TI 2010-122/3
Tinbergen Institute Discussion Paper

Education-related Inequity in Health Care with Heterogeneous Reporting of Health

Teresa Bago d'Uva[ab]
Maarten Lindeboom[bc]
Owen O'Donnell[abd]
Eddy van Doorslaer[ab]

[a] *Erasmus University Rotterdam;*
[b] *Tinbergen Institute;*
[c] *VU University Amsterdam;*
[d] *University of Macedonia.*

EDUCATION-RELATED INEQUITY IN HEALTH CARE WITH HETEROGENEOUS REPORTING OF HEALTH

Teresa Bago d'Uva[ab], Maarten Lindeboom[bc], Owen O'Donnell[abd] and Eddy van Doorslaer[ab]

a.   *Erasmus University Rotterdam*
b.   *Tinbergen Institute*
c.   *Free University Amsterdam*
d.   *University of Macedonia*

November 10, 2010

Abstract

Reliance on self-rated health to proxy medical need can bias estimation of education-related inequity in health care utilisation. We correct this bias both by instrumenting self-rated health with objective health indicators and by purging self-rated health of reporting heterogeneity identified from health vignettes. Using data on elderly Europeans, we find that instrumenting self-rated health shifts the distribution of doctor visits in the direction of inequality favouring the better educated. There is a further, and typically larger, shift the same direction when correction is made for the tendency of the better educated to rate their health more negatively.

1. INTRODUCTION

Throughout most of Europe there is a stated policy aspiration to equitable provision of health care. This is often interpreted as the horizontal equity principle of equal treatment for equal need. In empirical work, the prevailing approach is to measure horizontal inequity by the degree to which utilisation is related to socioeconomic status (SES) after controlling for differences in needs (Wagstaff and Van Doorslaer, 2000a,b). Existing European evidence reveals a bias favouring those of low SES in the distribution of primary care but inequity to the advantage of higher socioeconomic groups in specialist care (Van Doorslaer *et al*, 2000; Van Doorslaer et al, 2004).

The validity of this evidence is contingent on the adequacy of the measures of need employed. As comprehensive, objective health measures are seldom available in the general population surveys that provide data on health care utilisation and SES, researchers must usually rely on self-reported categorical health indicators. It is, however, possible that two equally healthy individuals report different levels of health because their conceptions of good health and their health expectations are contingent on their knowledge of disease and available medical treatments, as well as the observed health of their peers. To use a much-cited example from India, self-reported morbidity rates are highest in Kerala, the state with the highest life expectancy, the highest rate of literacy and the most extensive public health services (Dreze and Sen, 2002; Sen, 2002). They are higher still in the US (Murray and Chen, 1992; Sen, 2002). If health perceptions also vary with socioeconomic status, then self-reported health will not provide an unbiased benchmark of needs against which to measure inequity in health care utilisation. Then, as argued by van Doorslaer et al (2004), evidence of higher rates of utilisation of General Practitioner (GP) services in Europe among those of lower SES may not be indicative of inequity favouring the poor but, rather, a tendency of more disadvantaged groups to under-report morbidity. This has long been recognised (e.g. O'Donnell and Propper, 1991) but, hitherto, it has been difficult to do much about it. It has been shown that inclusion of a larger battery of health indicators (e.g. Van Doorslaer et al, 2000) or allowance for unobserved individual heterogeneity (e.g. Bago d'Uva et al, 2009) ameliorates the problem. But with data on self-reports only, the fundamental problem remains that differences in reporting behaviour cannot be disentangled from differences in true health.

Anchoring vignettes offer a potential solution to the problem. They identify reporting behaviour directly through survey respondents' ratings of case vignettes that describe fixed levels of functioning within a given health domain (Tandon et al, 2003; King et al, 2004). Systematic differences in these ratings by socio-demographic characteristics provide evidence of heterogeneity in reporting thresholds. Under assumptions, these differences in reporting styles can then be purged from the ratings of own health.

This study uses vignettes to identify differences in the reporting of health by education (and also by age, gender and country) and determines the impact of correcting for such differences on the measurement of education-related inequity in the utilisation of primary and specialist medical care. The focus on disparities by education is partly driven by the data used, which come from the Survey on Health, Aging and Retirement in Europe (SHARE) that samples from populations aged 50+. For older populations, education is a better indicator of long-term SES than income, which has been used in most previous European studies of inequity in health care utilisation (Van Doorslaer *et al*, 2000; Van Doorslaer *et al*, 2004; Bago d'Uva *et al*, 2009). Besides being a proxy for socioeconomic inequality, education disparities are of direct interest because of the strong association that exists between education and health (Smith and Kington, 1997; Mackenbach *et al.*, 2008; Cutler and Lleras-Muney, 2008). There is growing evidence from high-income countries that this relationship has a causal basis (Lleras-Muney, 2005; Oreopoulos, 2006; Van Kippersluis et al, 2009), unlike that between financial resources and health (Frijters et al, 2005; Smith, 2007). A potential causal mechanism is differential utilisation of medical care by education, which could run through income and health insurance coverage, and/or through the acquisition and utilisation of information on medical conditions and treatments (Glied and Lleras-Muney, 2008). One US study finds an education advantage in adoption of the most recent medicines that is not affected by control for income and health insurance coverage (Lleras-Muney and Lichtenberg, 2005). In the European context of universal and mostly publicly funded health insurance, financial barriers are likely to be even less important, relative to information, in explaining any education gradient in the use of medical care. Our aim in this paper is to establish the extent of the education disparity in utilisation, irrespective of the mechanisms through which it may operate, and so we purposefully do not control for income or supplementary health insurance coverage.

There are reasons to expect styles of reporting health to vary by education. Superior information acquisition skills possessed by more highly educated individuals increase the likelihood that they will recognise and report symptoms of disease. The strong correlation between health and education may generate peer effects, with low and high educated individuals reporting health relative to the average level observed among their respective associates. Both arguments imply that the higher educated will report their health more negatively. One strategy that has been adopted to test this hypothesis is to check whether self-reported health is a stronger predictor of mortality for more highly educated individuals. The evidence from both Europe and the US is mixed (Beam and Zajacova, 2007; Huisman et al, 2007; Jurges, 2008; Singh-Manoux et al, 2007). Adopting a different strategy, Lindeboom and Van Doorslaer (2004) find no education-related variation in reported health in Canada after conditioning on a presumed sufficiently comprehensive and objective indicator of health – the McMaster Health Utility Index. Vignettes offer a more direct test. Using the same SHARE data that are employed in the present paper, Bago d'Uva et al (2008) find that the more highly educated rate the health of vignette descriptions more severely in six of the eight countries examined. On the basis of this

evidence, we can predict that an observed education gradient in the response of health care demand to a given *reported* level of health will understate the true extent to which the better educated make greater use of health care for a given health condition. In this paper, we estimate the magnitude of this bias.

With one exception (Kapteyn et al, 2009), previous applications of the vignettes approach have been concerned with identification of and correction for reporting heterogeneity in the context of a descriptive analysis of the distribution of the error-ridden variable (eg, King et al, 2004; Kapteyn et al, 2007; Bago d'Uva et al, 2008). To our knowledge, this is only the second study to use vignettes to correct for measurement error in an independent variable and it is the first to do so in modelling health and health care. This involves development of a joint model of health care utilisation and health reporting.

Identification of reporting heterogeneity using vignettes rests on two assumptions – *vignette equivalence* and *response consistency* (see section 3.3). In a previous paper, we present evidence against both assumptions in the health domains of mobility and cognition for a sample of the older population of England (Bago d'Uva et al, 2009). These results cast some doubt on the validity of the approach and call for refinement of its implementation. In the present paper, we do not examine the method's validity but simply ask what impact it has when applied. *A priori* we expect that education-related inequity is underestimated when no account is taken of differences in reporting styles. In this context it is interesting to know whether the use of vignettes to 'correct' reporting heterogeneity shifts the estimates in the expected direction and by how much it does so. A partial, but by no means complete, alternative to vignettes is to instrument self-reported health using more objective health indicators (Bound, 1991). We examine the marginal impact of using vignettes to explicitly estimate systematic variation in reporting scales, in addition to exploiting objective health indicators. This allows us to gauge the marginal return to inclusion of a vignettes module in a health survey, in addition to measures of physical functioning, such as grip strength, and data on chronic conditions.

