Health and wealth: empirical findings and political consequences*

PREPARED FOR PERSPEKTIVEN DER WIRTSCHAFTSPOLITIK
PRELIMINARY DRAFT AUGUST 2005

Andrew M. Jones¹, Eddy van Doorslaer², Teresa Bago d’Uva³, Silvia Balia¹, Lynn Gambin¹, Cristina Hernandez Quevedo¹, Xander Koolman² and Nigel Rice⁴

¹ Department of Economics and Related Studies, University of York, United Kingdom
² Institute of Health Policy and Management, Erasmus University Medical Center, The Netherlands
³ Centre for Health Economics and Department of Economics and Related Studies, University of York, United Kingdom
⁴ Centre for Health Economics, University of York, United Kingdom

* The European Community Household Panel Users' Database, version of December 2003, was supplied by Eurostat. This paper derives from the project “The dynamics of income, health and inequality over the lifecycle” (known as ECuity III Project), which is funded in part by the European Community’s Quality of Life and Management of Living Resources programme (contract QLK6-CT-2002-02297).
1. Policy issues and the “ECuity Project”

There is increasing concern that equity in health and health care in Europe may suffer as a result of the expansion of the European Union and the ageing of its populations. This is reflected, for instance, in the recent commitment at EU level of member states to set up national action plans to combat poverty and social exclusion. Earlier work from the ECuity network of researchers has had substantial impact on the health dimension of the indicators proposed in the Atkinson Report (Atkinson et al, 2001) which recently set out recommendations for the development of indicators of social inclusion in the European Union. These were developed in response to the 2000 Lisbon European Council meeting’s resolution to “promote a better understanding of social exclusion through continued dialogue and exchanges of information and best practice, on the basis of commonly agreed indicators”. The Social Protection Committee is now in charge of the implementation of measures to fight against poverty and social exclusion as approved in the European Councils.

Europe has a long history of developing social protection systems which are rooted in the egalitarian tradition. Health care is no exception and most European Union member states have long achieved fairly universal coverage for their populations for a rather comprehensive package of health care services. Recent trends of rapidly ageing populations and expanding technological opportunities have not only challenged the affordability of these systems but also their commitment to equitable access for all, including socially disadvantaged groups. While this challenge is not unique to European countries, what is unique in the European setting is that the process of economic and monetary unification puts some pressure on countries to harmonize their social policies. At the special European summit in Lisbon in March 2000, for the first time, social policy was explicitly introduced as a distinct focus of attention for European cooperation. It was agreed that common objectives for eradication of poverty and social exclusion would be adopted, that national policies would be designed to meet these, and that progress would be monitored. As a result of this trend towards European social policy harmonization, cross-country comparative information on social inequalities and exclusion (in terms of health or other dimensions) has gained additional relevance in Europe.

The “ECuity project” is shorthand for a series of research projects carried out over the past decade and a half by collaborating European health economics research groups, that have been funded by the European Union (see http://www2.eur.nl/bmg/ecuity/). The previous findings of the ECuity project suggest that (a) while all countries in Europe show some income-related inequalities in health, some are faring much better than others, and (b) while most European countries have secured fairly equal access to a GP, if they continue to expand the options to “go private” alongside their essentially public health care systems, they are likely to exacerbate the pro-rich utilization patterns in volumes
and mixes of specialty services which are already observed, even in countries where such private options are currently not (or hardly) available.

Going beyond the descriptive evidence by estimating effects using causal modelling, rather than merely teasing out associations, may become possible now that the full eight waves of the European Community Household Panel (ECHP) for fourteen EU member states are available. That is the ambition of the current phase of the project, the “ECuity III project”. In order to help inform the policy debate about how to secure health equity in our ageing European societies, it pays particular attention to the key decisions about income, health and health care in age groups around the retirement age, as these prove to be crucial for a better understanding of cross-country differences in inequalities.

2. Evidence from the literature

Appropriate specification of the theoretical relationships between income, health and the inequality in each of these two key variables is crucial for any empirical attempts to shed light on the relationship between income and health inequalities. Recently, Contoyannis and Forster (1999a,b) have provided a theoretical model that generates testable predictions about the conditions under which changes in income level and inequality will generate changes in the level and inequality of health. They highlight the crucial role of the elasticity of health with respect to income. However, it is well known that the relationship between income and health is non-linear and that the association may partly be due to reverse causality and joint determination by unobserved third factors. Reliable estimates of the income elasticity of health therefore have to take account of this non-linearity and the potential of reverse causation.

2.1 The role of income in explaining health inequalities

The influence of material factors has played a fundamental role in research on health inequalities. The relationship between health and income has been highlighted by McKeown (1979) and his work on historical epidemiology. McKeown noticed that large declines in mortality had occurred in Britain prior to the development of key medical interventions. McKeown’s thesis is a diagnosis of exclusion (Evans et al., 1994). As no theory gave a reasonable explanation of the data he assumed that the decline in death rates must be due to the increase in income which led to better living standards and nutrition. Regardless of the manner in which the theory was developed, Evans et al. (1994) believe that McKeown’s central point - that the major decline in morality from most

Wilkinson (1996) examines the historical link between income and health. The increasing and concave non-linear relationship revealed by the data seems to have shifted up over time, illustrating the strong correlation between average income and increasing health across all nations (see also, Preston, 1975). However he also found that the association between health and income at the aggregate level only applies up to around $5,000 per capita in 1990. There is a large amount of evidence relating health to income across countries. The work of Adelman, over 30 years ago, recognised this fact. Adelman (1963) documented that there exists a negative long-run association between death rates and economic conditions. Pritchett and Summers (1996) find an association between wealth and health. Wealthier nations are healthier nations. They find that increasing a country's income will increase its health and their instrumental variable approach suggests that the income-mortality relationship is not an artefact of reverse causation.

