HEALTH ECONOMICS
Health Econ. **17**: 351–375 (2008)
Published online 13 August 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hec.1269

DOES REPORTING HETEROGENEITY BIAS THE MEASUREMENT OF HEALTH DISPARITIES?

TERESA BAGO D'UVA[a,d,e,*], EDDY VAN DOORSLAER[a,d,e], MAARTEN LINDEBOOM[b,d,e,f] and OWEN O'DONNELL[c,d]

[a] *Erasmus University Rotterdam, The Netherlands*
[b] *Free University of Amsterdam, The Netherlands*
[c] *University of Macedonia, Greece*
[d] *Netspar, The Netherlands*
[e] *Tinbergen Institute, The Netherlands*
[f] *HEB, Norway, IZA, Germany*

SUMMARY

Heterogeneity in reporting of health by socio-economic and demographic characteristics potentially biases the measurement of health disparities. We use anchoring vignettes to identify socio-demographic differences in the reporting of health in Indonesia, India and China. Homogeneous reporting by socio-demographic group is rejected and correcting for reporting heterogeneity tends to reduce slightly estimated disparities in health by education (not China) and to increase those by income. But the method does not reveal substantial reporting bias in measures of health disparities. Copyright © 2007 John Wiley & Sons, Ltd.

Received 31 August 2006; Revised 3 May 2007; Accepted 7 June 2007

KEY WORDS: health measurement; vignettes; self-reported health; reporting heterogeneity

JEL classification: D30; D31; I10; I12

INTRODUCTION

Self-reported health is a convenient and informative instrument, widely used in analyses of health inequality and the determinants of health, as well as in the examination of the economic consequences of ill health. It is a simple categorical measure of perceived health status that can be obtained from large-scale household surveys, which also provide information on socio-economic characteristics against which to assess health inequality. Self-reported health has been shown to be a powerful predictor of mortality (Idler and Kasl, 1995; Idler and Benyamini, 1997) and of medical care use (see, e.g., Van Doorslaer *et al*., 2000, 2004). Numerous studies have analysed the relationship between self-reported general health and socio-economic status (e.g. Ettner, 1996; Deaton and Paxson, 1998; Smith, 1999; Benzeval *et al*., 2000; Adams *et al*., 2003; Frijters *et al*., 2005), while others have used self-reported health status in the measurement of socio-economic inequality in health (see, e.g., Van Doorslaer *et al*., 1997; Van Ourti, 2003; Van Doorslaer and Koolman, 2004), which in turn has been shown to predict socio-economic inequality in mortality (Van Doorslaer and Gerdtham, 2003).

However, there is inevitably heterogeneity in the reporting of health. For a given true but unobserved health state, individuals will report health differently depending upon conceptions of health in general, expectations for own health, financial incentives to report ill health and comprehension of the survey questions. In many contexts, reporting heterogeneity need not be a major concern provided that it is

*Correspondence to: Department of Applied Economics (Room H13.09), Erasmus School of Economics, PB 1738, 3000 DR Rotterdam, The Netherlands. E-mail: bagoduva@few.eur.nl

random. Systematic differences in reporting behaviour are more problematic. Of primary concern in this paper is that measurement of health inequality will be biased if there are systematic differences in the way in which health is reported across the socio-economic characteristic, e.g. income or education, against which inequality is being assessed and/or across demographics that are correlated with this characteristic. The purpose of this paper is to test and correct for reporting bias in measures of health disparities in developing countries.

Differences in health disparities derived from self-reported and more objective indicators are suggestive of systematic variation in reporting behaviour. One frequently cited example is the tendency for Aboriginals to report better health than the general Australian population despite being seriously disadvantaged according to more objective health indicators, such as mortality (Mathers and Douglas, 1998). Discrepancy in health gradients measured by objective and subjective indicators is even more common in evidence from the developing world. In India, the state of Kerala consistently shows the highest rates of reported morbidity, in spite of having the lowest rates of infant and child mortality (Murray, 1996). Wagstaff (2002) notes that income-related inequalities in objective indicators of ill health, such as malnutrition and mortality, tend to be higher than those in subjective health. Moreover, the use of subjective health measures has led to some improbable health gradients in developing countries, with the rich reporting worse health than the poor (Baker and Van der Gaag, 1993), which seems quite inconsistent with substantial pro-rich inequality in infant and child mortality rate and in anthropometric indicators (Gwatkin *et al.*, 2000). Sen (2002) argues: 'There is a strong need for scrutinising statistics on self-perception of illness in a social context by taking note of levels of education, availability of medical facilities and public information on illness and remedy'.