The next section describes the SHARE data. Following that, we propose a model of doctor visits as a function of medical need and explain how both objective health indicators and vignettes can be used to reduced bias resulting from the use of self-rated health as a proxy for need. Section 4 presents our findings on the extent to which the two corrections for measurement error in need impact on the estimation of education-related inequity in primary and specialist care. The final section concludes.

2. DATA

2.1 SAMPLE

SHARE randomly samples from populations aged 50 years and over (plus spouses) in 12 countries (Börsch-Supan et al, 2005a). We use data from the first wave collected in 2004-05 (release version 2.0). Vignettes data were collected from supplementary probability samples, which also completed the full SHARE questionnaire, in all but four countries (Börsch-Supan et al, 2005b). The overall response rate in the vignette samples was 57.7% - highest in France (77%) and lowest in Belgium (42%). The countries where the vignettes were fielded and the respective sample sizes are as follows: Belgium (n=531), France (n=773), Germany (n=468), Greece (n=646), Italy (n=397), The Netherlands (n=489), Spain (n=414) and Sweden (n=380). We pooled these samples to give a total sample of 4098 individuals. Despite the fact that the sample size relative to population differs across countries, we do not apply weights to make the sample representative of the cross-country population. For modelling purposes, this is not necessary provided one conditions on country indicators.

2.2 HEALTH CARE UTILISATION

We examine variation both in the number of visits to a General Practitioner (GP) and visits to a specialist. Respondents are asked to report the number of times they saw or spoke with a medical doctor about their health in the last twelve months. A reminder is given of the month last year from which they should count. They are told exclude dental care and hospital stays but to include consultations in the emergency room and outpatient clinics. Of the total number of visits recorded, they are asked how many were with a GP or a doctor at a health centre. We use this as our measure of GP visits. The remainder is used as the measure of specialist visits.

While there is likely to be measurement error in the recall of the number of doctor visits over a period of one year, there is no obvious reason why this should be correlated with education, particularly since we control for cognition. Nonetheless, it should be kept in mind that, as with all previous studies, any correlated measurement error would bias the estimated education-related inequity in doctor visits.

The frequency of visiting a GP is highest in Italy, at just less than seven visits on average per year, and lowest in Sweden, where, on average, older individuals visit less than twice per year (Table 1). The standard deviation is generally large relative to the mean, particularly in Greece, Italy and Spain. In all countries except Greece, the Netherlands and Sweden, 80-90% of individuals consult a GP at least once a year. It is the low probability of any contact that is responsible for the relatively low mean number of visits in Greece, while both the probability and the intensity of visiting are low in both the Netherlands and Sweden. Mean visits are high in

Italy because of their intensity among those who make any consultation. Some substitution is suggested by the fact that Italy also has the lowest mean number of visits to a specialist. Spain, another country in which the GP acts as a gatekeeper to specialist care, also has a relatively low ratio of mean specialist to GP visits. But the Netherlands also operates a GP gatekeeper system and has the highest such ratio. Sweden is striking in having the lowest mean number of consultations with both GPs and specialists. Around 40% of the sample consults with a specialist each year in all of the other countries except Belgium and Germany, where more than half do so. Both GP and specialist visits are highly skewed with a few individuals consulting the doctor almost every week in most countries, and twice per week in some cases.

Table 1 here

2.3 EDUCATION & DEMOGRAPHICS

Educational attainment is measured according to the International Standard Classification of Education (ISCED 97) as: (i) finished at most primary education or first stage of basic education (ISCED 0-1 - *primary*); (ii) lower secondary or second stage of basic education (ISCED 2 – *lower secondary*), (iii) upper secondary education (ISCED 3-4 – *upper secondary*) and (iv) recognized third level education, which includes higher vocational education and university degree (ISCED 5-6 - *tertiary*) (UNESCO, 1997). In the case of Germany the data do not distinguish between levels (i) and (ii). The lowest education group is used as the reference category. As explained in the introduction, the purpose of the analysis is to measure education-related disparities in health care use conditional only on medical need. We therefore control only for indicators of health, age and gender. Exploratory analysis found no evidence of significant urban/rural differences in either health reporting or health care utilisation. Given the very small size of the rural samples in two countries (Greece and Spain), and the possibility that location is determined by education, we do not control for this characteristic.

2.4 SELF-REPORTED HEALTH AND VIGNETTES

A self-completion drop off questionnaire covered self-assessments of health and vignette ratings. Respondents were asked to rate problems or difficulties experienced in the preceding 30 days in each of six health domains – *mobility, cognition, pain, sleep, breathing* and *emotional health* – on a five-point scale running from *none* to *extreme*. The questions referred to problems/difficulties with: moving around (*mobility*); concentrating or remembering things (*cognition*); bodily aches or pains (*pain*); falling asleep, interrupted sleep or waking early (*sleep*); shortness of breath (*breathing*); feeling sad, low, or depressed (*emotional health*). For more detail on these questions and the choice of health domains the reader is referred to Bago d'Uva et al (2008) and references therein.

After rating their own health, respondents were presented with three vignettes within each domain, each of which is a brief description of a case corresponding to a given level of difficulty in that domain (see http://ije.oxfordjournals.org/content/37/6/1375/suppl/DC1). The respondent was asked to rate the level of difficulty, or problem, represented by each vignette using the same five response categories used to rate own-health. One of two versions of the vignettes module – differing in the implied genders (names) of the vignette and their ordering - was randomly assigned to respondents.

Ratings of the vignettes do indicate that reporting styles vary by education. To illustrate, in Table 2 we present the proportion of respondents across all countries who classify the vignette intended to correspond to the middle level of difficulty within each domain as experiencing *none* or *mild* difficulty. The proportions are standardised for age, gender and country and so do not reflect education-related differences in these characteristics. With the exception of cognition, there is a clear tendency for the lower education groups to be more likely to consider the health condition described in the vignette as being one that presents no, or only a mild degree, of difficulty. For example, while 24% of those with only primary level education consider pain in the arm that is relieved in the evening after work to represent no more than a mild (pain) problem, only 16% with tertiary education are so sanguine.

Table 2 here

In a related paper using the SHARE data (Bago d'Uva et al, 2008), some of us showed the implications of these education-related differences in health reporting for measurement of heath inequality. This analysis revealed significant reporting differences by education in 29 out of 48 domain-country cases examined, with the higher educated more likely to rate a given health state negatively. Correcting these differences generally increased measured inequality and resulted in the emergence of significant health disparities to the advantage of the better educated in 18 cases. Spain and Sweden displayed different patterns of health reporting by education. In the present study, we take account of such cross-country variation not only by allowing health reporting to vary by country but also permitting the education differences in reporting to vary by country.

2.5 HEALTH INDICATORS

Education-related differences in the propensity to report health problems is further evident from variation in objective health indicators within given categories of self-rated health. This is illustrated in Table 3 for the domains of *mobility, cognition* and *depression.* In each case, lower educated individuals score worse on corresponding objective measures of health than do higher educated individuals. This is true after standardizing for differences in age, gender and the proportion of respondents from each country across the education categories. For example, among individuals reporting no problems with mobility, measured mean grip strength,

a predictor of mobility limitations (Rantanen et al, 1999) as well as mortality (Rantanen et al, 2003; and references therein), is 34.8 for those with no more than primary education but 35.6 for those with tertiary level education. Even larger differences are observed for those reporting moderate problems with mobility. Among individuals reporting severe or extreme difficulty concentrating or remembering things, 36% of those with no more than primary education, but only 15% of those with tertiary education, are found to have some difficulty recalling the date, month, year or day of the week. Of the sample respondents reporting severe or extreme problems with feelings of sadness or depression, 66% of those with primary education versus 57% of those with tertiary education are identified as depression cases using the EURO-D depression scale (Prince et al, 1999).

Table 3 here

We exploit the health indicators for two main purposes. First, the additional health variables allow us to predict health within each domain from a model of reported health as a function of objective indicators and then substitute these predictions in a model of doctor visits. This instrumental variable approach removes errors-in-variables bias from the estimated effect of need on doctor visits (Bound, 1991). Second, the six health domains are unlikely to be fully comprehensive in capturing needs for primary and specialist medical care. Some dimensions of health - notably sight and hearing problems - are not covered by the six domains. Chronic conditions may be under medical management that requires regular consultation with a doctor for check-up and prescription of medication and yet be sufficiently well contained such that related health problems or difficulties are not reported.

