age2635	-0.049	-0.67	-0.087	-2.09	0.034	0.71	-0.204	-2.00
age3645	-0.197	-2.34	-0.154	-3.18	0.047	0.90	-0.290	-2.65
age4655	-0.169	-1.95	-0.137	-2.74	0.032	0.58	-0.274	-2.40
age5665	-0.248	-2.61	-0.147	-2.56	-0.073	-1.16	-0.157	-1.24
age66plus	-0.442	-3.19	-0.252	-3.05	-0.145	-1.77	-0.222	-1.10
male	-0.242	-5.44	-0.459	-19.00	-0.476	-16.68	-0.238	-4.65
sahgood	0.328	8.39	0.308	10.90	0.380	4.86	0.239	4.31
sahfair	0.946	14.79	0.595	17.50	0.743	9.06	0.663	9.16
sahbad	1,269	9.38	1,286	19.60	1,006	10.71	0.874	5.40
sahvbad	1,311	5.39	1,249	6.61	1,227	9.46		
hampsome	0.617	8.99	0.507	9.41	0.411	8.69	0.523	6.79
hampsev	0.353	2.80	0.465	4.86	0.462	6.98	1,113	6.84
selfemployed	-0.024	-0.34	-0.073	-1.60	-0.023	-0.50	-0.163	-1.60
student	-0.110	-0.96	0.081	1.65	0.021	0.33	0.167	0.87
unemployed	0.020	0.26	-0.033	-0.77	0.057	1.17	0.061	0.50
retired	0.310	2.73	0.099	2.02	0.167	2.71	-0.015	-0.11
housework	-0.050	-0.83	-0.002	-0.05	-0.011	-0.21	0.002	0.03
separated	-0.119	-0.76	-0.139	-1.51	-0.098	-0.81	-0.442	-1.89
divorced	-0.623	-1.34	0.077	0.57	-0.031	-0.30	0.081	0.86
widowed	-0.297	-2.28	-0.132	-1.81	-0.225	-3.54	-0.438	-2.33
never married	-0.207	-3.18	-0.313	-8.80	-0.201	-4.84	-0.194	-2.41
ln(income)	0.117	2.78	0.170	9.01	0.272	11.53	0.179	3.72
secondary educ	0.019	0.37	0.168	6.69	0.228	5.14	0.049	0.85
tertiary educ	0.089	1.17	0.203	4.51	0.336	4.70	0.148	2.17
_cons	-1,946	-5.98	-1,938	-11.14	-4,410	-13.89	-2,134	-5.31

Table 7.A.5 (continued) FIML regression coefficients

	Ireland		Italy		Portugal		UK	
	coeff	t-value	coeff	t-value	coeff	t-value	coeff	t-value
private insurance								
age2635	-0.111	-1.27	0.150	1.19	-0.253	-2.89	0.223	1.48
age3645	0.454	4.59	0.299	2.20	-0.134	-1.45	0.498	3.27
age4655	0.624	6.09	0.173	1.23	-0.330	-3.11	0.506	3.20
age5665	0.429	3.83	0.217	1.39	-0.436	-3.64	0.518	2.98
age66plus	0.150	1.01	0.171	0.83	-0.723	-3.40	0.592	2.25
male	-0.106	-2.07	0.165	2.86	0.153	2.83	0.026	0.42
sahgoodL	-0.057	-1.43	-0.097	-1.52	0.147	1.41	0.001	0.02
sahfairL	-0.149	-2.04	-0.044	-0.55	0.141	1.20	-0.208	-2.05
sahbadL	-0.354	-2.12	-0.458	-2.31	0.003	0.02	-0.423	-1.57
sahvbadL	-0.370	-1.21	-0.099	-0.29	-0.288	-0.82		
hampsomeL	-0.009	-0.11	0.203	1.76	0.235	2.21	0.003	0.03
hampsevL	0.015	0.09	0.112	0.49	0.285	1.90	-0.182	-0.70
selfemployedL	0.276	3.43	0.796	11.47	-0.177	-1.66	0.101	0.80
studentL	1,160	10.44	0.004	0.02	0.002	0.01	0.464	1.88
unemployedL	-0.587	-5.08	-0.183	-1.33	-0.294	-2.49	-0.465	-1.37
retiredL	0.145	1.19	-0.202	-1.57	-0.280	-1.65	0.169	0.86
houseworkL	-0.070	-1.06	-0.124	-1.17	-0.440	-2.80	0.175	1.68
separatedL	-0.858	-5.29	-0.137	-0.70	0.018	0.08	-0.554	-1.93
divorcedL	-0.728	-1.51	0.210	0.94	-0.005	-0.03	-0.588	-4.70
widowedL	-0.613	-4.42	-0.028	-0.14	-0.275	-1.47	-0.865	-2.80
never marriedL	-0.588	-8.17	-0.180	-2.30	-0.339	-4.34	-0.295	-3.07
ln(income)L	0.842	20.38	0.464	9.28	0.336	7.85	0.561	10.43
secondary educL	0.629	12.80	0.310	5.23	0.090	1.19	0.248	3.34
tertiary educL	1,075	14.99	0.367	4.12	0.446	4.85	0.385	4.80
employer benefit	0.437	9.62	0.320	6.54	0.594	11.12	0.945	16.08
_cons	-7,571	-21.85	-6,674	-14.11	-6,292	-11.20	-6,760	-14.60
	rho (e1,e2)	Prob > chi2	rho (e1,e2)	Prob > chi2	rho (e1,e2)	Prob > chi2	rho (e1,e2)	Prob > chi2
	-0.048	0.56	-0.144	0.33	-0.298	0.03	-0.052	0.63

Table 7.A.6 Private insurance probit regression coefficients

	Ireland		Italy		Portugal		UK	
	coeff	t-value	coeff	t-value	coeff	t-value	coeff	t-value
age2635	-0.123	-1.88	0.168	1.64	-0.212	-2.58	0.017	0.12
age3645	0.446	6.14	0.351	3.21	-0.070	-0.81	0.247	1.77
age4655	0.594	7.86	0.224	1.97	-0.304	-3.23	0.206	1.42
age5665	0.397	4.72	0.241	1.89	-0.466	-4.11	0.144	0.91
age66plus	0.082	0.70	0.165	0.93	-0.710	-3.75	0.194	0.80
male	-0.085	-2.16	0.170	3.66	0.192	4.03	0.054	0.92
sahgood	-0.107	-2.95	-0.078	-1.42	0.268	2.03	-0.011	-0.18
sahfair	-0.247	-3.90	-0.155	-2.30	0.409	2.95	-0.201	-2.21
sahbad	-0.513	-3.35	-0.324	-2.12	0.312	1.79	-0.281	-1.19
sahvbad	-0.309	-1.00	-0.059	-0.13	0.565	2.16		
hampsome	0.070	1.00	0.111	0.97	0.079	0.85	-0.018	-0.19
hampsev	0.176	1.29	0.146	0.69	0.075	0.49	-0.179	-0.78
selfemployed	0.100	1.69	0.691	12.04	-0.234	-2.70	-0.139	-1.23
student	1,162	12.49	0.070	0.56	0.071	0.67	-0.656	-1.50
unemployed	-0.555	-6.06	-0.110	-0.88	-0.434	-3.57	-0.416	-2.08
retired	0.091	0.91	-0.051	-0.50	-0.366	-2.54	-0.133	-0.84
housework	-0.088	-1.63	-0.079	-0.89	-0.372	-2.91	0.112	1.09
separated	-0.934	-6.39	-0.193	-1.13	0.010	0.05	-0.892	-2.47
divorced	-1,873	-2.62	0.017	0.09	-0.055	-0.37	-0.557	-4.55
widowed	-0.598	-5.83	-0.154	-0.95	-0.450	-2.21	-0.997	-3.42
never married	-0.612	-11.47	-0.204	-3.15	-0.293	-4.11	-0.411	-4.50
ln(income)	0.831	26.47	0.501	13.11	0.410	10.52	0.658	12.80
secondary educ	0.596	15.85	0.307	6.29	0.110	1.67	0.287	4.13
tertiary educ	1,047	19.50	0.346	5.09	0.529	6.66	0.346	4.61
_cons	-7,362	-27.96	-6,970	-18.96	-7,348	-13.60	-6,881	-15.29

Blundell, R.W. and R.J. Smith (1993), "Simultaneous microeconometric models with censored or qualitative dependent variables", In Maddala, G.S., C.R. Rao and H.D. Vinod, eds., Handbook of Statistics, Vol. 11. (Elsevier, Amsterdam) 117-143.