These studies are part of a debate over the association between health and socioeconomic status (SES): in particular health and education (see e.g., Grossman, 2000, Smith, 2004) and health and income or wealth (see e.g., Smith, 1999, 2004). Evidence of a positive association between health and SES is well-documented across many societies and periods (see e.g. Smith, 1999, Deaton, 2003). But the causal mechanisms underlying this relationship are complex and controversial. There can be a direct causal link from SES to health, for example, through the direct influence of material deprivation on the production of health and on access to health care, or of education on the uptake and compliance with medical treatments. There can be a direct causal link from health to SES, for example, through the impact of health shocks on labour market outcomes such as unemployment, early retirement and earnings. But there may also be pathways that link health and SES through “third factors”, for example time preference rates, that do not imply any causal link.

Fuchs (1974) found that there was no relationship between per capita income and age specific mortality rates across economically developed countries. Fuchs did find a relationship at the individual level between health and income, although he felt it was the result of reverse causation reflecting the fact that poor health created low income. Duleep (1995), using longitudinal data, found an individual-level non-linear relationship between income and health. She tries to control for potential reverse causation by controlling for health conditions that may affect income. She concludes that there is some evidence for reverse causation but still finds a significant causal link between income and health.
Ettner (1996), using data from the US, has studied the link between income and health at the individual level. She acknowledges that simple correlations between income and health are not conclusive evidence of a causal link between the two. Much of the work in the sociological and epidemiological fields has focused on either simple correlations or observed patterns controlling for known risk factors (Wilkinson, 1996, Marmot et al., 1978). As the evidence above suggests there is a relationship between health and labour force participation, with the ability to command higher wages, leading to the problems of reverse causality. Ettner models income as both exogenous and as simultaneously determined with health. When income is assumed to be exogenous she finds that income is positively related to self-assessed health and negatively associated with depressive symptoms, work limitations, functional limitations and bed days (Ettner, 1996). When she uses instrumental variable estimation she finds evidence that health and income are simultaneously determined. But the results still confirm that income exerts a significant and non-linear causal effect. These conclusions agree with those of Duleep (1995) and Fuchs (1974).

Adams et al. (2003) and the associated commentaries discuss the methodological issues involved in identifying causal effects of income on health. The problem facing such analyses is to overcome the biases, caused by reverse causality and selection, by identifying a source of exogenous variation in income. Meer et al. (2003) use inheritance as an instrument for wealth changes using the PSID from 1984-1999. They find a very small effect of changes in wealth requiring a quarter million dollar change to achieve an effect of around 2% points in probability of excellent of good health. Frijters et al. (2003) use the unanticipated shift in permanent income for East German households, following the reunification of Germany, as a source of exogenous variation in income. Using the GSOEP between 1991 and 1999 they find that, after controlling for heterogeneity, there is no evidence for a causal effect of income on health. Of course these findings do not rule out a cumulative and long-term relationship running from SES during childhood and early life to the gradient in adult health (see e.g., Case et al., 2002, Smith, 2004).

2.2 Health and earnings

There is little evidence concerning the impact of health on wages, particularly for developed economies. Very little evidence exists using recent European data. There are a number of reasons that health may have an impact on wages in a developed economy. Firstly, as noted by Mushkin (1962), Grossman and Benham (1974), Luft (1975) and Berkowitz et al. (1983), an increase in health leads to an increase in productivity, which should be reflected in an increased wage rate. Secondly, an employer may perceive health to be correlated with unobservable attributes that affect
productivity and hence offer higher wages to healthier individuals. Thirdly, an individual may be discriminated against because they are unhealthy, irrespective of their productivity.

The existing literature has in general found a positive impact of healthiness on wages and/or income. However, the vast majority of work in this area uses cross-sectional data and various forms of instrumental variable techniques. These techniques require finding valid instruments which are uncorrelated with the error term and predict well the endogenous variables for each equation estimated. As noted by Haveman et al. (1994), this becomes more problematic as the number of equations increases.

Berkowitz et al. (1983) examine a model where health capital influences productivity as well as labour supply. Using US data on white males from the survey of disabled and non-disabled adults and eight dichotomous indicators of impairments, they find that six indicators have a negative effect on wages using GLS. Lee (1982) estimates a general simultaneous equations model with multiple discrete indicators for unobserved health capital. Using US data from the National Longitudinal Survey of Men for 1966, he finds that unobserved health capital has a positive impact on wages both before and after accounting for endogeneity, with the coefficient 25% lower after accounting for the potential impact of wages on health. Lee’s estimates also indicate a potential bias in estimates of the return to schooling when health is considered exogenous. Haveman et al. (1994) estimate a simultaneous equations model for work-hours, wages, and health, using a Generalised Method of Moments estimator to account for simultaneity and to allow weak restrictions on the covariance-structure of the model. Using longitudinal data on 613 white males observed over 8 years from the US Panel Study of Income Dynamics, they find lagged ill-health (measured by a dichotomous indicator of work-limitations) to reduce wages and find a larger effect after accounting for endogeneity. This qualitative result was also found by Grossman and Benham (1974) using two-stage least squares.

Sundberg (1996) estimates a three equation simultaneous equations model of health, work hours and wages using Swedish data from 1991. She uses a self-assessed health variable and converts it into a continuous variable assuming a standard log-normal distribution for the latent index characterizing health. Using three-stage least squares Sundberg (1996) finds qualitatively similar results to Haveman et al. (1994), at least for men. For women the impact of self-assessed health on wages is insignificant. Using the first three waves of the British Household Panel Survey, Walker and Thompson (1996) estimated a model of hourly wages which included measures of disability. Applying both OLS and procedures to account for selective participation, they found disability to reduce years of schooling, wages and the probability of labour force participation. They found
disability to mainly affect participation rather than wages, and that once the endogeneity of schooling had been accounted for the effect of disability on wages was very small. Madden (1999) used cross-sectional data on 8747 couples from the UK 1995 Family Resources Survey, and attempts to decompose the healthy-unhealthy wage differential into productivity and discrimination components. He finds that health status is endogenous in wage equations for males and females, and that having taken into account the direct effect of health status on productivity, discrimination is, in general, an insignificant component of the observed mean wage differential.