The structural framework we adopt is one in which an individual consults a doctor either in response to a symptom, which may provoke reporting of a health problem within one of the six health domains, or for continued management of a chronic condition. Consistent with this, we specify doctor visits as a function of the need for curative care and for disease management, which are proxied by the six health domains and a list of diagnosed chronic conditions respectively. For each domain, self-rated health is specified as a function only of indicators that are symptoms of a health condition or disease within that domain, or that identify such a condition/disease. In the presence of comorbidities, an indicator may be correlated with reported health in a domain without being a symptom of any recognised medical condition within that domain. For example, in the SHARE data grip strength is correlated with cognitive problems. We do not, however, use grip strength to explain reported cognitive functioning since it is not a symptom of any cognitive problem for which medical advice may be sought. Following this logic, the choice of indicators per domain, which is listed in Table 4, was made following expert medical advice. The table also lists the diagnosed conditions that are permitted to impact on doctor visits directly, and not only through one or more of the health domains.

Table 4 here

8

3. JOINT MODEL OF DOCTOR VISITS AND HEALTH REPORTING

3.1 IDENTIFICATION PROBLEM

Our aim is to test for education-related variation in doctor visits conditional on a defined level of medical need. The problem is that need is fundamentally unobservable and its mis-measurement is likely to induce bias in the estimated education effect. This is analogous to the identification problem examined by Bound (1991) in the context of estimating (unobserved) disability and wage effects on labour supply. Let doctor visits (y) be determined as follows, $y=y(\eta_1,\eta_2,E)$, where η_1 represents need for treatment of current health problems and symptoms, η_2 is need for the medical management of diagnosed chronic conditions and E is education. Obviously, other determinants could be added. We will assume there is no problem with measurement of the second type of need. The set of diagnosed conditions listed in Table 4 is presumed to comprehensively represent all those for which an individual may consult for medication or check-up. Self-reported health problems (h) are potential proxies for the first type of need. There are two problems with this approach. First, substituting h for η_1 in a model of doctor visits will introduce errors-in-variables that will result in an underestimation of the need effect and will spill over to biased estimation of the education effect unless, conditional on self-rated health, there is no correlation between education and need. Second, as illustrated in Tables 2 and 3, it is likely that reporting styles vary with education, such that $h=h(\eta_1,E)$. This introduces an endogeneity bias to the estimated education effect on doctor visits that will be present even in the absence of any errors-in-variables bias.

Using objective health indicators, such as those listed in Table 4, to instrument self-rated health in a model of doctor visits does not solve the problem provided there is education-related reporting heterogeneity. The need effect would be consistently estimated but the education effect would not. As Bound (1991) demonstrates, such a model is only identified in the presence of outside information about reporting behaviour. We use vignettes to identify that information and then impose it in a model of doctor visits as a function of education and a measure of need purged of reporting heterogeneity. We compare the estimates from this model with those from two others. The first uses the raw self-rated health variables to proxy need and is vulnerable to bias both from errors-in-variables and from education-related reporting errors. The second instruments self-rated health with objective indicators and so deals with the first bias but not the second.

3.2 HURDLE MODEL OF DOCTOR VISITS

The descriptive statistics presented in Table 1 reveal a substantial proportion of individuals report no doctor visits, particularly specialist consultations, in the last year. In order to allow for the likelihood that the stochastic process responsible for observing zero visits differs from that generating a positive number of visits, we

estimate a hurdle model consisting of a logit specification of the probability of any visits and a truncated negative binomial 2 specification in the positive range of visits (Grootendorst, 1995; Gurmu, 1998; Winkelman, 2004). Let y_i denote the number of visits to the doctor (GP or specialist) made by individual i in the last year and $d_i=1(y_i>0)$, where 1() is the indicator function. The probability of observing a given number of doctor visits is (Gurmu, 1998; Winkelman, 2004),

$$f\ y_i\ =\ 1-f_{1i}\ ^{1-d_i}\left[\,f_{1i}f_T\ y_i\,|\,y_i>0\ \right]^{d_i} \tag{1}$$

where $f_{1i} = P\ y_i > 0; \mathbf{\beta}_1|\mathbf{X}_i\ = \lambda_{1i}/\ 1-\lambda_{1i}\ ,\ f_T\ y_i|y_i>0\ = f_2\ y_i\ /\!\left[1-f_{2i}\ 0\ \right],$

$$f_{2i}=P\ y_i;\alpha,\mathbf{\beta}_2\mid\mathbf{X}_i\ =\frac{\Gamma\ \alpha^{-1}+y_i}{\Gamma\ \alpha^{-1}\ \Gamma\ y_i+1}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda_{2i}}\right)^{\alpha^{-1}}\left(\frac{\lambda_{2i}}{\lambda_{2i}+\alpha^{-1}}\right)^{y_i}\quad,$$

$\lambda_{si}=\exp\ \mathbf{X}_i\mathbf{\beta}_s\ ,\ s=1,2$, $\Gamma(.)$ is the gamma function and $\alpha>0$ denotes an over-dispersion parameter that is estimated. It is assumed that the two parts of the model (i.e. the participation decision and the positive number of visits) are stochastically independent and so the log-likelihood factors into two components that can be estimated separately.

The vector $\mathbf{X}_i$ includes four types of variables: i) education, ii) demographics and country indicators, iii) health specific to the six reported domains ($\mathbf{H}_i$), and iv) diagnosed health conditions or diseases. Education is interacted with the country indicators in the more flexible models estimated. We estimate the model using alternative sets of domain-specific health variables. First, we simply use the chosen category within each of the health domains, $\mathbf{H}_i =\ \mathbf{H}_{1i}, \mathbf{H}_{2i}, ..., \mathbf{H}_{Di}$. In this case, self-reports are taken at face value and the estimates are vulnerable to both types of bias referred to above. Presuming doctor visits increase with true medical need and, as is likely, need is negatively correlated education, then the errors-in-variables bias will result in an underestimation of the effect of education on doctor visits (Bound, 1991, p.111). If the better educated understate their health, then this will further bias the education effect downwards. The degree of the latter bias will increase both with the degree of reporting heterogeneity and with the responsiveness of health care use to need.

Our second approach is to proxy need for curative care with domain-specific health scores derived from ordered probit models of self-rated health as a function of more objective health indicators, as well as education, demographics and country indicators. The category reported in domain d, H_{di}, is assumed to be generated by the position of a latent health index H_{di}^*, specified as

$$H_{di}^* = \mathbf{Z}_{di}\boldsymbol{\delta}_d + \mathbf{W}_i\boldsymbol{\gamma}_d + \varepsilon_{di}, \qquad \varepsilon_{di}/\mathbf{Z}_{di},\mathbf{W}_i \sim N\ 0,1\ , \qquad\qquad (2)$$

relative to a set of fixed cut-points $-\infty < \tau_d^1 < \dots < \tau_d^{K-1} < \infty$, such that,

$$H_{di} = k \quad \text{if} \quad \tau_d^{k-1} \le H_{di}^* \le \tau_d^k, \qquad k = 1,\dots K \qquad\qquad (3)$$

where $\mathbf{Z}_{di}$ is the corresponding vector of health indicators listed in Table 4 and $\mathbf{W}_i$ includes education, age, sex and country (interacted with education). The predicted latent health scores from these models, $\mathbf{H}_i = \left[\hat{H}_{1i}^*, \hat{H}_{2i}^*, \dots \hat{H}_{Di}^*\right]$, are entered into the hurdle model for doctor visits.

Besides non-linearity, identification relies on the exclusion of health indicators from the model of doctor visits. The exclusion restrictions follow from the presumption that individuals consult either for the treatment of current symptoms in the six health domains or for management of chronic conditions. All health indicators other than diagnosed chronic conditions are permitted to impact on utilisation only through health problems in the six health domains. For example, conditional on mobility problems, grip strength is assumed to have no direct influence doctor visits. Cognitive test scores are presumed to influence consultations only through problems of concentration and memory. The (excluded) instruments for each health domain are given by the indicators listed in the respective row of Table 4, with the exceptions of those also appearing in the bottom row that enter the doctor visits model directly.