Bound, J., D. Jaeger, and R. Baker (1995), "Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak", Journal of the American Statistical Association 90: 443-450.

Buchmueller, T.C., A. Couffinhal, M.Grignon and M. Perronnin (2004), "Access to physician services: does supplemental insurance matter? Evidence for France", Health Economics 13: 669-687.

Cameron, A.C., P.K. Trivedi, F. Milne and J.Piggott (1988), "A microeconometric model of demand for health care and health insurance in Australia", Review of Economic Studies 55: 85-106.

Cameron, A.C. and P.K. Trivedi (1991), "The role of income and health risk in the choice of health insurance", Journal of Public Economics 45: 1-28.

Chiappori, P-A. (2000) "Econometric models of insurance under asymmetric information" in Handbook of Insurance, G.Dionne (ed.), Kluwer Academic,:Boston.

Chiappori P-A., F. Durand and P-Y. Geoffard (1998), "Moral hazard and the demand for physician services: first lessons from a French natural experiment", European Economic Review 42: 499-511.

Colombo, F. and N. Tapay (2004). Private health insurance in OECD countries: the benefits and costs for individuals and health systems. In: OECD. Towards high-performing health systems: policy studies, chapter 6, OECD, Paris.

Coulson, N.E., J.V. Terza, C.A. Neslusan and B.C. Stuart (1995), "Estimating the moral-hazard effect of supplemental medical insurance in the demand for prescription drugs by the elderly", AEA Papers and Proceedings 85: 122-126.

Deb, P. and P.K. Trivedi (1997), "Demand for medical care by the elderly: a finite mixture approach", Journal of Applied Econometrics 12: 313-336.

Ettner, S.L. (1997), "Adverse selection and the purchase of Medigap insurance by the elderly", Journal of Health Economics 16: 543-562.

Gardiol, L., P-Y. Geoffard and C. Grandchamp (2005). Separating selection and incentive effects in insurance. Paris-Jourdan Sciences Economiques, Working paper n.2005-38.

Geoffard, P-Y. (2006), "Incentive and selection effects in health insurance", in Jones, A.M. (ed.), Elgar Companion to Health Economics. Edward Elgar, Cheltenham.

Gerdtham, U-G. (1997), "Equity in health care utilization: further tests based on hurdle models and Swedish micro data", Health Economics 6: 303-319.

Greene, W.H. (2000), Econometric Analysis, Prentice Hall: New Jersey.

Hakkinen, U., G. Rosenquist and S. Aro (1996), "Economic depression and the use of physician services in Finland", Health Economics 5: 421-434.

Harmon, C. and B. Nolan (2001), "Health insurance and health services utilization in Ireland", Health Economics 10: 135-145.

Heckman, J.J. (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. The Annals of Economic and Social Measurement, 5, 475-492.

Heckman, J.J., R.J. LaLonde and J.A. Smith (1999), "The economics and econometrics of active labor market programs. In Handbook of Labour Economics, O.Ashenfelter and D.Card (eds.) vol. IIIA, Elsevier: Amsterdam.

Holly, A., L. Garidol, G. Domenighetti and B. Bisig (1998), "An econometric model of health care utilization and health insurance in Switzerland", European Economic Review 42: 513-522.

Hurd, M.D. and K. McGarry (1997), "Medical insurance and the use of health care services by the elderly", Journal of Health Economics 16: 129-154.

Jofre-Bonet, M. (2000), "Public health care and private insurance demand: the waiting time as a link", Health Care management Science 3: 51-71.

Maddala, G.S. (1983), Limited-dependent and qualitative variables in econometrics, Cambridge University Press: Cambridge.

Manning, W.G., J.P. Newhouse, N. Duan, E.B. Keeler, A. Leibowitz and M.S. Marquis (1987), "Health insurance and the demand for medical care: evidence from a randomized experiment", American Economic Review 77: 251-277.

Meza, D. (1983), "Health insurance and the demand for health care", Journal of Health Economics 2: 47-54.

Mossialos, E. and S.M.S. Thomson (2002), "Voluntary health insurance in the European Union: a critical assessment", International Journal of Health Services 32: 19-88.

Nyman, J.A. (1999a), "The economics of moral hazard revisited", Journal of Health Economics 18: 811-824.

Nyman, J.A. (1999b), "The value of health insurance: the access motive", Journal of Health Economics 18: 141-152.

Nyman, J.A. and R. Maude-Griffin (2001), "The welfare economics of moral hazard", International Journal of Health Care Finance and Economics 1: 23-42.

Pauly, M.V. (1968), "The economics of moral hazard: comment", American Economic Review 58: 531-537.

Peracchi, F. (2002), "The European Community Household Panel: a review", Empirical Economics 27: 63-90.

Pohlmeier, W. and V. Ulrich (1995), "An econometric model of the two-part decisionmaking process in the demand for health care", Journal of Human Resources 30: 339-360.

Propper, C. (1989), "An econometric analysis of the demand for private insurance in England and Wales", Applied Economics 21: 777-792.

Rhine, S.L.W. and Y.C. Ng (1998), "The effect of employment status on private health insurance coverage: 1977 and 1987", Health Economics 7: 63-79.

Rodríguez, M. and A. Stoyonova (2004), "The effect of private insurance access on the choice of GP/specialist and public/private provider in Spain", Health Economics 13: 689-703.

Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects", Biometrika 70: 41-55.

Schellhorn, M. (2001), "The effect of variable health insurance deductibles on the demand for physician visits", Health Economics 10: 441-456

Shmueli, A. (2001), "The effect of health on acute care supplemental insurance ownership: an empirical analysis", Health Economics 10: 341-350.

Terza, J.V. (1998), "Estimating count data models with endogenous switching: sample selection and endogenous treatment effects", Journal of Econometrics 84: 93-127.

Van de Ven, W.P.M.M. (1987), "An econometric model for the simultaneous estimation of the demand for medical care and the demand for health insurance", Economics Letters 24: 299-303.

Van Doorslaer E., X. Koolman and F. Puffer (2002), "Equity in the use of physician visits in OECD countries: has equal treatment for equal need been achieved? In Measuring up: improving health systems performance in OECD countries. Chapter 11. OECD: Paris.

Van Doorslaer E., X. Koolman and A.M. Jones (2004), "Explaining income-related inequalities in doctor utilisation in Europe", Health Economics 13: 629-647.

Van Doorslaer, E., C. Masseria and the OECD Health Equity Research Group. Income-related inequality in the use of medical care in 21 OECD countries. In: OECD (2004). Towards high-performing health systems: policy studies, chapter 3, OECD, Paris.

Vera-Hernández A.M. (1999), "Duplicate coverage and the demand for health care: the case of Catalonia", Health Economics 8: 579-598.

Vera-Hernández A.M. (2003), "Structural estimation of a principal-agent model: moral hazard in medical insurance", RAND Journal of Economics 34: 670-693.

Wilde,J. (2000), "Identification of multiple equation probit models with endogenous dummy variables", Economic Letters 69: 309-312.

Windmeijer, F.A.G. and J.M.C. Santos Silva (1997), "Endogeneity in count data models; an application to demand for health care", Journal of Applied Econometrics 12: 281-294.

Wagstaff, A. and E. van Doorslaer (2000), "Equity in health care finance and delivery" in Culyer A.J. and J.P. Newhouse (eds.) Handbook of Health Economics Volume 1b, Elsevier: Amsterdam.

Wagstaff, A, E. van Doorslaer, and N. Watanabe (2003), "On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam", Journal of Econometrics,. 112 (1): 207-223

Waters, H.R. (1999), "Measuring the impact of health insurance with a correction for selection bias – a case study of Ecuador", Health Economics 8: 473-483.

8 CHAPTER

Discussion

8.1 INTRODUCTION

Health and health care inequities have been studied for many years. While many methodological advances have been made in the past, state-of-the-art advances may still be incapable of accurately capturing societal notions of justice or societal preferences. That is because societal notions of justice or societal preferences are not clear or even stable across a span of countries, or within countries (Gakidou, Murray et al. 2003). As a consequence methods to measure inequalities require assumptions that may or may not be supported by society. In fact, as time progresses, researchers may develop a better understanding of the properties that underlie popular measures, and thus the assumptions implied by such measures (Wagstaff 2002; Bleichrodt and Van Doorslaer 2005 and chapter 2 and 3). This thesis builds upon the body of work relating to the measurement of income as related to health and also health care inequity and aims to extend the frontier of measurement methods.