Contoyannis and Rice (2001) consider the effect of self-assessed general and psychological health on hourly wages using longitudinal data from the six waves of the British Household Panel Survey. They employ single equation fixed effects and random effects instrumental variable estimators suggested by Hausman and Taylor (1981), Amemiya and MacCurdy (1986), and Breusch, Mizon and Schmidt (1989). Their results show that reduced psychological health reduces the hourly wage for males, while excellent self-assessed health increases the hourly wage for females. They also confirm the findings of previous work by Cornwell and Rupert (1988) and Baltagi and Khanti-Akom (1990), which suggested that the majority of the efficiency gains from the use of the instrumental variables estimators fall on the time-invariant endogenous variables, in their case academic attainment, and add further support to the hypothesis of a negative correlation between educational attainment and individual characteristics which affect wages.

2.3 Health and retirement

Health is undoubtedly an important factor in the decision to retire. Economic theory on the relationship between retirement and health (for example Lazear, 1986) states that agents have preferences over current and future leisure which depend in part on current and expected health status. Poorer health reduces the probability of continued work because it may increase the disutility of work; reduce the return from work (via lower wages); entitle the individual to benefits and other non-wage income that is contingent on not working. A possible counteracting effect is that poorer health may increase consumption requirements (for example via increased health care costs). However, if poorer health also reduces life expectancy then the annualised consumption available from existing wealth is raised, and this may still lead to earlier retirement.

There is a growing literature on health and work, and for older workers the retirement decision is a key part of this. Health effects operate alongside the effects of the pensions and benefits system, and there is an enormous literature on the importance of these financial incentives in determining retirement behaviour (Lumsdaine and Mitchell, 1999). However, Lindeboom (2006) in a
comprehensive review of the work in this area, argues that a number of empirical studies have suggested that health is the most important determinant of an older persons labour supply; a finding rejected by other studies, which point to problems in finding an appropriate measure of health and problems arising from the endogeneity of health in models of retirement.

There are many reasons why one may expect biases to arise when modelling the impact of health on the timing of the retirement transition. First, self-reported measures of health are based on subjective judgements and there is no reason to believe that these judgements are comparable across individuals. Secondly, self-reported health may not be independent of labour market status. Thirdly, since ill-health may represent a legitimate reason for a person of working age to be outside the labour force, respondents not working may cite health problems as a way to rationalize behaviour. Fourthly, for individuals for whom the financial rewards of continuing in the labour force are low there exists a financial incentive to report ill-health as means of obtaining disability benefits, this is often cited as the ‘disability route into retirement’ (Riphahn, 1997; Blundell, Meghir and Smith, 2002). For example, in a study of social security benefit programmes in the Netherlands, Kerkhofs and Lindeboom (1995) show that recipients of disability insurance systematically overstated their health problems.

Bound (1991) identifies the bias that results from each of the above problems. Reporting heterogeneity resulting in a lack of comparability across individuals in self-assessed health (SAH) represents measurement error that leads to an underestimation of the impact of health on labour force participation. Conversely, endogeneity in the health-retirement relationship will lead to an overestimation of the impact of health. These biases will also have implications for the estimation of the impact of socio-economic characteristics that are correlated with SAH. Indeed, Bound argues that SAH is in part, determined by socio-economic factors and that should the impact of health on labour market activity be correctly estimated the impact of economic variables may still be biased.

Bound (1991) also points out, however, that we cannot be sure that objective measures are any better predictors of the relationship between health and labour market status; objective measures of health are unlikely to be perfectly correlated with the aspect of health that affects an individual’s capacity for work. As such, objective measures of health will suffer from an error in variables problem leading to downwardly biased estimates of the impact of health on retirement. Whereas the biases involved in using self-assessed measures of health act in a way so as to counteract each other, the bias associated with objective measures of health operates in one direction only and hence may be more problematic in empirical applications.
Empirical studies on the role of health on retirement provide mixed conclusions about the endogeneity of SAH and the extent of the bias provided through measurement error. For example, while Kerkhofs, Lindeboom and Theeuwes (1999) find that the choice of health measure (SAH versus more objective measures) does affect the estimates of health on labour market outcomes, Dwyer and Mitchell (1999) conclude that SAH is not endogenous and their models of labour market retirement do not suffer significantly from measurement bias. Further, by applying a direct test for measurement error Au, Crossley and Schellhorn (2005) report significant error in their SAH variable. However, when this measure was used to predict retirement behaviour it gave similar results to those obtained from using a more objective measure of health and to those obtained through instrumental variable approaches.

Of further relevance is whether change in labour market status (into retirement) is best identified by a ‘shock’ to an individual’s health or by a levels effect (for example, a slow deterioration in health status). It is often argued that modelling health ‘shocks’ is a convenient way to eliminate one source of potential endogeneity bias caused through correlation between individual-specific unobserved factors and health (see for example, Disney, Emmerson and Wakefield, 2004). This is due to the identification of a health ‘shock’ through, for example, differencing the data over consecutive time periods which consequently implies eliminating unobserved individual effects. Disney, Emmerson and Wakefield (2004) find that health shocks are an important determinant of retirement behaviour in the UK, and that positive and negative health shocks tend to have symmetric effects.

Riphahn (1999) has also investigated the dynamic effect of health shocks on the employment and income of older workers. She finds significant effects of a health shock on leaving employment and finds small reductions in individual and equivalent household income. Another interesting fact is that health shocks seem to happen more often to those individuals/households which are already at the lower end of the income distribution before the health shock occurs. For the US, Bound et al. (1999) use 3 waves of the Health and Retirement Study to consider the retirement behaviour of men and women aged 50-62. They find that changes in health are as important as the long-term level of health in determining the retirement age.