Constancy of the cut-points across individuals implies an assumption of reporting homogeneity. If this does not hold, in particular, if the cut-points vary with some of the covariates, then the predicted health scores will reflect not only information on true health but also variation in the reporting of that health. So, while this instrumental variables approach corrects errors-in-variables bias deriving from the mis-measurement of need, it does nothing to deal with the bias due to education-related reporting errors. Under the assumptions stated above, the education effect will continue to be underestimated, although not to the same extent as when raw self-reported health is used to proxy need.

3.3 HETEROGENEOUS REPORTING BEHAVIOUR

Our third specification of the domain-specific health variables uses predictions of latent health scores from an extended ordered probit model in which the reporting thresholds are made functions of individual characteristics and so the parameters of the latent index represent true health effects, and not a mixture of health and reporting effects. This hierarchical ordered probit (HOPIT) model (King et al, 2004) is identified using the information from the vignette ratings. Validity of the approach rests on two assumptions - *vignette equivalence* and *response consistency.* The first corresponds to the assumption that, up to random

11

measurement error, all respondents understand the vignette description as corresponding to the same level of functioning on a uni-dimensional scale. This is required in order that responses to a given vignette can be interpreted as reflecting heterogeneity in the reporting of a given level of functioning. In the present context, the assumption is made more plausible by the separation of health into six domains with vignettes being rated within each of these. It does require that language translation does not distort vignette descriptions, resulting in differences in their interpretation across countries. Given improvement of cross-country comparability was the main motivation for the inclusion of vignettes in SHARE, one presumes sufficient care was taken with their translation. The second assumption is that respondents rate the vignettes in the same way as they do their own health. If this did not hold then it would not be valid to impose the thresholds identified from the vignettes ratings on the reporting of own health, and so the true health effects would not be identified. To date, there has been little formal testing of these identifying assumptions. On vignette equivalence, see Murray et al (2003), Kirstensen and Johansson (2008), Bago d'Uva et al (2009), and Rice et al (2009). On response consistency, see Van Soest et al (2007), Bago d'Uva et al (2009) and Datta Gupta et al (2010).

Under the assumption of *vignette equivalence,* the perceived latent health level of vignette *j* in domain *d*, V_{jdi}^*, depends solely on an indicator of the vignette and random error (and *not* on characteristics of the respondent):

$$V_{jdi}^* = v_{jd} + \upsilon_{jdi}, \qquad \upsilon_{jdi} \sim N\ 0,1\ . \tag{4}$$

The observed vignette ratings V_{jdi} relate to V_{jdi}^* through the thresholds:

$$V_{jdi} = k \quad \text{if} \quad \tau_{di}^{k-1} \leq V_{jdi}^* \leq \tau_{di}^k,\ k = 0,....,K, \tag{5}$$

$-\infty < \tau_{di}^1 < ... < \tau_{di}^{K-1} < \infty$, which are now defined as functions of the same covariates that enter the latent index of own-health in (2),

$$\tau_{di}^k = \mathbf{Z}_{di}\mathbf{\theta}_d^k + \mathbf{W}_i\mathbf{\phi}_d^k \tag{6}$$

Inclusion of the individual's characteristics in the cut-points is possible because the assumption of *vignette equivalence* ensures that all the systematic variation in the vignette ratings can be attributed to reporting behaviour. In principle, it would be possible to include an error term in (6) representing unobservable heterogeneity in reporting styles. We do not do so since there are only three vignette ratings within each domain from which to identify the individual effects. With relatively small samples identification is likely to be weak.

Under the assumption of *response consistency,* the same heterogeneous cut-points govern the observation mechanism that relates latent own-health, H_{di}^*, to the observed categorical response, H_{di}. It is assumed that the error terms in the vignette and own latent health equations, υ_{jdi} and ε_{di} respectively, are independent for all *i, j* and *d*. King and Wand (2007) introduce a nonparametric method of analyzing vignettes data that, while maintaining the basic assumptions of *vignette equivalence* and *response consistency*, avoids the distributional assumptions of the HOPIT model.

The predicted latent own-health scores from the HOPIT model are used to proxy need for curative care in the hurdle model of doctor visits. By both instrumenting reported health with objective indicators and purging reporting heterogeneity using the vignettes, this procedures deals with both sources of bias that are present when the raw self-reports are used to proxy need. Since both biases are expected to be downward, we expect the education effects estimated using this need proxy to be largest. In addition to the exclusion restrictions on the health indicators and the vignette equivalence and response consistency assumptions, identification also requires that the vignette ratings do not directly explain doctor consultations. The latter implies that perceptions of health in general, as opposed to perceptions of own health, do not impact on health care seeking behaviour. While it by no means amounts to a formal test, we found no evidence of any effect of the vignette responses when these were entered into a model of doctor visits.

In order to take account of the sampling variability of the predicted health scores that are used in the hurdle models of doctor visits, we bootstrap the whole procedure, using 50 replications, to obtain standard errors.

4. RESULTS

Estimates of the ordered probit and HOPIT models are not presented. Using the same SHARE data, estimates of education-related disparities in each of the six self-reported health domains with and without adjustment for reporting heterogeneity using the HOPIT model are presented in Bago d'Uva et al (2008). Hurdle models are estimated separately for GP and specialist visits. In addition to need for curative care proxied in three alternative ways from self-rated health, all models include the full set of diagnosed conditions listed in Table 4, as well as age, gender, country dummies and education. Only the education coefficients are presented in Tables 5-8, both from restricted models in which the education effects are held constant across countries and from more flexible models in which they are allowed to vary.

4.1 GENERAL PRACTITIONER VISITS

The top panel of Table 5 gives the education coefficients, constrained to be equal across countries, from logit models of the binary decision to visit a GP. Using the reported category within each health domain, without instrumenting using health indicators or purging reporting heterogeneity using vignettes, to proxy need for curative care (results columns 1-3), individuals with tertiary level education appear significantly less likely than those with no more than primary education to visit a GP. If one were to presume that self-rated health and diagnosed conditions are adequate proxies for need, then this would indicate inequity in access to GP care to the disadvantage of the better educated. However, when self-rated health is instrumented using the objective health indicators (columns 4-6), the magnitude of the coefficient on tertiary education falls and it loses significance. Correcting, in addition, for reporting heterogeneity using predictions of health from HOPIT models identified from the vignettes results in a further large fall in the magnitude and significance of the coefficient on tertiary education. In this specification, the education categories are clearly not jointly significant. The loss of significance of education is not due to increases in the standard errors, which are relatively constant across the specifications. It is the fall in the magnitude of the tertiary education effect that is responsible. The positive coefficients on the secondary education categories increase as more adjustments are made to the reported health variables but they never reach significance. These results are entirely consistent with our predictions that using self-rated health to proxy need will bias estimates away from showing an educational advantage in health care use due both to errors-in-variables and education-correlated reporting errors and that instrumenting reported health using objective health indicators will only partially correct the bias.

The tests at the foot of the table indicate that the restricted model is rejected relative to one that allows education effects to vary by country. For most countries, the results are broadly consistent with those from the restricted model. There is an apparent lower propensity for the higher education groups to consult with a GP when raw self-rated health is used to proxy need, which is eroded as reported health is instrumented and then corrected for reporting heterogeneity. This pattern is most apparent for Belgium, France, Greece and Spain. The Netherlands is clearly an exception. Even using raw self-rated health there is some evidence that higher education groups are more likely than those with primary education to visit the GP. When adjustment is made for reporting heterogeneity, but not so much when self-rated health is instrumented, the pro-higher education advantage becomes much more marked both in magnitude and significance. Sweden also starts off with some indication of an advantage to the better educated. Note that the average probability of visiting a GP is lower in Sweden and the Netherlands than in all other countries but for Greece (Table 1). It could be that the higher educated are less constrained by the more restricted access to GP services. Unlike the Netherlands, the education disparity does not rise in Sweden when adjustment is made for reporting heterogeneity. This is

because the reporting bias is in the opposite direction in Sweden – the higher educated are less likely to report a given condition as representing a health problem (Bago d'Uva et al, 2008).