This thesis also adds to earlier work on international variations in the size of inequalities in health and health care (Van Doorslaer and Wagstaff 1992; Mackenbach, Kunst et al. 1997; Van Doorslaer, Wagstaff et al. 1997; Cavelaars, Kunst et al. 1998; Van Doorslaer, Wagstaff et al. 2000; WHO 2000). These studies often used data from different sources and consequently suffered from lack of comparability. This study is based on the European Community Household Panel (ECHP), which was designed to maximize comparability.

It extends previous comparative studies by decomposing the variation as measured by the concentration index. The decompositions allow a better understanding of the underlying patterns of socio-economic inequality in health and health care. Special attention is given to health-related drop-outs that may affect the health income association and to private health care insurance as a cause of income-related inequity in health care.

8.2 MAIN CONTRIBUTIONS TO INTERPRETATION AND METHODS

8.2.1 A redistribution interpretation of the concentration index

Prior to this thesis several studies showed that complex measures of health and health care variation – such as the concentration index or the relative index of inequality - had attractive properties (Wagstaff, Paci et al. 1991; Wagstaff, Van Doorslaer et al. 1991; Mackenbach and Kunst 1997), but suffered from their complex interpretation (Mackenbach and Kunst 1997). For the unstandardised (standardised) concentration index it was known it was bounded between -1 and +1 (-2 and +2), but it remained difficult to interpret the extent of the inequality. The complex interpretation became more problematic when we computed inequality indices based on health rather than ill-

health and the standardised concentration index reduced almost tenfold. Even though statistical significance remained similarly strong, concentration indexes of about 0.003 (Van Doorslaer and Koolman 2000) raised questions about the ethical relevance of the observed variation.

Chapter two facilitates interpretation of the concept of a concentration index by adding a new redistribution interpretation. It reveals the percentage of the total amount of health or health care that needs redistributing from rich to poor (or vice versa) to achieve a concentration index equal to zero. The total percentage to redistribute is approximately 75 times the concentration index. In other words, given that the health concentration index for the Netherlands in 1996 was equal to 0.0034 (chapter four), it would have to redistribute 0.25 percent of health in order to obtain a concentration index equal to zero.

While the redistribution results in a concentration index that equals zero, the post redistribution concentration curve may not lie on the diagonal. It is possible that post redistribution income-related health inequality remains in some parts of the income distribution, which will always offset each other.

8.2.2 Relationship concentration index and relative index of inequality

The concentration index and Pamuk's relative index of inequality have been shown to be related through a multiplication by a factor with a probability limit of 1/6 for large samples, resulting in the approximate equality: concentration index ≈ 1/6 relative index of inequality (Wagstaff, Van Doorslaer et al. 1991). As a consequence they were assumed to share the same properties when based upon large samples. However, the current the relative index of inequality is often not based on the concept introduced by (Pamuk 1985). Chapter three shows that due to a redefinition of the relative index of inequality by (Mackenbach and Kunst 1997) - one that more resembled relative risk - the relationship between the two becomes much more complex. Comparability is hampered further by the common practice of basing the relative index of inequality on an odds-ratio when the outcome variable has a categorical nature (Huisman, Kunst et al. 2003).

8.2.3 Prevalence proportion ratio based relative index of inequality

Odds-ratios lack an intuitive interpretation when the prevalence of the health state under study is higher than a few percent. Chapter three proposes a novel method to compute the relative index of inequality on a prevalence proportion ratio rather than an odds-ratio, irrespective of the type of outcome variable and thus statistical model used. This new procedure allows researchers to compute the concepts suggested in demographic and epidemiologic research (Pamuk 1985; Mackenbach and Kunst 1997) rather than its (odds-ratio based) approximation. It shows that the relative index of

inequality based on a prevalence proportion ratio may lead to very different results from an approach based on odds-ratios. The new method results in a different ranking of the compared countries. The differences are especially great when prevalences are much higher than a few percent. This is frequently the case in health inequality research.

8.2.4 Statistical inference

Statistical inference on concentration indices was based often computed using the convenient linear regression suggested by (Wagstaff, Van Doorslaer et al. 1991). However, in practice the relationship between rank of income and health or health care may not be linear, and thus the expected value of the error term may depend upon the rank, causing a rank-related serial correlation. A correction for serial correlation was presented by (Kakwani, Wagstaff et al. 1997). Another cause for the biased estimation of the standard errors follows from the variation in both health and health care which is often related to rank of income, leading to heteroscedasticity. Chapter two suggests a solution to both serial correlation and heteroscedasticity in the form of Newey-West standard errors. However, this approach does not correct the inherent underestimation of the standard errors that was due to the use of predicted quantities to construct the dependent variable of the convenient regression equation. These predicted quantities suppressed the variability of the constructed dependent variable. The suppressed variability is particularly problematic when the convenient regression is used to construct standard errors around the contributions estimated in a decomposition exercise. Moreover, the Newey-West standard errors failed to correct the effect of popular complex sampling frames (such as stratified multi-stage sampling) on standard errors. We therefore use bootstrap techniques to compute the t-statistics (see chapter four and six). These resampling techniques allow for a simultaneous correction for serial correlation, heteroscedasticity, a complex sampling frame, and suppressed variability on dependent variable.

8.2.5 Decomposing income-related health inequalities

(Wagstaff, Van Doorslaer et al. 2003) created a useful tool when they introduced a decomposition of concentration index into the contribution of factors. The contributions allow for a further decomposition in health elasticities and concentration indices of the explanatory variables. Chapter applies of the decomposition technique in combination with an interval scale outcome variable, where the interval values are assumed to have ratio scale properties. It is also applies of the decomposition to explain differences in health inequality in European countries. This chapter also applies bootstrap techniques to compute the t-statistics for each of the contributions.

Results of chapter four are based on the third wave of a panel. Panels could suffer from informative non-response and attrition. In chapter five we evaluated whether non response was related to health using the full eight waves of the ECHP. Descriptive evidence shows that there is health-related non-response in the data, with those in very poor initial health more likely to drop out, and variable addition tests provide evidence of non-response bias in the panel data models of SAH. We both test and correct for non-response in empirical models with the impact of socioeconomic status on self-assessed health. Nevertheless a comparison of estimates - based on the balanced sample, the unbalanced sample and corrected for non-response using inverse probability weights – shows that, on the whole, there are no substantive differences in the average partial effects of the variables of interest. Similar findings have been reported concerning the limited influence of non-response bias in models of various labour market outcomes.

Chapter six is similar to chapter four in that it employs the decomposition method to the concentration index of health care inequality. However, because health care utilisation is modelled using non-linear model, and the decomposition is based upon the decomposition of a linear function of determinants of health care, we could not apply the standard decomposition technique. For that reason we present a new approach based on the average partial effects of all factors concerning the utilisation of health care estimated from non-linear models. These average partial effects depend on the distribution of all other factors in the model, which explains why they cannot be aggregated without creating an approximation error. This error depends on the extent of the non-linearity and thus the extent of interaction of the factors in predicting the outcome. Since the contributions are based on average partial effects the summation of contributions is also sensitive to approximation errors. The extent of these errors may render the summation – and, consequently, the decomposition – to be inadequate.

However, approximation errors do not affect individual factor contributions when all other factors are held constant. These contributions can be further decomposed into their average partial effect on health care, their distribution by income, and their average level. Furthermore, we decompose the overall inequity in physician visits into the inequity of general practitioner (GP) visits and the inequity in specialist visits. A further decomposition allows a look into inequality in the probability of visiting a specialist at least once, and the number of visits conditional on at least one visit. This decomposition provides new insights and is also used in a follow up paper by (Van Doorslaer, Masseria et al. 2006).

8.3 NEW CROSS-COUNTRY COMPARATIVE EVIDENCE IN HEALTH AND HEALTH CARE INEQUALITY

8.3.1 Income-related inequality in health

We present income-related differences in less than good self-assessed health for the elderly population from the European Union member states as measured by the relative index of inequality. These results are constructed to be comparable with (Huisman, Kunst et al. 2003). They are however not population-weighted and may therefore not be representative for the elderly populations. But more importantly they are measures of inequality in the less preferable measure of ill-health rather than good health (see chapter one). Therefore, these results are not compared with the results from chapter four.

Chapter four provides new evidence on the sources of differences in the degree of income-related inequalities in self-assessed health in 13 European Union member states. We use new and more comparable data from the 1996 wave of the European Community Household Panel (ECHP). Significant inequalities in health (utility) as measured by the concentration index favouring the higher income groups emerge in all countries, but are particularly high in Portugal and - to a lesser extent - in the UK and in Denmark. By contrast, relatively low health inequality is observed in the Netherlands and Germany, and also in Italy, Belgium, Spain Austria and Ireland. There is a positive correlation with income inequality per se but the relationship is weaker than in previous research. Health inequality is not merely a reflection of income inequality. A decomposition analysis shows that the (partial) income elasticities of the explanatory variables are generally more important than their unequal distribution by income in explaining the cross-country differences in income-related health inequality. This is especially true for the relative health and income position of non-working Europeans like the retired and disabled explains a great deal of 'excess inequality'. We also find a substantial contribution of regional health disparities in socioeconomic inequalities, primarily in the Southern European countries.