2.4 Socioeconomic inequalities in health and health care in Europe

Van Doorslaer and Koolman (2004) have provided new evidence on the sources of differences in the degree of income-related inequalities in self-assessed health in thirteen European Union
member states. They went beyond earlier work by measuring health using an interval regression approach to compute concentration indices and by decomposing inequality into its determining factors. Comparable data were used, taken from the 1996 wave of the European Community Household Panel. Significant inequalities in health favouring higher income groups emerged in all countries, but were found to be particularly high in Portugal and – to a lesser extent – in the UK and in Denmark. By contrast, relatively low health inequality was observed in the Netherlands and Germany, and also in Italy, Belgium, Spain, Austria and Ireland. There was a positive correlation with income inequality per se but the relationship was weaker than in previous research. Health inequality is therefore 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”. They also find a substantial contribution of regional health disparities to socio-economic inequalities, primarily in the Southern European countries.

Van Doorslaer, Koolman and Jones (2004) have presented new international comparative evidence on the factors driving inequalities in the use of GP and specialist services in 12 EU member states using data are taken from the 1996 wave of the European Community Household Panel (ECHP). They examined 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. They 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. Van Doorslaer, Masseria et al (2004) have added to this evidence by including a number of other OECD countries, including Australia, Canada, Hungary, Mexico, Norway, Switzerland and the United States. While in most OECD countries general practitioner care is found to be distributed fairly equally and often even pro-poor, the very pro-rich distribution of specialist care tends to make total doctor utilisation somewhat pro-rich. This phenomenon appears to be universal, but offering private insurance or private care options reinforces this tendency. A number
of country-specific studies have provided further in-depth understanding of, for instance, the varying public-private insurance or patient arrangements in countries like Italy (Atella et al., 2004), Spain (Rodriguez and Stoyanova, 2004) and France, (Buchmueller et al., 2004).

3. Data and methods for ECUity III

3.1 The ECHP

The European Community Household Panel Users Database (ECHP-UDB) is a standardised annual longitudinal survey, designed and coordinated by the European Commission's Statistical Office (EUROSTAT). It provides 8 waves (1994 - 2001) of comparable micro-data about living conditions in the European Union Member States. 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 (Peracchi, 2002). “National Data Collection Units” implemented the survey in each country. Approximately, 60,000 households and 130,000 adults across the European Union were interviewed at each wave. The survey covers a wide range of topics including demographics, income, social transfers, individual health, housing, education and employment. The information provided by the ECHP-UDB can be compared across countries and over time.

The first wave covered the EU-15 Member States with the exception of Austria, Finland and Sweden. Austria joined in 1995 and Finland in 1996. In the periods covering the first three waves, the ECHP ran parallel to existing national panel surveys in Germany, Luxembourg and the United Kingdom. From the fourth wave onwards, the ECHP samples were substituted by data harmonized ex-post from these three surveys. Hence, there were two versions of the ECHP database for Germany, Luxembourg and United Kingdom. Although Sweden did not take part in the ECHP, the Living Conditions Survey is included in the UDB, together with comparable versions of the British Household Panel Survey (BHPS), the German Socioeconomic Panel (GSOEP) and the Panel Survey for Luxembourg (PSELL).

3.2 Measurement and explanation of socioeconomic inequalities in health

3.2.1 Concentration and Gini indices

In order to measure socioeconomic or income-related inequality in health, economists have borrowed tools from the income inequality literature. Foremost among these is the health concentration index, which provides a measure of relative income-related health inequality.
The health concentration index is derived from the health concentration curve; there are various ways of expressing the concentration index \((C)\) algebraically. The one that is most convenient for our purposes is,

\[
C = \frac{2}{\mu} \sum_{i=1}^{N} (y_i - \mu)(R_i - \frac{1}{2}) = \frac{2}{\mu} \text{cov}(y_i, R_i)
\]

This shows that the value of the concentration index is equal to the covariance between individual health \((y)\) and the individual’s relative rank \((R)\), scaled by the mean of health in the population \((\mu)\).

Then the whole expression is multiplied by 2, to ensure the concentration index lies between -1 and +1. Writing the concentration index in this way emphasises that it is an indicator of the degree of association between an individual’s level of health and their relative position in the income distribution. Concentration indices are sometimes criticised for being hard to interpret; what does a value of, say, 0.04 mean? A recent contribution by Koolman and van Doorslearer (2004) helps to clarify the situation. They show that, if the concentration index is interpreted in terms of a hypothetical linear redistribution from rich to poor, it can be given a Robin Hood-type interpretation. This interpretation implies that 75 times the concentration index is the percentage of total \(y\) that would have to be redistributed from individuals in the richest half to individuals in the poorest half of the population to achieve an equal distribution.

Socioeconomic inequality in health is cited widely as a concern for health policy makers, however it may not be the whole story. Recent work at the World Health Organisation through their Evidence for Health Policy programme has argued that policy makers should also be concerned about other sources of inequality, and that measurement should focus on total health inequality (e.g., Gakidou et al., 2000). This can be analysed using health Lorenz curves and inequality can be measured using the Gini coefficient of health inequality, \(G\) (Le Grand, 1989 and Wagstaff, Paci and van Doorslaer, 1991). The attraction of this approach is that there is a direct relationship between the concentration index and the Gini coefficient for health: the concentration index is proportional to the Gini coefficient, where the factor of proportionality is given by the ratio between the correlation coefficient for health and income rank and the correlation coefficient between health and health rank (Kakwani, 1980; van Doorslaer and Jones, 2003). This means that it is easy to move between these measures of socioeconomic and pure health inequality.

The inequality literature makes a distinction between partial orderings, based on Lorenz or concentration curves, and complete orderings, based on index numbers such as the Gini and concentration indices. A partial ordering means that some combinations of distributions can be
ranked unambiguously, but not all. The ambiguity arises if the Lorenz or concentration curves for
two distributions cross each other. In order to obtain a complete ordering of distributions, Gini
coefficients and concentration indices embed particular normative judgements about the weight
given to individuals at different points in the income distribution and, hence, they embody a
particular degree of inequality aversion. Sensitivity of the results to inequality aversion can be
assessed by using extended Gini or concentration indices (Yitzhaki, 1983; Lerman and Yitzhaki,
1984; Wagstaff, 2002). These add an extra parameter that can range from inequality neutrality (no
concern for inequality) to extreme inequality aversion (Rawlsian lexi-min).