Results from the truncated negative binomial model of the (positive) number of GP visits are even more consistent with our expectations about the effect of correcting measurement error (Table 6). In the results from the model with no education-country interactions there are no significant education effects when self-rated health is used to proxy need. The coefficients of all three education categories increase when reported health is instrumented using objective health indicators and it emerges that those with upper secondary education visit the GP significantly more than those with primary schooling. Adjusting for reporting heterogeneity leads to large increases in the coefficients and all three become significant. On average across countries, GP utilisation is significantly greater for all education groups in comparison with the least educated. There is clear evidence of inequity in the intensity of GP care that is not evident when reported health categories are used to proxy need.

The tendency for a higher education bias to become more evident as cumulative adjustments are made to self-rated health is apparent for most countries, particularly Belgium, Greece and the Netherlands, although significance is much lower due to the small cell sizes in the models with country-education interactions. In France and Italy the results are similar to those from the logit model – an apparent disparity to the disadvantage of the highest education group is removed once account is taken of the tendency of the highly educated to report their health more negatively. Greece is the country with the greatest education-related inequality in GP visits. It also has one of the lowest mean rates of GP consultation (Table 1), possibly reflecting both the lack of a family doctor service and the very heavy reliance on out-of-pocket payments for health care (OECD, 2010). While primary care is free to those with public health insurance, long waiting times for consultations may lead patients, at least those prepared to pay, to turn to providers who do not accept public insurance patients.

Tables 5 & 6

4.2 SPECIALIST VISITS

Consistent with the findings of previous European studies (Van Doorslaer *et al*, 2000; Van Doorslaer et al, 2004), access to specialist care displays greater bias in favour of higher education groups that does GP care. On average across countries, the probability of seeing a specialist rises significantly and monotonically with education even when self-rated health in used to proxy need (Table 7, top panel, columns 1-3). Instrumenting with health indicators using an ordered probit model of reported health has little impact on the education effects, which remain highly significant (columns 4-6). Adjusting for reporting heterogeneity using the HOPIT

model does increase the coefficients, particularly for the top education category. So, again, there appears to be a marginal impact of using the vignettes to correct for reporting heterogeneity over and above that achieved by exploiting information on objective health indicators.

Unlike for GP care, the likelihood ratio tests do not indicate substantial cross-country variation in the education effects on specialist care (Table 7 bottom row). Indeed, the consistent educational advantage in access to specialist care is striking. Even using raw self-rated health to proxy need, there is a significant education gradient in the probability of accessing specialist care in every country except the Netherlands, which stands out by having a gradient in access to GP, but not specialist, care. This is consistent with GP gatekeepers in the Netherlands being effective in ensuring equitable access to specialist care, possibly withstanding pressure from higher social groups seeking privileged access. See Bago d'Uva and Jones (2009) for further discussion of the role of gatekeepers, and other European health system characteristics, in explaining inequality in the distribution of GP and specialist visits. In general, the country-specific education effects on the probability of using specialist care increase when adjustment is made for reporting heterogeneity.

The impact of education on the number of specialist visits is very different from that on the probability of having at least one visit. Conditional on coming into contact with a specialist, there is little or no evidence that the intensity of consultations varies systematically with education (Table 8). In the restricted model with homogeneous education effects, the education coefficients move in the direction of an education advantage when adjustment is made for measurement error in reported health, but they do not come close to reaching significance. As with the binary specialist decision, homogeneity of the education effects across countries is not clearly and consistently rejected. After adjustment for reporting heterogeneity, but not before, at least one of the education groups use specialist care more intensively than those with primary education only in Belgium, France, Italy and the Netherlands. The lack of significance of the effects in this model could be due to the smaller sample size, given that less than half of respondents use specialist care. But analysis of data from the European Community Household Panel Survey (ECHP) also reveals larger income elasticities in the probability of contacting a specialist than in the conditional number of contacts (Bago d'Uva and Jones, 2009). It is plausible that inequity is at the extensive rather than the intensive margin. Consistent with the hypothesis advanced in the introduction that information disparities may be responsible for the education gradient in health and health care, the lower education groups may lack information on the type of specialist care from which they could benefit and consequently fail to make contact with an appropriate specialist. But once contact is made, the doctor will be largely responsible for the course of treatment and the number of consultations. On this interpretation, the problem is not one of unequal treatment of equal *presented* need but unequal presentation for treatment. That is, the problem is on the demand side rather than the supply side.

Tables 7 & 8 here

16

5. CONCLUSION

Equal treatment for equal need is a founding principle of many European health care systems. Monitoring the performance of systems with respect to this goal is a major challenge. A weakness of previous research has been the failure to identify individuals in equal need using the self-reported health measures typically available in large household surveys. In this paper, we addressed this weakness by using two methods to correct bias in the estimated education gradient in doctor visits conditional on need. The first approach instruments self-rated health with objective health indicators. This deals with bias arising from the fact that self-rated health is an imperfect proxy for medical care need. The second method goes a step further by purging self-rated health of education-related reporting errors using information on reporting styles from the rating of health vignettes.

Analysis of SHARE data on elderly Europeans confirms it is likely that bias is present in studies that use self-rated health as a proxy for need with the goal of estimating inequity in the distribution of doctor visits. Instrumenting self-reported health with objective health indicators generally shifts the distribution of health care conditional on measured need in the direction of inequality favouring more highly educated groups. There is a further, and typically larger, shift the same direction when correction is made for the observed tendency of the more highly educated to rate their health more negatively. On average across countries, while the probability of contacting a General Practitioner conditional on self-rated health is lower for the higher educated, this perverse inequity is no longer apparent when self-rated health is purged of both sources of bias. Using self-rated health to proxy need, there is no apparent inequality in the number of GP visits. But inequity to the advantage of the better educated emerges when corrections are made to reported health. The probability of accessing specialist care displays inequity to the advantage of the better educated even without any correction for measurement error, but the disparity becomes even larger once differential reporting thresholds are taken into account.

These results are a warning against complacency in the equity performance of European health systems. The distribution of primary care is perhaps not as equitable as is often believed and specialist care is even less equitably distributed than has hitherto been realised. But the implications of the results potentially stretch beyond the monitoring of inequity to the understanding of its causes. With universal coverage ensured and relatively low financial barriers to access in most European countries, one may rightly wonder how it can possibly be that equity remains an elusive goal. Our hypothesis is that variation in health expectations, as reflected in the rating of vignettes, at least in part, drives inequality in the utilisation of health care. Given that the better educated report a given level of health more negatively, it seems likely that they will also be quicker to seek health care for any given condition. Viewed in this way, reporting heterogeneity is not simply a nuisance

to be purged from health measures in order to better evaluate equity in health care use, it is potentially an important factor in improving understanding of health seeking behaviour.

REFERENCES

Bago d'Uva, T., Jones, A.M. and van Doorslaer, E. (2009) Measurement of horizontal inequity in health care utilisation using European panel data. *Journal of Health Economics*, 28(2), 280-289.

Bago d'Uva, T., O'Donnell, O. and van Doorslaer, E. (2008) Health reporting by educational level and its impact on the measurement of health inequalities in Europe. *International Journal of Epidemiology,* 37(6), 1375-83.

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M. and O'Donnell, O. (2008) Does reporting heterogeneity bias the measurement of health disparities? *Health Economics,* 17(3), 351-75.

Beam Dowd, J. and Zajacova, A. (2007) Does the predictive power of self-rated health for subsequent mortality risk vary by socioeconomic status in the US? *International Journal of Epidemiology,* 36, 1214–21.

Börsch-Supan, A., Jürges, H. (2005a) The *Survey of Health, Ageing and Retirement in Europe - Methodology*. Mannheim: MEA.

Börsch-Supan, A., Hank, K., Jürges, H. (2005b) A new comprehensive and international view on ageing: introducing the survey of health, ageing and retirement in Europe. *European Journal of Ageing* 2, 245–253.

Bound, J. (1991) Self reported versus objective measures of health in retirement models. *Journal of Human Resources,* 26, 107-137

Cutler, D. M. and Lleras-Muney, A**.** (2008) Education and Health: Evaluating Theories and Evidence. In *Making Americans Healthier: Social and Economic Policy As Health Policy* (eds. R. F. Schoeni, J. S. House, G. A. Kaplan and H. Pollack). New York: Russell Sage Foundation.