8.3.2 Income-related inequities in health care

While we present income-related variation in dental visits for the EU member states in chapter two, we cannot say much about income-related inequities because we were unable to measure, and therefore correct for, differences in need. And it is quite likely that need varied with income due to income-related differences in individual hygiene, eating and drinking habits, dental prosthesis etcetera. Moreover, measures of need may be hard to construct because, from our ethical perspective we are only interested in

need that may affect our concept of health, while much of the demand for dental care may result from esthetical considerations.

Chapter six presents new international comparative evidence on the factors driving inequities in the use of GP and specialist services as measured by the concentration index in 12 EU member states. New and more comparable data from the 1996 wave of the ECHP are used. We examine two types of utilisation (the probability of a visit and the conditional number of positive visits) for two types of medical care: general practitioner and medical specialist visits using probit, truncated negative binomial and generalised negative binomial regression models. We find little or no evidence of income-related inequity in the probability of a GP visit in these countries. Conditional upon at least one visit, there is even evidence of a somewhat pro-poor distribution. By contrast, substantial pro-rich inequity emerges in virtually every country with respect to the probability of contacting a medical specialist. Despite their lower needs for such care, wealthier and higher educated individuals appear to be more likely to see a specialist at least once in a year than the less well-off. This phenomenon is universal in Europe, but stronger in countries where either private insurance cover or private practice options are offered to purchase quicker and/or preferential access. Pro-rich inequity in subsequent visits adds to this access inequity but appears more related to regional disparities in utilisation than to other factors. Despite decades of universal and fairly comprehensive coverage in European countries, utilisation patterns suggest that rich and poor are not treated equally.

Both (Van Doorslaer, Koolman et al. 2002) and the above indicate that private health insurance coverage may well increase the health care inequity as measured by the standardised concentration index. However, both previous studies focus on a partial association rather than the causal effect of private insurance in physician visits. In chapter eight we try to retrieve the causal impact on the use of specialist visits in four selected European countries: Ireland, Italy, Portugal and the UK. The central questions are whether access to private insurance encourages greater utilisation and whether it contributes to horizontal inequity in the use of specialists. Empirical analysis of this issue is complicated by the fact that the decision to purchase voluntary health insurance is an individual choice that is likely to be influenced by risk selection. We compare different estimators to allow for this cause of residual confounding or endogeneity. The empirical results show that the probability of having private insurance increases with income and with better reported health. Private insurance has a positive association with the probability of specialist visits in all countries. The magnitude of the utilisation effect is sensitive to corrections for selection in two of the countries. These findings imply that private insurance contributes to 'pro-rich' horizontal inequity in the probability

of visiting a specialist, and may explain why countries with private insurance experience some of the highest inequity in at least one specialist visit a year.

8.4 ASSUMPTIONS, LIMITATIONS AND ALTERNATIVE APPROACHES

Below we discuss a small selection of issues that may compromise our results. First, we look into the requirements that concentration index demands from the data such as on measurement scales and absence of a measurement error. Then we focus on some crucial assumptions and alternative approaches chosen by others.

8.4.1 Measurement scales

The health concentration index can be written as (see chapter two):

$$C = 2\sigma_r^2 \frac{SII}{\overline{y}} \qquad (8.1)$$

where $2\sigma_r^2$ has a probability limit equal to 1/6, SII indicates the slope index of inequality, $\overline{y}$ indicates the mean of *y*, and *y* indicates health. This shows that the concentration index is a ratio measure – some refer to it as the concentration ratio (Rao 1969). Like other relative measures such as relative risk, risk ratio, rate ratio and odds-ratio, it therefore requires health to be measured on a ratio scale. Ratio scales have four properties:

1. Equality property: If we measure health with an interval scale, then if $y_a = y_b$ we can be sure that *a* and *b* have the same health.
2. Ordinality property: If we measure health with an interval scale, then if $y_a > y_b$ we can be sure that *a* has better health than *b*.
3. Interval ratio property. If $y_a - y_b = 0.1$ and the difference $y_b - y_c = 0.2$, then we know the difference in health between *b* and *c* is greater than the difference in health between *a* and *b*. In fact we know that the difference between *b* and *c* is twice as much as between *a* and *c*.
4. Value ratios property. If we measure health with a ratio scale, then if *a* has twice the level of health as *b*, then $y_a = 2y_b$ and vice versa, i.e., ratios of the measured values need to correspond to ratios of the actual measured quantities.

It is worth noting that cardinal scales at least meet properties 1-3 and that ratio scales allow only proportionality transformations ($f(y) = \theta y$ where θ is a constant).

In practice, health inequalities are often measured using lower quality measurement scales; i.e., scales that may not meet one or more of the properties of a ratio scale. Biases resulting from possible failure to meet requirements one and two are discussed under the section Reporting heterogeneity. When health is measured on a two point

ordinal scale (healthy/sick) researchers commonly measure inequality in terms of the probability of being one or the other. Probabilities require properties one and two to hold and meet properties three and four. When health is measured on a multi-point ordinal scale (e.g., Likert-scale), then one could (1) dichotomize the dependent variable and express inequality in terms of the probability to have one or the other, or (2) study the inequality for each cut-off point. Alternatively, one could (3) add information to the ordinal variable or use (strong) assumptions to convert it into a ratio scale.

The first option is straightforward but comes at the price of the loss of efficiency as information is discarded and, more importantly, results may be sensitive to the arbitrary choice of cut-point[a] (Wagstaff and Van Doorslaer 1994; Van Doorslaer and Koolman 2000). The second option will result in as many inequality indicators as cut-points and may have little statistical power for individual tests of inequality for cut points with few observations on one side.

The third option involves mapping the ordinal response categories onto a ratio scale (Wagstaff and Van Doorslaer 1994). This ratio scale is typically based on an assumed distribution of latent health, such as the log-normal distribution of ill-health, from which the ordered responses are drawn (Wagstaff and Van Doorslaer 1994), or on an empirical distribution of health that is measured on a ratio scale. For the latter scales concepts such as the Euroqol 5D (EQ5D) or the Health Utility Index (HUI) Mark III can be used (Van Doorslaer and Jones 2003), even though the latter is an interval rather than ratio scale (Horsman, Furlong et al. 2003).

It is easy to see how this could become problematic by considering the equation (8.1). Imagine that for some of the population average health equals zero, then the ratio cannot be calculated. If we take a less extreme example, we can see that when average health tends to zero, the concentration index will go to infinity. In other words if health is measured on an interval rather than a ratio scale then the concentration index is unbounded and thus looses its interpretation. No simple solution exists, because removing all negative values, which indicate a health state worse than death, would make the sample less the representative of the underlying population. Converting all negative values into zero would not do justice to negative health states. Absolute measures of inequality - such as the generalised concentration index and the slope index of inequality - merely require interval scales. Such scales have only the first three properties listed above, and therefore require no natural zero point.

8.4.2 Mapping ECHP SAH into Canadian Health Utility Index values

Because we lacked datasets for all member states of the European Union that allowed us to do the mapping exercise for all countries, we applied the cut points from Canada to all countries in our dataset. In doing so we assumed that the answer categories

of the Canadian self-assessed health question (bad, fair, good, very good, excellent) are equal to those in the ECHP (very bad, bad, fair, good, very good[b]) despite the difference in wording. We argued that this mapping may be acceptable because the relative frequencies of the European-wide 1996 ECHP responses are remarkably close to the relative frequencies for the response categories in the Canadian 1994 NPHS despite the different wording. By using the same cut-point values for all countries we assume that differences in responses to the self-assessed health question do reflect differences in health, contrary to earlier work by (Van Doorslaer, Wagstaff et al. 1997). (Lecluyse and Cleemput 2006) however have shown that cut-points may differ for the Belgium and may also affect the magnitude of the CI. Clearly, future research should aim to relax the assumption we had to impose.

As explained in chapter one and four, we use the Canadian Health Utility Index values to convert the ordinal measure of health into a health utility measure with interval scale properties, which we then use as if it had ratio scale properties. The internal validity of the mapping was put to the test by (Van Doorslaer and Jones 2003). We assume that (1) the approach that performed best in Canada would also be the best in other countries, (2) different wording of the self-assessed health question in the European panel from the Canadian survey give the answer categories the same meaning in terms of health utility and that thus Canadian cut-points are applicable to the European panel and (3) that a self-assessed health category means the same in all ECHP countries.