Gini and concentration indices are measures of relative inequality and do not address the equity-
efficiency trade-off. This trade-off can be captured by generalized Lorenz or concentration curves.
These multiply the Lorenz or concentration curve by the absolute level of health. A classic result
from the income equality literature – the Kakwani-Kolm-Shorrocks theorem – shows that
generalized Lorenz dominance is equivalent to a distribution having a greater level of social welfare,
for any welfare function that is increasing and concave in income. The generalized concentration
index, \( \mu(1-C) \), gives a single index that captures the trade-off between the mean of the distribution
(\( \mu \)) and the level of inequality. This can be combined with different degrees of inequality aversion,
through the extended concentration index, to give what Wagstaff (2002) calls an index of health
achievement. This index summarises the equity-efficiency trade-off for different degrees of
inequality aversion.

The proceeding analysis assumes that a cardinal measure of health is available. This is relatively
straightforward for indicators of illness, such as the presence of chronic conditions, as the
concentration index or Gini coefficient can be based on the head-count of the number of
individuals experiencing the illness. It is more difficult when health is measured using self-reported
subjective scales. Self-assessed health (SAH) is widely available in many general population surveys
and has been used extensively in the ECuity project. The problem with this measure is that
respondents are asked to describe their health in ordered categories and the variable is inherently
ordinal rather than cardinal. In the past researchers have dealt with ordinal measures of health
either by dichotomizing the variable so that individuals are described as either healthy or non-
healthy, or by imposing some sort of scaling assumption. The problem with the former is that
information is lost and not all of the health variation contained in the original SAH variable is used.
Evidence shows that comparisons of inequality over time or across populations may be sensitive
because the results differ depending on the choice of the cut-point between healthy and non-
healthy. A variety of methods have been used to re-scale the ordinal measure of health into a
cardinal measure. Early work in the ECuity project imposed a lognormal distribution on self-
assessed health. More recently external information such as, the average level of health utility within categories of self-assessed health have been used in the rescaling. A third approach is to adopt an appropriate econometric specification, such as the ordered probit model, and use the predictions from this model as a scaled measure of individual health.

Van Doorslaer and Jones (2003) suggest an approach that combines the use of external information with the ordered probit model. This relies on having a dataset that includes both self-assessed health and a cardinal index of health: in their case the Canadian National Population Health Survey (NPHS), which includes SAH and the McMaster health utility index (HUI). This is used to construct a mapping from HUI to SAH. On the assumption that there is a systematic relationship between the two measures of health - such that those at the bottom of the distribution of self-assessed health will also be those at the bottom of the distribution of health utility - it is possible to scale the cut-points for categories of self-assessed health using health utility values. These cut-points can then be incorporated into the ordered probit model and self-assessed health can be estimated as an interval regression, where the values of the cut-points are treated as known. The attraction of this approach is that predictions from the interval regression model are on the same scale as health utility.

The concentration index measures income-related inequality in health. This is not the same thing as inequity in health. For example, variations in health that are attributable to age and gender may be seen as unavoidable and hence legitimate sources of inequality. The same argument applies to measures of inequality in the use of health care (see e.g., van Doorslaer, Koolman and Jones, 2004). 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. 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 (see Jones and Rice (2005) for further discussion).

### 3.2.2 Decomposing inequality indices

Like the Gini coefficient of income inequality, the concentration index has the attraction that it can be decomposed by factors (Rao, 1969, Kakwani, 1980). For example this property has been used in the past to decompose the concentration index for health care financing into different sources of health care payments such as taxation, social insurance contributions, user charges etc. A recent paper by Wagstaff, van Doorslaer and Watanabe (2003) exploits the result that if a regression model for the association between health and a set of factors is additively separable,
\[ y_i = \alpha + \sum_k \beta_k x_{ki} + \epsilon_i, \]

then, because the concentration index is additively decomposable - which stems from the fact that the covariance of a linear combination is equal to the linear combination of covariances - the overall concentration index for health can be written as follows,

\[ C = \sum_k \left( \frac{\beta_k \bar{x}_k}{\mu} \right) C_k + GC_{\mu} / \mu = C_y + GC_{\mu} / \mu \]

This has the convenient form that \( C \) can be split into two parts, the first term can be thought of as the explained component \( \{C_y\} \) and the second term as the unexplained component. Within the explained component there is a contribution for each of the regressors \( (x) \) and this is made up of the product of two terms. The first term is the elasticity of health with respect to that variable, for example the income elasticity of health, and the second term is the concentration index of that variable, for example in the case of income this would be the Gini coefficient.

### 3.2.3 Measurement of inequality and mobility with panel data

Up to now we have focused on methods for the measurement and explanation of socioeconomic inequalities in health that have been designed for use with cross sectional data. Jones and López Nicolás (2003) explore what more can be gained by using panel data. Again it is possible to borrow from the income inequality literature. Work on income mobility has focused on comparing the distribution of income using two perspectives, first of all a cross sectional or short-run perspective and secondly a long-run perspective where income is aggregated over a series of periods. If an individual’s income rank differs between the short-run and the long-run there is evidence of income mobility. One way of measuring this phenomenon is through the index of income mobility proposed by Shorrocks (1978).

The aim of the paper by Jones and López Nicolás (2003) is to apply the same principles to income-related health inequality. They show that the long-run concentration index can be written as the sum of a weighted average of short-run concentration indices plus a term that captures the covariance between levels of health and fluctuations in income rank over time. This differs from income inequality in that income-related health inequality can be either greater or smaller in the long-run than the short-run but, once again, these changes can be measured through an index of health-related income mobility which is based on the familiar tools of the concentration index. The paper shows that this mobility index can be decomposed using the contribution of different factors.
through a regression model for health and this is illustrated using the GHQ measure of subjective well-being from the first nine waves of the British Household Panel Survey (BHPS). This shows that, after nine waves, the weighted average of short-run measures underestimates the long-run measure by 15% for men and 5% for women.