Datta Gupta, N., Kristensen, N. and Pozzoli, D. (2010) External validation of the use of vignettes in cross-country health studies. *Economic Modelling*, 27(4), 854-865.

Drèze J. and Sen, A.K. (2002) *India: Development and Participation*. Oxford: Oxford University Press.

Frijters, P., Haisken-DeNew, J.P., Shields, M. (2005) The causal effect of income on health: Evidence from German reunification. *Journal of Health Economics,* 24(5), 997-1017.

Glied, S. and Lleras-Muney, A. (2008) Technological Innovation and Inequality in Health. *Demography* 45(3), 741-761.

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

Gurmu, S. (1998) Generalized hurdle count data regression models. *Economics Letters* 58(3), 263-268.

Huisman, M. van Lenthe, F., Mackenbach, J. (2007) The predictive ability of self-assessed health for mortality in different educational groups. *International Journal of Epidemiology,* 36, 1207–13.

Jurges, H. (2008) Self-assessed health, reference levels and mortality. *Applied Economics,* 40(5), 569-582.

Kapteyn, A., Smith, J., van Soest, A. (2007) Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review,* 97(1), 461–473.

Kapteyn, A., Smith, J., van Soest, A. (2009) Work Disability, Work, and Justification Bias in Europe and the U.S. *IZA Discussion Paper No. 4388.* Bonn: IZA.

King, G., Murray, C. J. L., Salomon, J. and Tandon, A (2004) Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98(1), 184-91.

King, G. and Wand, J. (2007) Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes. *Political Analysis* 15(1), 46-66.

Lleras-Muney, A**.** (2005) The Relationship Between Education and Adult Mortality in the United States. *Review of Economic Studies,* 72(1), 189-221.

Lleras-Muney, A. and Lichtenberg, F. (2005) The Effect Of Education On Medical Technology Adoption: Are The More Educated More Likely To Use New Drugs? *Annales d'Economie et Statistique* 79/80.

Mackenbach, J. P., Stirbu, I., Roskam, A-J.R., Schaap, M.S., Menvielle, G., Leinsalu, M. and Kunst, A.E. (2008) Socioeconomic Inequalities in Health in 22 European Countries. *The New England Journal of Medicine* 358(23), 2468-2481.

Murray, C. J. L. and Chen, L. C. (1992) Understanding Morbidity Change. *Population and Development Review* 18(3): 481-503.

Murray, C.J.L, Ozaltin, E., Tandon, A. and Salomon, J. (2003) Empirical evaluation of the anchoring vignettes approach in health surveys. In Health Systems Performance Assessment: Debates, Methods and Empiricism, (Murray CJL, Evans, DB eds). Geneva: World Health Organization.

O'Donnell O., Propper, C. (1991) Equity and the distribution of UK National Health Service resources. *Journal of Health Economics*, 10, 1-19.

OECD (2010) *OECD Health Data 2010*. Paris: OECD.

Oreopoulos, P. **(**2006) Estimating Average and Local Average Treatment Effects of Education when Compulsory School Laws Really Matter. *American Economic Review*, 96 (1), 152-175.

Prince, M. J., Reischies, F. et al. (1999) Development of the EURO-D scale--a European, Union initiative to compare symptoms of depression in 14 European centres. *The British Journal of Psychiatry,* 174(4), 330-338.

Rantanen, T., Guralnik, J. M., et al. (1999) Midlife Hand Grip Strength as a Predictor of Old Age Disability. *JAMA,* 281(6), 558-560.

Rantanen, T., Volpato, S., et al. (2003) Handgrip Strength and Cause-Specific and Total Mortality in Older Disabled Women: Exploring the Mechanism. *Journal of the American Geriatrics Society,* 51(5): 636-641.

Salomon, J.A., Mathers, C.D., Chatterji, S., Sadana, R., Üstün, T.B., Murray, C.J.L. (2003) Quantifying Individual Levels of Health: Definitions, Concepts, and Measurement Issues. In *Health systems performance assessment: debates, methods and empiricism (*eds. C.J.L. Murray and D.B. Evans). Geneva: World Health Organization.

Sen, A. (2002) Health: perception versus observation. *BMJ* 324(7342), 860-861.

Singh-Manoux, A., Dugravot, A., Shipley, M.J. *et al.* (2007) The association between self-related health and mortality in different educational groups. *International Journal of Epidemiology,* 36, 1222–28.

Smith, J. P. (2007) The Impact of Socioeconomic Status on Health over the Life-Course. *Journal of Human Resources,* XLII(4), 739-764.

Smith, J.P. and Kington, R. (1997) Demographic and economic correlates of health in old age. *Demography* 34 (1), 159-170.

Tandon, A., Murray, C.J.L., Salomon, J.A., King, G. (2003) Statistical models for enhancing cross-population comparability. In *Health systems performance assessment: debates, methods and empiricism (*eds. C.J.L. Murray and D.B. Evans). pp. 727-46. Geneva: World Health Organization.

UNESCO (1997). *International Standard Classification of Education* 1997, Paris: UNESCO.

van Kippersluis, H., O'Donnell, O. and van Doorslaer, E. (2009) Long run returns to education: Does schooling lead to an extended old age? *Tinbergen Institute Discussion Paper* no. 09-037/3, Amsterdam: Tinbergen Institute.

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), 553-583.

Van Doorslaer, E., Koolman, X., Jones, A. (2004) Explaining income-related inequalities in doctor utilisation in Europe. *Health Economics,* 13, 629-647.

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

Wagstaff, A. and van Doorslaer, E. (2000b) Equity in health care finance and delivery. In *Handbook of Health Economics* (eds. A.J. Culyer and J.P. Newhouse) p.p. 1803-1862. Amsterdam: North-Holland.

Winkelmann*,* R*. (*2004) Health Care Reform and the Number of Doctor Visits – An Econometric Analysis*. Journal of Applied Econometrics,* 19*,* 455-472*.*

TABLES

Table 1: Descriptive statistics of General Practitioner and Specialist visits by country

	General Practitioner visits					Specialist visits				
	Mean	Std. Dev.	P(visits>0)[†]	Mean|visits>0[‡]	Max.	Mean	Std. Dev.	P(visits>0)	Mean|visits>0	Max.
Belgium	5.011	5.613	0.887	5.650	53	2.215	4.833	0.544	4.069	54
France	4.886	4.573	0.891	5.482	50	1.551	2.865	0.468	3.312	34
Germany	4.154	4.820	0.870	4.776	49	2.667	5.013	0.575	4.639	60
Greece	3.043	4.571	0.605	5.028	36	1.944	5.407	0.390	4.984	73
Italy	6.662	10.756	0.811	8.214	98	1.446	3.070	0.406	3.565	30
Netherlands	2.145	2.431	0.767	2.797	24	1.573	4.844	0.405	3.884	89
Spain	6.246	9.484	0.845	7.389	90	1.942	6.055	0.440	4.418	98
Sweden	1.655	2.274	0.650	2.547	26	0.971	2.901	0.316	3.075	40
Total	4.211	6.155	0.794	5.307	98	1.805	4.519	0.447	4.034	98

† Proportion with any visit. ‡ Mean visits among those with any visit.

Table 2: Proportion of respondents who classify vignette corresponding to the middle level of difficulty in each domain as having "none" or "mild" difficulty by educational level[†]

ISCED Education Category	Mobility	Cognition	Pain	Sleep	Breathing	Emotional Health
Primary	0.1669	0.3172	0.2398	0.1372	0.1032	0.1821
Lower secondary	0.1503	0.3355	0.2030	0.1248	0.0626	0.1554
Upper secondary	0.1226	0.3235	0.1950	0.0932	0.0668	0.1404
Tertiary	0.1009	0.3118	0.1600	0.1008	0.0465	0.1236

†Cross-country pooled sample. Rates are standardised for age, gender and country.