8.4.3 Informative measurement error (reporting heterogeneity)

Assumption 1 above has not been validated in other studies, but does not seem very problematic. The second and the third assumption, however, appear to be strong (Lecluyse and Cleemput 2006). Consider, also, the distribution of the self-assessed health of Denmark and Portugal. Table 8.1 presents the population weighted shares in the different SAH categories from the third wave of the ECHP of two extreme distributions: Denmark and Portugal.

From table 8.1, it seems that the Danish are in much better health than the Portuguese. Yet the Portuguese population was younger at the time of the survey and life expectancy for women was equal and for men only two years lower (Mackenbach, Bakker et al. 2002). The difference could be due to Portuguese suffering a much longer part of their life from adverse health conditions. It is possible that the differences in health are largely differences in the perception of health. In theory, the differences in perception could represent differences in health utility derived from a more similar objective health state such that the differences in the distribution of self-assessed health reflect the true difference in health-related utility. If the latter is not true then this reporting heterogeneity may have affected the results presented in chapter four.

Table 8.1 Shares in SAH categories		
	Denmark	Portugal
very bad health	1.7	4.7
bad health	4.6	17.6
fair health	17.0	29.6
good health	31.7	44.5
very good health	45.1	3.6

Reporting heterogeneity may be the most severe between countries - which is why we studied variation within countries rather than within Europe as a whole - but it may also exist within a country (Simon 2002; Van Doorslaer and Gerdtham 2003; Lindeboom and Van Doorslaer 2004; Simon, De Boer et al. 2005) and bias our estimates. Many current research projects using vignette evaluation focus on reporting heterogeneity. This type of research may be very helpful if the vignettes accurately describe all health states that are of moral interest and if the health states are appropriately valued for different groups in society.

8.4.4 Random measurement error in health or health care

The association between health and SES could be affected by measurement error. As is well known, a random measurement error in the dependent variable in OLS does not bias the coefficients but does increase the confidence intervals. In chapter two, we show that the concentration index can be computed using OLS, and thus that random measurement error in the dependent variable is unlikely to affect the concentration index. Non-random measurement error, e.g., reporting heterogeneity related to a socio-economic class may, however, well affect our estimates (Sen 2002).

8.4.5 Measurement error in explanatory variables

Measurement error in the explanatory variable of interest, fractional rank of income, could affect the results. Breen and Moisio show that measurement error in the measurement of income in ten ECHP country surveys is substantial and so are the differences in measurement error between countries (Breen and Moisio 2004).

Under the classical errors in variables (CEV) assumption [Cov(X,e)=0, where X the true value of the variable measured with measurement error and e is measurement error], measurement error leads to attenuation bias, which is equal to:

$$p\lim\hat{\beta} = \beta\left(\frac{\sigma_X^2}{\left(\sigma_X^2 + \sigma_e^2\right)}\right) \qquad (8.2)$$

where beta-hat is the estimated coefficient. From this it becomes clear that variance in measurement error of the magnitude observed by the authors could lead to (seriously) underestimated coefficients. Concentration indices may thus have been underestimated and relative positions of countries may have been affected.

8.4.6 Modelling drop out and conditional independence

In chapter five we try to allow for (health-related) drop-out by assuming selection on observables. Due to the problem of high dimensionality - finding identical individuals in terms of observables for those that drop-out in the data is impossible - techniques such as direct matching and direct standardisation are ruled out. Alternative approaches involve the modelling of the drop-out process and conditioning upon just one dimension: the probability of dropping out. Because of this approach, 'comparable' respondents will often have very different characteristics. This may even lead to an unbalanced distribution of characteristics at group level as we have shown using the propensity score approach in chapter seven. It remains possible that characteristics with equal power to predict drop-out, are related to very different associations between income and health. Unbalanced distribution of such characteristics could hamper the ability of the inverse probability weighted sample to represent the population.

8.4.7 Informative drop out

Our procedure to correct for drop-outs requires the assumption that respondents who dropped out from the sample, have equal health as those who remained in the sample with comparable characteristics up to the year before (selection on observables). This assumption seems reasonable when people with lower health are less likely to answer the questionnaire, but those who decide not to answer have not experienced a change in health like those who still answer the questionnaire. If, however, the individual does not answer the questionnaire due to a health shock that differs from comparable individuals who remain in the sample, such as death, then this assumption breaks down, and the inverse probability weighted sample fails to represent the population of interest.

8.4.8 Standardising by regression

Variation in health or in health care contains both ethically relevant variation and ethically irrelevant variation. Filtering out irrelevant variation is often referred to as standardising, and standardisation by regression is the most popular technique at present. Standardising by regression however only correctly removes the effect of the standardising factors - such as age and sex - if the effects of the standardising factors are correctly estimated by the regression model. In other words, if the estimated effects accurately capture the causal effect of the standardising factors on health or health

care. If this is not the case then residual confounding may cause other explanatory variables to be correlated with the error term, which would lead to biased estimates, and ultimately biased concentration indices.

Estimating causal effects from observational data is no easy task. Epidemiological research has produced many examples where efforts in causal modelling did not resemble the results from more reliable experiments. Many results from observational studies that were disproven by experiments seem to have suffered from unmeasured confounding (omitted variable bias). But bias also results from reverse causality, selection and information bias, measurement error and false statistical model assumptions.

By implication, knowledge about content matter (i.e. what is known about the causes of variation in health) is necessary to identify all confounders and is likely to be of crucial importance when constructing models. These causes could be economical, educational, societal, psychological, geographical, medical etc. This makes the building of structural models, or more specifically, arguing and defending model assumptions (identifying restrictions), an inherently interdisciplinary exercise. An exercise which is hampered by the way research is organised.

8.4.9 Decompositions

In this thesis we have applied two decompositions of the concentration index: Milanovich decomposition and the decomposition into contribution of factors. Strictly speaking the latter is not a true decomposition since to decompose means to separate something into its constituent parts or elements or into simpler compounds (Romo 2003). As such, a decomposition could be additive or multiplicative, but the parts, elements or compounds should ideally be independent. However, the components of the presented decompositions are often not completely independent, and thus may not qualify for decomposition in the strict sense of the term. For example, the components of the decomposition in factors are often not independent as the removal of the contribution of one factor could lead to a change in the health effect and the concentration index of other factors and a change in the average level of the dependent variable (which is part of each contribution). Similarly, other presented decompositions contain factors that, if manipulated, are likely to affect other factors. Other decompositions include residual contributions or a contribution due to approximation errors without an intuitive interpretation and of the which the magnitude may depend on the other contributions. As such, we use the term decomposition loosely.

8.4.10 Alternative factor decompositions with non-linear explanatory models

The decomposition of the concentration index in factors requires y (health or health care) to be a function that is additively separable into factors; (k), i.e., if for $i = 1,2,...n$ and

$j = 1,2,...m$, $y = f\left(k_i^1, k_i^2, ...k_i^m\right) + \varepsilon_i = g^1\left(k_i^1\right) + g^2\left(k_i^2\right) ... + g^m\left(k_i^m\right) + \varepsilon_i$. If y is not additively separable into all factors then the decomposition will not work out. In chapter six, we compute the contributions of all factors, while holding all other factors constant at their observed values. The sum of all contributions no longer equalled the concentration index. The differences - or approximation errors - are often substantial, indicating the limitations of decomposition methods. So, while adding up contributions may not always be meaningful, the contributions themselves show the importance of the factor while holding the other factors constant. From a policy making point of view, that is still very relevant information.

Morris et al. (2005) discuss two alternatives. They point out that the non-linear functions are often transformations of linear functions, e.g., $y = f\left(\dot{y}, \varepsilon_i\right) = f\left(k_i^1, k_i^2, ..k_i^m, \varepsilon_i\right)$, where $\dot{y}$ is referred to as a latent variable and is a additively separable function of factors k. Indeed they show that a concentration index based on $\dot{y}$ can be decomposed without approximation errors. However, such a decomposition is meaningless because (1) latent variables have no ratio scale properties (see above), and consequently no boundaries or interpretation and (2) magnitude or ranking of the contributions retrieved this way may be unrelated to the 'true' but unattainable contributions.