The distinction between the short-run and long-run will be of interest to policy makers whose ethical concern is with inequalities in long-run health. For example, the “fair innings” perspective suggests that equity should be defined in terms of a person’s lifetime experience of health (Williams and Cookson [28, p.1899]). In practice, this lifetime experience could be measured using DALYs (Murray and Lopez, 1996) or QALYs (Williams, 1997).

4. New findings from ECuity III

4.1 Health-related attrition

Evidence of health-related attrition has been explored 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 its consequences for models of the association between socioeconomic status and self-assessed health (see Jones, Koolman and Rice, 2005). Attrition may be important as there is a risk of survivorship bias: long-term survivors who remain in the panel are likely to be healthier on average than the sample surveyed at wave 1. To address this issue Jones, Koolman and Rice describe the pattern of health-related attrition revealed by the BHPS and ECHP data. Descriptive evidence shows that there is health-related attrition in the data, with those in poor initial health more likely to drop out. Variable addition tests provide evidence of attrition 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 - does not show substantive differences in the average partial effects of the variables of interest – income and education. So, while health-related attrition exists, it does not appear to distort the magnitudes of the estimated average partial effects of socioeconomic status. Similar findings have been reported concerning the negligible influence of attrition bias in models of various labour market outcomes; the authors discuss possible explanations for their results.

4.2 Reporting bias and heterogeneity

There is a concern that ordered responses on health questions may differ across populations or even across subgroups of a population. This reporting heterogeneity may invalidate group comparisons and measures of health inequality. Lindeboom and van Doorslaer (2005) propose a
test for differential reporting in ordered response models which makes a distinction between cut-point shift and index shift. The method is illustrated using Canadian National Population Health Survey data and the McMaster Health Utility Index (HUI) is used as a more objective health measure than the simple 5-point scale of self-assessed health. They find clear evidence of cut-point shifting for age and gender, but not for income, education or language.

Hernandez Quevedo, Jones and Rice (2005) explore reporting bias and heterogeneity in the measure of self-assessed health (SAH) used in the British Household Panel Survey (BHPS). The ninth wave of the BHPS includes the SF-36 general health questionnaire, which incorporates a different wording to the self-assessed health variable used at other waves. Considerable attention has been devoted to the reliability of SAH and the scope for contamination by measurement error; the change in wording at wave 9 provides a form of natural experiment that allows them to assess the sensitivity of panel data analyses to a change in the measurement instrument. In particular, they investigate reporting bias due explicitly to the change in the question. They show how progressively more general specifications of reporting bias can be implemented using panel data ordered probit and generalised ordered probit models. The results suggest that the distribution of SAH does shift at the ninth wave but there is little evidence that this varies with socio-economic characteristics at an individual level.

4.3 Long-run inequality and mobility in health limitations

A paper by Hernandez Quevedo, Jones, López Nicolás and Rice (2005) contributes to the literature on income-related inequalities in health across European Union Member States. The analysis is based on the European Community Household Panel Users' Database (ECHP-UDB) and uses two binary measures of health limitations for the full 8 waves of available data. Short-run and long-run concentration indices together with mobility indices are derived for indicators of severe health limitation and any health limitation. Results suggest the existence of “pro-rich” inequality in health across Member States in both the short-term and the long-term, with health limitations concentrated among those individuals with lower incomes. For many countries, short-run indices suggest income related inequalities in health are widening.

The ECHP dataset contains information on a wide range of health and health related variables, from health outcomes to health care utilisation. In our study, we are interested in the information on health limitations, in particular responses provided to the question: “Are you hampered in your daily activities by any physical or mental health problem, illness or disability?”. Three possible answers are available for the respondent: “Yes, severely”, “Yes, to some extent” and “No”. The
study focuses on two binary measures of health problems that have been derived from the responses to the health limitations question. From these responses, two dummy variables are constructed. The first, labelled HAMP1, represents an indicator of any limitations (severe or to some extent) versus no limitations; the second (HAMP2) represents an indicator of severe limitations versus no limitations or limited to some extent.

Figure 1 shows the long-run CIs and MIs for HAMP1 and HAMP2 for each country. For HAMP1 long-term concentration indices are negative for all the countries; hence, there are long-term income-related inequalities in health, with health limitations concentrated among those with lower incomes in the European Union. The estimates reveal Ireland as having the highest level of long-term “pro-rich” inequality in health (0.279), followed by Greece (0.223) and Denmark (0.201), while Germany (0.090), followed by Italy (0.108) and Finland (0.109), present the lowest levels. For HAMP2, the results are similar with Ireland (0.409) followed by Denmark (0.373) and Belgium (0.325) again exhibiting the highest levels of pro-rich inequality in health. For each country, the long-term CI is greater for severe limitations than it is for limitations to some extent indicating that inequalities become more pronounced when considering more severe health problems.

**INSERT FIGURE 1**

In terms of the mobility estimates, the mobility index for each country is negative, which suggests that there is greater long-run income related inequality in both HAMP1 and HAMP2, than would be inferred by the average of short-run indices, for all countries. Care should be taken in interpreting the results as some countries have less than eight waves of data. If the size of the mobility index is compared across countries that have the full eight waves, it can be seen that Belgium, followed by Ireland and Spain, have the highest mobility indices in absolute terms for HAMP1, while the lowest levels correspond to Italy and Spain. For HAMP2, Belgium, followed by Portugal and Ireland, have the highest mobility indices, while Denmark and Italy have the lowest estimates.