Table 3: Means of objective health indicators by education and self-rated health[†]

ISCED Education Category

	Primary	Lower Secondary	Upper Secondary	Tertiary	Observations
Mean grip strength					
Self-rated mobility problem					
None	34.7882	35.1748	35.6109	35.5989	2440
Mild	33.3737	33.3920	35.0635	34.3488	913
Moderate	31.4700	31.7039	32.5558	34.4061	515
Severe / Extreme	29.7960	32.7974	29.1409	30.2602	230
Proportion with less than very good orientation in time (score on date recall test)					
Self-rated difficulty concentrating /remembering					
None	0.1066	0.0791	0.0588	0.0936	1834
Mild	0.1430	0.1203	0.0798	0.0884	1445
Moderate	0.2021	0.2072	0.1644	0.0914	653
Severe / Extreme	0.3646	0.2765	0.1599	0.1514	166
Proportion identified as depressed on EURO-D scale (Prince et al, 1999)					
Self-rated problem of sadness/depression					
None	0.1094	0.0946	0.0783	0.0747	2068
Mild	0.2683	0.2682	0.2674	0.2034	1172
Moderate	0.5445	0.5935	0.4744	0.4736	590
Severe / Extreme	0.6662	0.6756	0.6028	0.5775	275
Observations	1402	780	1154	762	4098

†Cross-country pooled sample. Standardised for age, gender and country.

Table 4: Health indicators used to predict latent health within each domain and diagnosed conditions permitted to directly influence doctor visits

Health Domain	Health Indicators
Mobility	i) whether obese ; ii) maximum grip strength; iii) # limitations in Activities of Daily Living; iv) # limitations in mobility, arm function and fine motor function; v) whether has any of the following diagnoses: stroke, arthritis/rheumatism, hip/femoral fracture or Parkinson's disease; vi) whether has any of the following symptoms: swollen legs, falling down, fear of falling down, dizziness, faints or blackouts
Cognition	All five measured tests of *cognitive functioning* performed by SHARE respondents: i) orientation in time (score of date recall test); ii) & iii) immediate and delayed word recall[†]; iv) word-finding and verbal fluency (# animals named within time interval); v) numeracy[‡]
Pain	i) whether bothered by pain in back, knees, hips or other joints; ii) whether takes medication for joint pain or inflammation (d); iii) whether takes medication for other pain or for stomach burn; iv) whether has any of the following diagnoses: cancer[§], stomach, duodenal or peptic ulcer, arthritis or rheumatism, hip or femoral fracture
Sleep	i) whether Euro-D depression score (Prince et al, 1999) indicates trouble with sleeping; ii) whether is bothered by sleeping problems; iii) whether takes medication for sleeping problems; iv) whether obese; v) whether has any of the following diagnoses: asthma, bronchitis, etc., cancer of oral cavity, larynx, pharynx and lung (dropping cases diagnosed longer than 5 years ago)
Breathing	i) whether is bothered by breathlessness; ii) whether bothered by persistent cough; iii) whether has diagnosis of chronic lung disease or asthma; iv) whether takes medication for asthma and/or chronic bronchitis
Emotional Health	i) score on Euro-D depression scale (Prince et al, 1999); ii) whether takes medication for depression

Diagnosed conditions permitted to directly explain doctor visits	
Whether has	i) heart attack or other heart problems; ii) high blood pressure; iii) high blood cholesterol; iv) stroke or cerebral vascular disease; v) diabetes; vi) chronic lung disease; vii) asthma; viii) arthritis or rheumatism; ix) osteoporosis; x) cancer or malignant tumour (except if diagnosed more than 5 years ago); xi) stomach, duodenal or peptic ulcer; xii) Parkinson's disease; xiii) cataracts; xiv) hip or femoral fracture; xv) any other diagnosed condition.

† Respondents are presented orally with 10 common words and asked to remember them. Word recall is tested immediately and after a short delay during which other cognitive tests are performed.

‡ Respondents are asked to solve up to three problems requiring simple mental calculations based on real-life situations. Those who fail the first question are asked an easier one (and given a total score of 1 if they get the second one wrong too or 2 if they get it right). Those who answer the first question correctly are asked two progressively more difficult questions (and given a total score of 3 if they get the first wrong, of 4 if they get only the first right and of 5 if they get both right).

§ Excluding cancer that doesn't cause pain (skin, testicle and thyroid) and also cases diagnosed more than four years previous.

Table 5: Education effects on propensity for any General Practitioner visit – Logit coefficient estimates from first part of hurdle model. Pooled cross-country sample (N=4098).

	Education	Reported health category			Latent health index from ordered probit			Latent health index from HOPIT		
		Coeff.	SE[†]	Joint signif.[‡]	Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.
Restricted model with homogeneous education effects (1)										
	Lower secondary	0.004	0.132	0.049	0.027	0.139	0.103	0.091	0.135	0.319
	Upper secondary	0.015	0.121		0.100	0.143		0.201	0.140	
	Tertiary	-0.287**	0.129		-0.186	0.128		-0.044	0.132	
Unrestricted model with country specific education effects (2)										
Belgium	Lower second.	-0.718*	0.436	0.401	-0.754	0.525	0.783	-0.627	0.550	0.647
	Upper second.	-0.337	0.454		-0.404	0.484		-0.234	0.495	
	Tertiary	-0.508	0.450		-0.437	0.498		-0.221	0.530	
France	Lower second.	-0.480	0.448	0.076	-0.289	0.455	0.076	-0.114	0.447	0.620
	Upper second.	-0.214	0.315		-0.137	0.350		0.060	0.365	
	Tertiary	-0.774**	0.308		-0.663*	0.350		-0.466	0.407	
Italy	Upper second.	-0.120	0.340	0.599	-0.087	0.374	0.237	0.033	0.394	0.488
	Tertiary	0.261	0.332		0.494	0.438		0.583	0.510	
Germany	Upper second.	0.085	0.431	0.201	0.174	0.500	0.842	0.313	0.506	0.622
	Tertiary	-0.485	0.464		-0.315	0.539		-0.109	0.578	
Greece	Lower second.	0.617*	0.339	0.017	0.503	0.427	0.002	0.722	0.623	0.464
	Upper second.	-0.413*	0.213		-0.317	0.233		-0.263	0.244	
	Tertiary	-0.215	0.250		-0.063	0.299		0.012	0.346	
NL	Lower second.	0.435	0.348	0.245	0.436	0.396	0.940	0.700	0.432	0.049
	Upper second.	0.726*	0.389		0.838**	0.401		1.293**	0.477	
	Tertiary	0.221	0.376		0.265	0.335		1.018**	0.482	
Spain	Upper second.	0.536	0.413	0.018	0.624	0.483	0.735	0.581	0.489	0.094
	Tertiary	-0.691**	0.336		-0.631	0.400		-0.567	0.407	
Sweden	Lower second.	0.020	0.334	0.257	-0.025	0.407	0.218	0.077	0.412	0.871
	Upper second.	0.579*	0.318		0.410	0.408		0.298	0.418	
	Tertiary	0.314	0.307		0.149	0.313		0.240	0.330	
Likelihood ration tests of restricted (1) against unrestricted (2) model										
		LR	df	p-value	LR	df	p-value	LR	df	p-value
		36.91	20	0.0120	33.85	20	0.0272	36.33	20	0.0140

Notes: All models include controls for health in six domains, diagnosed chronic conditions, age, gender and country. The reference education category is primary school or less except for Italy, Germany and Spain. The first panel contains estimated education coefficients from a restricted model with no education-country interactions. The second panel contains country-specific education effects from a model with interactions. The ordered probit and HOPIT models used to predict the health indices include the same age, gender, education, country and education-country interactions, plus the health indicators listed in Table 4.

† – Bootstrap standard error. ‡ – p-value from test of joint significance of education categories.

*, **, *** denotes significance at 10%, 5%, 1%.

Table 6: Education effects on positive number of General Practitioner visits – Truncated Negative Binomial 2 coefficient estimates from second part of hurdle model. Pooled cross-country sample (N=3252).