As an alternative, they discuss the use of linear models in the context of a dichotomous y variable. They mention two drawbacks of the linear probability model: (1) statistical inference will be sensitive to the heteroscedasticity implied by the model and (2) the model may produce impossible estimated probabilities. The first is not an issue if one uses the bootstrap methods proposed in chapter six. The second is not problematic if the model includes one series of dummies only. In practice models often include both continuous variables and an independent series of dummies from several categorical variables. However, if no causality is required then one could adjust the model such that it contains just one set of dummies. Equally, for non dichotomous variables, one could opt for a linear model that may not describe the data just as well, as long as no causality is implied. In chapter six, however, we aimed to standardise for the standardising factors, which implies that we needed to estimate effects that resemble causal effects.

8.3.11 Short-run versus long-run socioeconomic inequality in health

If the ultimate aim is to measure inequity in health-related well-being over the life span, then measuring variation in cross sectional health utility, health utility at one point in time, may not be a good approximation. For example, the may not be any association between health and income at any point in time among the living, yet life expectancy could still be strongly related to income. Alternatively, reductions in health could correlate with reductions in income and consequently a cross section inequity measure displays pro rich health inequity, yet each individual could still have an equal

quality adjusted life and income over the a lifetime. (Gerdtham and Johannesson 2000) have shown that concentration indices for income-related inequality in remaining life-years is indeed of a different and greater magnitude. Recent work by Andrew Jones and Angel Lopez Nicolas - who also use the European Community Household Panel - to compute concentration indices based on average health and average income is a step to bridge the gap between the cross-sectional inequity and longitudinal inequity (Jones and Lopez 2004).

8.5 CONCLUSION

This thesis adds methodological advances to the measurement and explanation of health and health care inequities. These and other recent advances are subsequently used to gather new cross-country comparative evidence on income-related inequality in health and equity in the use of health care. Some advances help to interpret the extent of the inequality as measured by the concentration index or relative index of inequality. Others help to clarify the difference between the concentration index and the relative index of inequality, both popular measures of socioeconomic differences in health. Methods are suggested that may help to make both measures more comparable.

Advances are made in the measurement of income-related health inequality. These measurements that now reflect inequalities in health rather than ill-health. Cross-country differences can be understood better by looking at the contributions of several policy relevant factors in a decomposition analysis. It is shown that cross-country differences between all countries and the country with the smallest income-related inequalities are mostly related by variation in health elasticities – the effect of a factor on health – rather than differences in the income distribution of these factors. For example the fact that countries perform poorly is due mostly to the retired people reporting worse health, rather than that the retired are more concentrated among the poor.

Similar advances are made in the files of inequities in health care consumption. Results allow a better insight into the consumption patterns of the different income groups. Evidence shows that some of these patterns – such as pro-rich inequity in at least one specialist visit a year - are robust to system characteristics, yet their extent sometimes depends on whether or not private insurance is endorsed by the system.

Many questions have, however, remained unanswered. Questions such as whether the implicit ethical consequences of the concentration index are endorsed by the societies that are evaluated, or whether self-assessed health is fully comparable within and between populations. And to what extent do different consumption patterns and possible health care substitutions indicate inequity? These are questions that will surely be answered in future research.

Notes

a To obtain more efficiency one could impose that the measure for inequality does not depend upon cut-point, for example by estimating an odds-ratio using an ordered logit. A test for irrelevance of the choice of cut point for odd-ratio's (proportional odds assumption) is offered by Brant (1990). "Assessing proportionality in the proportional odds model for ordinal logistic regression." Biometrics 46: 1171-1178.

b Except for the French questionnaire in which respondents were asked to indicate the satisfaction with their general health in six categories, ranging from very satisfied to very dissatisfied.

Bleichrodt, H. and E. Van Doorslaer (2005). A welfare economics foundation for health inequality measurement., Erasmus University.

Brant (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 46: 1171-1178.

Breen, R. and P. Moisio (2004). Poverty dynamics corrected for measurement error. Journal of Economic Inequality 2(3): 171-191.

Cavelaars, A.E.J.M., A.E. Kunst, et al. (1998). Differences in self-reported morbidity by educational level: A comparison of 11 western European countries. Journal of Epidemiology and Community Health(52): 219-227.

Gakidou, E.E., C.J.L. Murray, et al. (2003). Quality and Equity: Preferences for Health System Outcomes. Health Systems Performance Assessment. C.J.L. Murray and D.B. Evans. Geneva, World Health Organization.

Gerdtham, U. G. and M. Johannesson (2000). Income-related inequality in life-years and quality-adjusted life-years. Journal of Health Economics 19: 1007-1026.

Horsman, J., W. Furlong, et al. (2003). The Health Utility Index (HUI): concepts, measurement properties and applications. Health and Quality of Life Outcomes 1(54).

Huisman, M., A.E. Kunst, et al. (2003). Socioeconomic inequalities in morbidity among the elderly; a European overview. Social Science & Medicine 57(5): 861-873.

Jones, A. and A. Lopez (2004). Measurement and explanation of socioeconomic inequality in health with longitudinal data. Health Economics 13: 1015-1030.

Kakwani, N., A. Wagstaff, et al. (1997). Socioeconomic inequalities in health: Measurement, computation and statistical inference. Journal of Econometrics 77: 87-103.

Lecluyse, A. and I. Cleemput (2006). Making health continuous: implication of different methods on the measurement of inequality. Health Economics Forthcoming.

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

Mackenbach, J. and A. Kunst (1997). Measuring the magnitude of socio-economic inequalities in health: An overview of available measures illustrated with two examples from Europe. Social Science and Medicine 44(6): 757-771.

Mackenbach, J., A. Kunst, et al. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. Lancet 349(9066): 1655-1659.

Mackenbach, J.P., M.J. Bakker, et al. (2002). Socioeconomic inequalities in health in Europe; An overview. Reducing inequalities in health; A European perspective. J. P. Mackenbach and M. J. Bakker. London, Routledge.

Morris, S., M. Sutton, et al. (2005). Inequity and inequality in the use of health care in England: an empirical investigation. Social Science and Medicine 60(6): 1251-1266.

Pamuk, E. (1985). Social class inequality in mortality from 1921-1972 in England and Wales. Population Studies 39: 17-31.

Rao, V.M. (1969). Two decompositions of concentration ratio. Journal of Royal Statistical Society Series A 132(3): 418-425.

Romo, V.C. (2003). Decomposition methods in demography. Groningen, Rijksuniversiteit Groningen.

Sen, A.K. (2002). Health: perception versus observation. Britisch Medical Journal 324: 860-861.

Simon, J., J. De Boer, et al. (2005). How is your health in general? A qualitative study on self-assessed health. European Journal of Public Health 15(2): 200-208.

Simon, J.G. (2002). How is your health in general? Department of Public Health. Rotterdam, Erasmus University Rotterdam.

Van Doorslaer, E. and U.G. Gerdtham (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. Social Science and Medicine 57(9): 1621-1629.

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

Van Doorslaer, E. and X. Koolman (2000). Income-related inequalities in health in Europe: evidence from the European Community Household Panel. ECuity Project II Working Paper 1. Erasmus University, Rotterdam, ECuity Project II Working Paper.

Van Doorslaer, E., X. Koolman, et al. (2002). Equity in the use of physician visits in OECD countries: Has equal treatment for equal need been achieved? Measuring up: Improving health systems performance in OECD countries. Paris, OECD: 225-248.

Van Doorslaer, E., C. Masseria, et al. (2006). Inequalities in access to medical care by income in developed countries. Canadian Medical Association Journal Forthcoming.

Van Doorslaer, E. and A. Wagstaff (1992). Equity in the delivery of health care: Some international comparisons. Journal of Health Economics 11(4): 389-411.

Van Doorslaer, E., A. Wagstaff, et al. (1997). Income-related inequalities in health: Some international comparisons. Journal of Health Economics 16(1): 93-112.

Van Doorslaer, E., A. Wagstaff, et al. (2000). Equity in the delivery of health care in Europe and the US. Journal of Health Economics 19(5): 553-583.

Wagstaff, A. (2002). Inequality aversion, health inequalities and health achievement. Journal of Health Economics 21: 627-641.

Wagstaff, A., P. Paci, et al. (1991). On the measurement of inequalities in health. Social Science and Medicine 33(5): 545-557.

Wagstaff, A. and E. Van Doorslaer (1994). Measuring inequalities in health in the presence of multiple category morbidity indicators. Health Economics 3(4): 281-291.

Wagstaff, A., E. Van Doorslaer, et al. (1991). On the measurement of horizontal inequity in the delivery of health care. Journal of Health Economics 10: 169-205.

Wagstaff, A., E. Van Doorslaer, et al. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. Journal of Econometrics 112: 207-223.