Several conclusions can be inferred from this analysis. First, there is evidence that income-related inequalities in health limitations exist among all Member States included in our analysis, both in the short-term and long-term. These socioeconomic inequalities favour the rich over the poor in each society. Secondly, there is evidence that inequalities in health are increasing in almost all the countries studied. Thirdly, there is an important difference between long-term and short-term measures of inequality, even over the relatively short span of 8 years covered by the ECHP-UDB. This highlights the importance of utilising a longitudinal perspective where feasible when measuring and interpreting socioeconomic inequalities in health.
4.4 Mortality, lifestyle and socioeconomic status.

Inequalities in health are partly explained by differences in lifestyle and living conditions, and lifestyle can vary between groups depending on economic circumstances. To better explain inequalities in health, it is appropriate to use a behavioural model, which contains socio-economic characteristics but also individual health decisions, and all potential determinants of health. Investments in health are assumed to be endogenous and to influence longevity. Balia and Jones (2005) investigate the relationship between individual socio-economic characteristic and mortality, emphasizing the role of lifestyles. The paper uses the British Health and Lifestyle Survey (HALS, 1984-1985) data and the longitudinal follow-up of May 2003 to investigate the determinants of premature mortality risk in Great Britain. They find that lifestyles and unobservable individual heterogeneity strongly contribute to inequality in mortality, measured by the Gini coefficient, reducing the contribution of socio-economic factors. A statistically significant correlation exists between the unobservable factors affecting the mortality equation and those affecting some of the lifestyle equations. This motivates the assumption of endogeneity of lifestyles and emphasizes the role of individual heterogeneity in a model of mortality. Contrasting with the more recent literature, which suggests that people in poor health select into unhealthy lifestyles (such as smoking), their model finds a strong evidence of selection of frailer individuals into non-smoking. Moreover, the evidence is that people who have a healthier style of life are less likely to die even if they are frailer. Individual's choices about their lifestyle may induce variations in health status and affect premature mortality. Health-related behaviours mediate the relationship between mortality and socio-economic characteristics.

4.5 Health and wages

The impact of income and earnings on health has been well-examined in the health economics literature while the impact of health on wages has been less studied. Even rarer in previous work is the possible difference between the influences of health on wages for men versus women. Gambin (2005a) attempts to fill this apparent gap in the literature. She augments the well-established earnings function to include a number of health indicators and estimate equations for men and women using eleven waves of the British Household Panel Survey. The paper considers a range of estimation procedures, including pooled ordinary least squares, random and fixed effects, and Hausman-Taylor instrumental variables approaches. The impact of health is found to differ slightly by sex and is more strongly related to women's wages than men's.
As there is such a divergence between men and women in developed countries regarding both wages and health, studying the interaction of health and wages and how the relationship differs by gender is an important addition to our understanding of the complex relationship between health and labour market outcomes. The analysis in Gambin (2005b) draws on individual level data from up to eight waves of the European Community Household Panel (ECHP). Estimation procedures are applied to unbalanced panels from 14 different countries. The samples consist of employed adults aged 24 to 64 years. The data is used in estimation of Mincer-type wage functions where the natural logarithm of an individual’s hourly wage is function of a number of individual specific characteristics such as age, education, work experience, type of job, and health. Two health variables are included: self-assessed health status and an indicator of chronic illness or disability. The first estimates are obtained from pooled ordinary least squares. Further estimates are obtained from random effects and fixed effects panel models. A gender-related difference in the association between health and wages has been found in several of the countries examined however; these differences are not the same in magnitude. For a number of countries, there appears to be no significant gender difference. Overall, self-assessed health has greater effects on men’s wages than women’s, while chronic illness appears to be more significant for women. The largest “gender-gap” seems to exist in France, Portugal, Spain and the United Kingdom.

4.6 Health and retirement

Along with most developed countries the UK is experiencing population ageing combined with, at least until very recently, increasingly early exit from the labour market of older workers. Health is often cited as an important factor in determining the timing of retirement decisions and Roberts, Rice and Jones (2005) investigate this phenomenon using 12 waves of the British Household Panel Survey (BHPS). They track individuals deemed to be at risk of labour market exit from work to retirement. Using discrete-time transition models they estimate the impact of health on the retirement hazard while controlling for confounding factors including income and pension entitlement. In tracking the same individuals over time they attempt to overcome the problems of endogeneity and unobservable individual heterogeneity that plague investigations of these causal relationships. They use a variety of measures of health and also construct a health stock variable that has the advantage of removing measurement error that may be inherent in subjective measures of general health status. Emphasis is also placed on the possibility that the health of a spouse (or partner) may influence the timing of the retirement decision. Results show a positive effect of health on retirement.
4.7 The use of health care

Bago d’Uva (2005a) models access to and utilisation of primary care using data from the British Household Panel Survey for the period 1991-2001. A latent class panel data framework is adopted to model individual unobserved heterogeneity in a flexible way in which individual effects are approximated using a discrete distribution. This framework offers an alternative representation of heterogeneity, where individuals are drawn from a finite number of latent classes. Accounting for the panel structure of the data leads to a substantial improvement in fit, and permits the identification of latent classes of users of health care. Analysis by gender shows that men and women respond differently to some factors, in particular, to age and income. There is evidence of a positive impact of income on the probability of seeking primary care. This effect is especially significant in the case of women. For both genders, the marginal effect of income on the propensity to visit a GP is greater for individuals who are less likely to seek primary care. A latent class aggregated count data model for the number of GP visits classifies individuals in three latent classes and shows a positive income effect particularly amongst those with lower levels of utilisation.

Bago d’Uva, Jones and van Doorslaer (2005) exploit the ECHP data to explore the evolution on inequalities and inequities in the use of health care over recent years. They analyse health care utilisation over the previous year, represented by the number of visits to a GP and the number of visits to a specialist. The paper presents concentration curves, concentration indices (short-run and long-run), and mobility indices for the number of visits to a GP and to a specialist, by country across time. Additionally, they present concentration curves for the average number of visits across waves, using as ranking variable the average income across periods (see Tables 1 and 2). Firstly, they analyse the inequalities for each country and wave. Secondly, they look into more detail at the long-run income-related inequalities, providing a comparison across countries. To compute indices of horizontal equity that exploit the panel data dimension of the ECHP they estimate latent class hurdle models, a model developed by Bago d’Uva (2005b).