	Education	Models differentiated by means of controlling for need in six health domains								
		Reported health category			*Latent health index from ordered probit*			*Latent health index from HOPIT*		
		Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.
Restricted model with homogeneous education effects (1)										
	Lower secondary	0.058	0.055	0.154	0.102	0.065	0.133	0.149**	0.067	0.019
	Upper secondary	0.044	0.051		0.119**	0.057		0.206***	0.067	
	Tertiary	-0.068	0.059		0.029	0.070		0.141*	0.081	
Unrestricted model with country specific education effects (2)										
Belgium	Lower second.	0.144	0.130	0.569	0.144	0.153	0.783	0.213	0.164	0.316
	Upper second.	0.162	0.129		0.164	0.179		0.329*	0.190	
	Tertiary	0.064	0.131		0.085	0.158		0.228	0.178	
France	Lower second.	-0.187	0.150	0.011	-0.091	0.166	0.076	0.090	0.200	0.300
	Upper second.	-0.038	0.089		0.019	0.118		0.179	0.124	
	Tertiary	-0.353***	0.111		-0.274**	0.112		-0.080	0.146	
Italy	Upper second.	0.065	0.137	0.097	0.251	0.185	0.237	0.281	0.200	0.222
	Tertiary	-0.253*	0.133		-0.059	0.195		-0.057	0.210	
Germany	Upper second.	-0.107	0.131	0.538	-0.051	0.159	0.842	0.036	0.162	0.877
	Tertiary	-0.179	0.163		-0.101	0.173		-0.047	0.205	
Greece	Lower second.	0.301*	0.163	0.019	0.205	0.164	0.002	0.246	0.160	0.003
	Upper second.	0.304**	0.126		0.409***	0.138		0.420***	0.143	
	Tertiary	0.367**	0.151		0.499***	0.156		0.562***	0.174	
NL	Lower second.	0.033	0.185	0.981	-0.006	0.171	0.940	0.104	0.193	0.412
	Upper second.	0.061	0.203		0.095	0.192		0.261	0.226	
	Tertiary	-0.009	0.210		0.028	0.184		0.373	0.238	
Spain	Upper second.	-0.074	0.129	0.783	-0.076	0.141	0.735	-0.077	0.140	0.750
	Tertiary	-0.079	0.155		0.073	0.221		0.070	0.231	
Sweden	Lower second.	0.471**	0.226	0.205	0.653*	0.343	0.218	0.645*	0.355	0.337
	Upper second.	0.239	0.197		0.287*	0.165		0.155	0.183	
	Tertiary	0.224	0.216		0.278	0.220		0.149	0.232	
Likelihood ration tests of restricted (1) against unrestricted (2) model										
		LR	df	p-value	LR	df	p-value	LR	df	p-value
		31.67	20	0.0470	35.24	20	0.0189	35.11	20	0.0195

Notes: As Table 5

Table 7: Education effects on propensity for any Specialist visit – Logit coefficient estimates from first part of hurdle model. Pooled cross-country sample (N=4098).

	Education	Models differentiated by means of controlling for need in six health domains								
		Reported health category			_Latent health index from ordered probit_			_Latent health index from HOPIT_		
		Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.
Restricted model with homogeneous education effects (1)										
	Lower secondary	0.387***	0.107	0.000	0.368***	0.115	0.000	0.414***	0.118	0.000
	Upper secondary	0.521***	0.099		0.497***	0.103		0.554***	0.106	
	Tertiary	0.779***	0.109		0.782***	0.131		0.910***	0.145	
Unrestricted model with country specific education effects (2)										
Belgium	Lower second.	0.327	0.268	0.103	0.315	0.287	0.081	0.405	0.293	0.033
	Upper second.	0.477*	0.268		0.481	0.318		0.579*	0.328	
	Tertiary	0.656**	0.274		0.708**	0.290		0.829***	0.295	
France	Lower second.	0.421	0.296	0.010	0.429	0.295	0.037	0.513*	0.306	0.033
	Upper second.	0.198	0.186		0.172	0.218		0.261	0.231	
	Tertiary	0.697***	0.212		0.690***	0.239		0.755***	0.257	
Italy	Upper second.	0.564**	0.288	0.000	0.556**	0.270	0.001	0.609**	0.284	0.002
	Tertiary	1.072***	0.269		1.147***	0.302		1.192***	0.347	
Germany	Upper second.	0.359	0.270	0.060	0.334	0.272	0.044	0.387	0.275	0.049
	Tertiary	0.758**	0.322		0.784**	0.314		0.872**	0.358	
Greece	Lower second.	-0.318	0.349	0.000	-0.383	0.305	0.000	-0.207	0.335	0.000
	Upper second.	0.881***	0.209		0.866***	0.195		0.888***	0.205	
	Tertiary	1.050***	0.245		1.127***	0.235		1.125***	0.261	
NL	Lower second.	0.011	0.310	0.993	0.030	0.359	0.994	0.139	0.391	0.848
	Upper second.	-0.065	0.343		-0.043	0.373		0.164	0.392	
	Tertiary	-0.016	0.343		0.008	0.382		0.417	0.520	
Spain	Upper second.	0.723***	0.274	0.017	0.753***	0.237	0.006	0.706***	0.247	0.016
	Tertiary	0.518*	0.297		0.529	0.391		0.498	0.429	
Sweden	Lower second.	0.469	0.354	0.057	0.409	0.404	0.063	0.455	0.418	0.033
	Upper second.	0.509	0.321		0.440	0.318		0.361	0.333	
	Tertiary	0.856***	0.316		0.811***	0.301		0.938***	0.319	
Likelihood ration tests of restricted (1) against unrestricted (2) model										
		LR	df	p-value	LR	df	p-value	LR	df	p-value
		27.97	20	0.1102	30.84	20	0.0573	21.88	20	0.3468

Notes: As Table 5.

Table 8: Education effects on positive number of Specialist visits – Truncated Negative Binomial coefficient estimates from second part of hurdle model. Pooled cross-country sample (N=1833).

		Models differentiated by means of controlling for need in six health domains								
		Reported health category			*Latent health index from ordered probit*			*Latent health index from HOPIT*		
	Education	Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.	Coeff.	SE	Joint signif.
Restricted model with homogeneous education effects (1)										
	Lower secondary	0.177[*]	0.107	0.121	0.149	0.128	0.484	0.180	0.127	0.541
	Upper secondary	-0.067	0.097		-0.026	0.110		0.050	0.114	
	Tertiary	-0.024	0.105		-0.010	0.119		0.103	0.135	
Unrestricted model with country specific education effects (2)										
Belgium	Lower second.	0.312	0.250	0.428	0.361	0.280	0.531	0.551[*]	0.284	0.270
	Upper second.	-0.044	0.248		0.121	0.267		0.305	0.239	
	Tertiary	0.074	0.247		0.126	0.252		0.378	0.264	
France	Lower second.	0.295	0.285	0.273	0.380	0.421	0.266	0.554	0.430	0.114
	Upper second.	0.138	0.184		0.160	0.221		0.379[*]	0.216	
	Tertiary	-0.208	0.201		-0.194	0.208		0.030	0.244	
Italy	Upper second.	0.286	0.296	0.624	0.383	0.238	0.269	0.482[*]	0.264	0.169
	Tertiary	0.056	0.264		0.312	0.378		0.356	0.401	
Germany	Upper second.	-0.212	0.235	0.500	-0.275	0.236	0.504	-0.152	0.243	0.719
	Tertiary	-0.017	0.270		-0.123	0.293		0.063	0.318	
Greece	Lower second.	-0.163	0.400	0.354	-0.460	0.374	0.103	-0.295	0.439	0.179
	Upper second.	-0.028	0.206		-0.178	0.256		-0.202	0.269	
	Tertiary	0.348	0.237		0.287	0.249		0.261	0.249	
NL	Lower second.	0.410	0.312	0.105	0.479	0.371	0.451	0.782[**]	0.400	0.234
	Upper second.	-0.159	0.358		0.163	0.411		0.596	0.454	
	Tertiary	-0.068	0.351		0.076	0.375		0.878[*]	0.532	
Spain	Upper second.	-0.256	0.251	0.174	-0.546	0.402	0.339	-0.605	0.392	0.296
	Tertiary	-0.523[*]	0.295		-0.732	0.540		-0.654	0.567	
Sweden	Lower second.	-0.310	0.417	0.439	-0.252	0.438	0.801	-0.215	0.427	0.889
	Upper second.	0.041	0.354		0.119	0.324		-0.002	0.334	
	Tertiary	-0.501	0.366		-0.232	0.314		-0.246	0.334	
Likelihood ration tests of restricted (1) against unrestricted (2) model										
		LR	df	p-value	LR	df	p-value	LR	df	p-value
		20.46	20	0.4295	27.05	20	0.1339	32.05	20	0.0428

Notes: As Table 5.