WHO (2000). The World Health Report 2000. Health Systems: Improving Performance. Geneva, World Health Organization.

SAMENVATTING

Ongelijkheden. Soms zijn ze gewenst en leveren ze de broodnodige stimulans, soms zijn ze ongewenst en worden ze als onrechtvaardig ervaren. Uiteindelijk bepaalt de samenleving of ongelijkheden wenselijk of onwenselijk zijn. Vele samenlevingen vinden gezondheidsongelijkheden onwenselijk. Dat geldt echter niet voor alle oorzaken van gezondheidsongelijkheid. Indien ouderen ongezonder zijn dan jongeren, dan is dat misschien jammer maar volgens velen niet direct onrechtvaardig. Ook het beoefenen van risicovolle sporten, zoals skieen, kan leiden tot gezondheidsverschillen, zonder dat dit tot morele verontwaardiging zal leiden.

Dat is anders voor oorzaken die buiten de invloed van het individu tot ongezondheid leiden. Zo was de verontwaardiging groot toen defensie staatssecretaris B.J.M. baron Van Voorst tot Voorts wel kennis had van de aanwezigheid van asbest en haar gevaren in de Cannerberg (Maastricht) maar volgens TNO onvoldoende maatregelen nam om zijn personeel ter plaatse te beschermen. Sociaal-economische verschillen in gezondheid worden veelal ook als onwenselijk gezien zoals blijkt uit de uitspraak van de Engelse minister van volksgezondheid Frank Dobson (1997): “Ongelijkheid in gezondheid is de ergste ongelijkheid van alle ongelijkheden. Er is geen ernstiger ongelijkheid dan te weten dat je eerder zult sterven omdat je armer bent.”

Het meten van de ongelijkheden is een taak van onderzoekers. Een taak waar artsen, epidemiologen, demografen, sociologen, economen en andere onderzoekers zich al meer dan 150 jaar aan wijden. Ondanks deze lange onderzoekstraditie is de vraag hoe ongelijkheden te meten en vervolgens te verklaren nog altijd actueel. Dit proefschrift is geschreven om bij te dragen aan de methoden van ongelijkheidsonderzoek. De bijdragen zijn ieder voorzien van een toepassing en worden hieronder genoemd en toegelicht.

Ongelijkheden worden vaak in relatieve termen uitgedrukt, bijvoorbeeld rijken hebben een 10% hogere levensverwachting dan armen. Hierbij kunnen onderzoekers kiezen voor een positieve gezondheidsmaat (levensverwachting) of een negatieve gezondheidsmaat (kans op sterfte). In dit proefschrift wordt gesteld dat een positieve gezondheidsmaat te preferen is boven een negatieve indien men geinteresseerd is in de ongelijkheid in (voor kwaliteit gecorrigeerde) levensverwachting. Een negatieve gezondheidsmaat verreist een ijkpunt, bijvoorbeeld 80 jaar in volle gezondheid, en die standaard is arbitrair. De keuze van dit ijkpunt is evenwel zeer bepalend voor de gemeten omvang van de ongelijkheid, rangorde tussen landen en trend over tijd (positief of negatief). In hoofdstuk 1 blijkt dat indices van ongelijkheid gebaseerd op negatieve gezondheidsmaten niet met indices gebaseerd op positieve maten overeen hoeven te komen. De resultaten kunnen elkaar zelfs tegenspreken. Hierdoor kan vergelijkend onderzoek naar de ongelijkheid

in bijvoorbeeld sterfte volkomen misleidend zijn als benadering van de ongelijkheid in levensverwachting.

Een van de meest gebuikte maten van ongelijkheid is de concentratie-index. Onder gebruikelijke voorwaarden kan deze index varieren tussen -2 en +2, maar de waarden uit onderzoek zijn soms niet groter dan 0,0034. Dergelijke waarden zijn statitisch significant, maar roepen de vraag op of ze wel relevant zijn. Omdat de concentratie-index geen eenvoudige interpretatie kende, is hoofdstuk 2 in dit proefschrift gewijd aan de interpretatie van de concentratie-index. Daaruit blijkt dat aan de concentratie-index een herverdelingsinterpretatie kan worden gegeven. Bij benadering geldt dat indien 75 maal de waarde van de concentratie-index als percentage wordt herverdeeld volgens een bepaalde sleutel, dat de concentratie-index gelijk wordt aan nul. Met andere woorden, als de concentratie-index gelijk is aan 0,0034 dan kan de ongelijkheid worden opgeheven door 0,25% van alle gezondheid opnieuw te verdelen. Hierdoor is het mogelijk geworden om te beoordelen welke waarde van de concentratie-index aanleiding geeft tot zorgen of niet.

Een andere veel gebruikte maat van ongelijkheid is de odds-ratio. De odds-ratio is te interpreteren als een relatief risico, bijvoorbeeld de rijken hebben een 10% hogere kans om goede gezondheid te rapporteren dan de armen. Deze interpretatie is enkel geldig indien de kans op de uitkomst goede gezondheid lager is dan 5%. In gezondheidsongelijkheid onderzoek is die kans veelal groter dan 5%, ongeacht of positieve of negatieve gezondheidsmaten worden gebruikt. En, indien die kans groter is, dan zal de odds-ratio het relatief risico altijd overschatten maar de mate waarin varieert. Zo kan het voorkomen dat de ongelijkheids-rangorde van landen en trends binnen een land over tijd, sterk afhankelijk zijn van de gekozen maat. Ondanks de interpretatieproblemen wordt de odds-ratio vaak gebruikt omdat onderzoekers geen relatief risico kunnen berekenen indien gezondheid dichotoom (0-1) is gemeten. In hoofdstuk 3 wordt daarom een methode gepresenteerd om op basis van dichotome gezondheidsmaten toch een relatief risico uit te rekenen.

Eenmaal gemeten volgt de vraag wat de oorzaken van de waargenomen ongelijkheid zijn. Causale relaties kunnen veelal het best onderzocht worden door een experiment op te zetten. Echter, experimenten zijn voor deze vraagstelling in de praktijk ondoenlijk. Wel kunnen we mensen ‘observeren’ met behulp van vragenlijsten en bestuderen of factoren met elkaar samenhangen. Samenhang kan oorzakelijk zijn, maar kan ook het resultaat zijn van een niet-representieve steekproef of door andere vertekening. Kortom, de oorzaken van ongelijkheid uiteenrafelen is nog niet mogelijk. Wel kunnen we met behulp van zogenaamde decompositietechnieken in kaart brengen in welke mate bevolkingsgroepen bijdragen aan ongelijkheid. Verder kunnen we binnen een groep, bijvoorbeeld werklozen, uitsplitsen of de bijdrage het gevolg is van een slechte

gezondheid of juist een lage inkomenspositie. Vergelijkend onderzoek kan uitwijzen of de bijdragen van specifieke groepen in het ene land groter is dan in een ander land, en waar dit mee samenhangt. De resultaten van vergelijkend onderzoek voor 13 Europese landen op basis van de derde jaar van het Europese Unie Huishoud Panel (European Community Household Panel) staan gepresenteerd in hoofdstuk 4.

De resultaten in hoofdstuk 4 zijn verminderd betrouwbaar indien de steekproef in de derde jaar geen goede weerspiegeling is van de totale bevolking boven de 16 jaar. Het is bijvoorbeeld mogelijk dat mensen met gezondheidsproblemen de vragenlijst in het derde jaar minder vaak hebben ingevuld, waardoor een te positief beeld ontstaat over de gezondheid van de bevolking. In hoofdstuk 5 laten wij zien dat mensen met gezondheidsklachten inderdaad minder geneigd zijn om deel te nemen aan de enquête. We gebruiken inverse probability weighting om te corrigeren voor deze selectieve uitval. De correctie laat zien dat, door de bank genomen, de selectieve uitval de relatie tussen inkomen en opleiding aan de ene kant en gezondheid aan de andere kant, niet wezenlijk beïnvloedt.

Gezondheidsverschillen zouden ook het gevolg kunnen zijn van ongelijkheid in de toegang tot zorg. Zo zouden rijkeren, bij een vergelijkbare gezondheid, gemakkelijker toegang kunnen hebben tot de gezondheidszorg dan armeren. Vele Europese overheden stellen zich daarom expliciet tot doel hebben om de toegang tot zorg 'rechtvaardig' te laten zijn. Daarmee wordt bedoeld dat het zorggebruik dient af te hangen van de gezondheidssituatie en niet van andere factoren zoals de inkomenssituatie. In hoofdstuk 6 blijkt in welke mate overheden geslaagd zijn in dit doel. De contacten met de huisarts blijken, grosso modo, niet met inkomen samen te hangen of zelfs geconcentreerd te zijn bij de lagere inkomens. Toegang tot de specialist lijkt echter wel positief samen te hangen met inkomen. En deze ongelijkheid blijkt vooral te bestaan rondom het eerste contact. De ongelijkheid in het aantal vervolgcontacten is minder groot.