5. The future

Previous work has shown that significant inequalities favouring the better off exist in all European countries, both with respect to the use of health care and with respect to the distribution of health itself, and that the degree of inequality is particularly associated with the way each society treats its aged population in terms of both income and health protection (Van Doorslaer and Jones, 2004).
The association between income and health is a consequence of (i) the impact of health and ageing on income, (ii) the reverse effects of income protection on health and health care use, and (iii) the joint determination of life cycle profiles of income and health by social and other factors.

As a result of the Amsterdam Treaty’s requirement that “the EU Council adopt measures for the production of statistics where necessary for the performance of the activities of the Community” a new source of data has been created for all members of the enlarged EU, i.e. the 25 current members and the 2 candidate entrants (Bulgaria and Romania): the EU Survey on Income and Living Conditions (or EU-SILC). Its first wave is to become available in 2006 and provide new and nationally representative household level data on income, labour, education, health status and housing for about 120,000 households in these countries. Health care utilization information is limited to questions about instances where specialist or dental care was not sought when it was needed. Data are collected in a rotating panel, implying that one quarter of the sample will be renewed each year, resulting in (a maximum of) four-year panel data observations, for most households. Data collection has started in all countries in 2004 or 2005 and public use versions should be released two years after collection. This means that cross-sectional data for the countries that started the first wave in 2004 ought to become available at the end of 2006. By 2007, the first wave data for all countries should be released.

While the ECHP and the EU-SILC are attractive and useful for cross-country comparisons, they often lack country-specific institutional detail (e.g. on health insurance coverage) and as general socio-economic surveys can collect only limited health information. We have therefore decided to complement the common data set-based comparative work with investigations using national country-specific survey data. Several country teams have access to other (panel) datasets and/or extended versions of the common core datasets that allow for analyses. Some countries (e.g. Denmark, Sweden, Norway, UK) can link survey data with mortality statistics, while other countries (e.g. Netherlands, Ireland, Panel Study of Belgian Households, German Socio-economic Panel) have either longer panels or have included more extensive health information. Still others (e.g. Finland) have linkages with hospital registrations.

The next phase of the project aims to address pertinent questions regarding the causal mechanisms underlying these systematic associations, by exploiting the significant variation in relative income and health positions, especially around the retirement age. It will take advantage of the availability of new longitudinal household level data sources for all EU member states and of the significant recent changes occurring in both new and EU15 member countries to examine these relationships.
References


Gambin, L. (2005a) Gender differences in the effect of health on wages in Britain


data. mimeo.


Figure 1: Long-run inequality and mobility, ECHP-UDB (source: Hernández Quevedo, Jones, López Nicolás and Rice, 2005)

a) HAMP1

![Graph for HAMP1](image1)

b) HAMP2

![Graph for HAMP2](image2)
Table 1: Long-run concentration indices ($CI^T$), and mobility indices ($MI^T$) for number of GP visits in ECHP-UDB (source: Bago d'Uva, Jones and van Doorslaer, 2005)

<table>
<thead>
<tr>
<th>Country</th>
<th>$CI^T$</th>
<th>$MI^T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>-0.090</td>
<td>-0.147</td>
</tr>
<tr>
<td>Belgium</td>
<td>-0.158</td>
<td>-0.102</td>
</tr>
<tr>
<td>Denmark</td>
<td>-0.104</td>
<td>-0.100</td>
</tr>
<tr>
<td>Finland</td>
<td>-0.033</td>
<td>-0.395</td>
</tr>
<tr>
<td>Germany</td>
<td>-0.081</td>
<td>0.011</td>
</tr>
<tr>
<td>Greece</td>
<td>-0.166</td>
<td>-0.244</td>
</tr>
<tr>
<td>Ireland</td>
<td>-0.159</td>
<td>-0.104</td>
</tr>
<tr>
<td>Italy</td>
<td>-0.076</td>
<td>-0.131</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>-0.077</td>
<td>0.057</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-0.070</td>
<td>0.037</td>
</tr>
<tr>
<td>Portugal</td>
<td>-0.092</td>
<td>-0.199</td>
</tr>
<tr>
<td>Spain</td>
<td>-0.109</td>
<td>-0.122</td>
</tr>
<tr>
<td>UK</td>
<td>-0.115</td>
<td>-0.070</td>
</tr>
</tbody>
</table>

Table 2: Long-run concentration indices ($CI^T$), and mobility indices ($MI^T$) for number of specialist visits in ECHP (source: Bago d'Uva, Jones and van Doorslaer, 2005)

<table>
<thead>
<tr>
<th>Country</th>
<th>$CI^T$</th>
<th>$MI^T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>0.042</td>
<td>-0.224</td>
</tr>
<tr>
<td>Belgium</td>
<td>-0.038</td>
<td>0.030</td>
</tr>
<tr>
<td>Denmark</td>
<td>-0.011</td>
<td>0.376</td>
</tr>
<tr>
<td>Finland</td>
<td>0.102</td>
<td>-0.129</td>
</tr>
<tr>
<td>Germany</td>
<td>0.013</td>
<td>-0.672</td>
</tr>
<tr>
<td>Greece</td>
<td>-0.076</td>
<td>-0.452</td>
</tr>
<tr>
<td>Ireland</td>
<td>0.027</td>
<td>-0.251</td>
</tr>
<tr>
<td>Italy</td>
<td>0.042</td>
<td>0.130</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>-0.029</td>
<td>0.272</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-0.025</td>
<td>0.388</td>
</tr>
<tr>
<td>Portugal</td>
<td>0.110</td>
<td>0.061</td>
</tr>
<tr>
<td>Spain</td>
<td>0.015</td>
<td>0.117</td>
</tr>
<tr>
<td>UK</td>
<td>-0.018</td>
<td>-0.056</td>
</tr>
</tbody>
</table>