Uit hoofdstuk 6 komt ook naar voren dat landen waar het mogelijk is om een private gezondheidszorgverzekering af te sluiten, veelal de landen zijn waar de toegang tot specialisten ongelijker is verdeeld. In een poging om met behulp van vragenlijsten toch iets te zeggen over het causale verzekeringseffect, is in hoofdstuk 7 gebruik gemaakt van een combinatie van identificatietechnieken waaronder de instrumentele variabele techniek, recursieve modellering en de veronderstelling van een specifieke niet-lineaire relatie tussen factoren. De resultaten van deze modellering suggereren dat het causale verzekeringseffect op zorggebruik groter is dan de associatie die is gemeten in hoofdstuk 6. De verschillen zijn zeer relevant, maar in dit geval niet statistisch significant. Het toelaten van private verzekeringen lijkt de inkomensgerelateerde ongelijkheid in toegang tot de zorg te vergroten.

ACKNOWLEDGEMENTS / DANKWOORD

I would like to thank many people for making this thesis possible. First, I am indebted to the European Union for providing the funds that allowed me to work on this great topic, and to Adam Wagstaff, Eddy van Doorslaer, and Andrew Jones for their efforts to acquire these funds. I was fortunate to work with Eddy van Doorslaer, with whom I have had many long and often pleasant discussions about the measurement of inequality, and Andrew Jones, whose extensive knowledge about health econometrics is unsurpassed. Eddy also indefatigably pointed at the space between the goal posts at times when I was - once again - looking another way. Andrew's impossible-to-disturb willingness to give any organic life form a second chance has been used gratefully by 'this organic life form'.

The ECuity II and III team members and participants to the ECuity workshops have also encouraged me and given me great support in numerous ways. Many allowed me to discuss methodological issues (Matthew Sutton, Carol Propper, Brian Nolan, Hugh Gravelle, and others) and stimulated me to explore methodological questions (Thomas Buchmueller, Franco Peracci, Maarten Lindeboom, and others); while others rose important but sometimes ignored institutional issues (Terkel Christiansen, Vincenzo Attela, Angel Lopez, and others). I have always felt that I could have used their knowledge more than I did - a thing I hope to do in the near future.

Hindrik Vondeling, Evertjan Jansen and Erik Schut have always functioned as ethical beacons in the competitive world of research. Evertjan, most of all, succeeded many times in getting me to defend the indefendable. Anton Kunst has often stimulated me with his questions and views. I would also like to thank my office mates and direct colleagues: Michiel Hol for the very funny but often unproductive times, Piet Stam for teaching me everything I know about the functioning of Dutch health insurers, Piet Stam and Tom Van Ourti for the many econometric discussions, and Cristina Masseria and Cristina Hernandez Quevedo for their pleasant company. Peter Huberts and Elly Stolk provided great help with the lay-out of this thesis.

During periods with doubts, I have sought the advice of senior scientists. They have all been very willing to give me the fatherly advice I was looking for. It did not always silence my internal disputes, but was much appreciated. For this, I would like to express my gratitude to some people in particular: Angus Deaton for his refusal to accept my reasons to work mono-disciplinary in the field of health inequalities and for his views about the use of instrumental variables, James Robins for his views about social processes within the group of econometricians and about the different motivation of students in statistics (methods) and epidemiology (content) and Gerhard Pfann and Peter Smith for their opinions on the functioning of peer groups within a specific field of expertise. The conversations with these men have changed the way I look at scientific world to a great extent. When in methodological doubt, I could always turn to the excellent textbooks

of Jeffrey Wooldridge, J. Scott Long, Moyses Szklo and Javier Nieto, which helped me a great deal during the first years of working on this thesis.
The iBMG provided an environment far enough from the daily reality of health care provision to concentrate on more abstract issues, yet close enough not to forget the very reason why health care research is done in the first place. The iBMG students have certainly contributed to my concern for the real world. It is a pleasure to teach them.
Hoewel genoemd op het einde, is alles begonnen bij mijn ouders. Bij mijn vader omdat hij vindt dat een 'echte man' van twee kapotte motoren één werkende kan maken en oog heeft voor de praktische noden van anderen, en bij mijn moeder omdat zij vindt dat een 'echte man' voor zijn geliefde een jurk kan naaien en aandacht kan opbrengen voor de emotionele problemen van anderen.
Hoewel de motor en jurk zijn gelukt, zal het helpen en steunen van anderen zoals mijn ouders dat kunnen, voor altijd een doel blijven. In de afrondingsfase van dit proefschrift heb ik dankbaar gebruik gemaakt van de gastvrijheid en zolder van mijn schoonouders Sybrand en Bep. Tot slot wil ik mijn vrouw Dika bedanken voor de ruimte die ze me gaf om bij tijden helemaal op te gaan in het werk, ruimte waarvan ik me de prijs pas realiseerde toen zij aan haar boek werkte. Ik verheug me enorm op het volgende hoofdstuk in ons leven waaraan zij nu zo hard 'werkt'.

SPECIFIC ACKNOWLEDGEMENTS

CHAPTER 2

We are grateful to Andrew Jones and Tom Van Ourti for stimulating discussions from which the chapter benefited considerably and Evertjan Jansen and two anonymous referees for useful suggestions and comments on an earlier version of the chapter.

CHAPTER 3

We are grateful to Martijn Huisman, Dika Luijendijk, James Deddens, Eddy van Doorslaer and three anonymous referees for helpful comments.

CHAPTER 4

We are grateful to Andrew Jones and participants at ECuity Project workshops in Aix-en-Provence and Rome, and at seminars given at the universities of Oxford, York, Bergen, McGill, Montreal and the London School of Economics for comments on an earlier version of this chapter.

CHAPTER 5

We are grateful to the EU for financial support and to Hugh Gravelle, Tom van Ourti, and participants at the ECuity Project Workshops in Rome, 2002, and Belgirate, 2003, and two anonymous referees for helpful comments on earlier drafts.

CHAPTER 6

We are grateful to Maarten Lindeboom, John Wildman and participants at the RSS/ESRC Research Methods Programme conference on Statistical methods for attrition and non-response in social surveys, London May 2004, the ECuity III Project workshop, Helsinki June 2004, and the York Seminars in Health Econometrics (YSHE) for their comments on earlier versions of this paper.

CHAPTER 7

We are grateful to Jan Erik Askildsen, Tom Buchmueller, Angel López Nicolás, Brian Nolan, Owen O'Donnell, participants at ECuity Workshops in Rome and Helsinki, the Workshop on Health Insurance and Equality of Access in Leuven and the York Seminars in Health Econometrics (YSHE) and two anonymous referees for their comments on earlier versions.

CURRICULUM VITAE

Xander Koolman (1970) was born and raised in Eindhoven (the Netherlands). At the age of 17 he became politically engaged and joined the Jonge Democraten, a political youth organisation related to Democrats 66 (a centre-left political party) for which he quickly became chairman of the regional department of Noord-Brabant. In 1989 he finished his preparatory university education (VWO) at the Eckart College in Eindhoven and started the study of Economics and Business Administration at the University of Maastricht, specialising in Economics. As a student assistant, he taught mathematics, statistics, and economics to national and international students. He extended his studies by studying Economics at the University College Cork (Ireland) for one year; he also attained a research traineeship and later a research position at the University Science Shop (Wetenschapswinkel). He finished his Master's Degree in Economics and Business Administration cum laude in 1997.

Following his studies, he worked as a researcher in the field of medical technology assessment at the Vrije Universiteit. From 1998 onwards he researched in the field of health inequality at the institute of Health Policy and Management (iBMG), Erasmus University Rotterdam. The institute joined the Erasmus University Medical Center in 2003. He has been affiliated from 2003 until 2005 with the York Seminars in Health Econometrics (YSHE), and since 2005 with the Health Econometrics and Data Group (HEDG), both from the Department of Economics and Related Studies, University of York. He referees papers for the British Medical Journal, Health Economics, Health Policy, Journal of Health Economics, Social Science and Medicine, and other related journals.

He remains actively involved in Democrats 66 and holds a position in the supervisory board of the health care foundation Boog, which employs about one hundred health care professionals in an ambulatory health care setting. Xander is married to Dika Luijendijk, and she is expecting their first child.