Formal testing of reporting heterogeneity by demographic and socio-economic status has been undertaken in recent studies, albeit not in an exhaustive way, and not for less developed countries. Van Doorslaer and Gerdtham (2003) use Swedish data to assess the extent to which the capacity of self-reported health to predict mortality varies across socio-demographic groups. Self-reported health is found to be a very strong predictor of subsequent mortality risk. The relationship varies with demographic and disease characteristics but not by socio-economic status. Lindeboom and Van Doorslaer (2004) assume that the McMaster Health Utility Index provides an objective and comprehensive health indicator and test whether, conditional on this, there is variation in stated health in Canada that can be attributed to reporting behaviour. The results are consistent with those of Van Doorslaer and Gerdtham; there is evidence of reporting heterogeneity for age and sex, but not for education and income. Etilé and Milcent (2006), using French data, also test for reporting heterogeneity by examining the relationship between income and self-assessed general health conditional on a battery of reported, but putatively more objective, health indicators. They find reporting heterogeneity accounts for more of the correlation between income and reported health at higher incomes and at middle/high levels of health. They suggest that a dichotomous indicator of poor/non-poor health may avoid much of the reporting bias in income-related health inequality.[1] While this evidence is somewhat encouraging for the measurement of socio-economic inequalities in health in developed countries, it says nothing about the effect of reporting heterogeneity on the measurement of health inequality in developing countries, where differences in conceptions of illness by education and income levels and between urban and rural locations may be greater.

The studies discussed in the previous paragraph test for reporting heterogeneity through examination of variation in health reporting conditional on some 'objective' measure of health. One problem is that

[1] Jürges (2007) tests for reporting heterogeneity across European countries by conditioning on diagnosed conditions and measured health indicators. There is also a substantial literature that examines incentives to report poor health deriving from entitlements for disability transfers and justification of non-employment (Stern, 1989; Bound, 1991; Kerkhofs and Lindeboom, 1995; Benitez-Silva *et al.*, 1999; Kreider, 1999; Disney *et al.*, 2006). This literature is concerned with bias created by health reporting heterogeneity in models of employment and retirement rather than with the measurement of inequality in health that is the primary motivation for the present paper.

DOI: 10.1002/hec

objective indicators, for example, mortality, may not be available. Less objective indicators, such as health conditions, are more likely to be available but are also self-reported and are subject to error (Baker *et al*., 2004). The test might uncover different types of reporting heterogeneity in different indicators rather than deviations from a purely objective benchmark of health. A further disadvantage of using 'objective' indicators to test and correct for reporting heterogeneity is that this strips out any socio-economic related variation in self-reported health conditional on the objective indicators. If the self-reported health contains information on true health, conditional on objective indicators, then this is lost. If self-reported health does not contain additional information, then one might as well examine the relationship between 'objective' indicators and socio-economic characteristics from the outset.

Rather than attempt to identify reporting behaviour from variation in self-reported health beyond that explained by 'objective' indicators, an alternative is to examine variation in the evaluation of given health states represented by hypothetical case vignettes (Tandon *et al*., 2003; King *et al*., 2004; Salomon *et al*., 2004).[2] The vignettes represent fixed levels of latent health and so all variation in the rating of them can arguably be attributed to reporting behaviour, which can be examined in relation to observed characteristics. Under the assumption that individuals rate the vignettes in the same way as they rate their own health, it is possible to identify a measure of health that is purged of reporting heterogeneity. Murray *et al*. (2003) evaluate this approach to the measurement of health, in the domain of mobility, using data from 55 countries covered by World Health Organisation (WHO) surveys. The principal objective of their analysis is to obtain comparable measures of population health that are purged of cross-country differences in the reporting of health.[3] Besides country, reporting of health is allowed to vary with age, sex and education, but there is no detailed examination of these dimensions of reporting heterogeneity or of the impact on measured health disparities. Using the vignettes method, Kapteyn *et al*. (forthcoming) find that about half of the difference in rates of self-reported work disability between the Netherlands and the US can be attributed to reporting behaviour.

Our concern in this paper is not with the cross-country comparability of health measures but with the comparability of self-reported health across demographic and socio-economic groups *within* a country and the consequences of any systematic differences in reporting behaviour for measures of health disparities between socio-economic groups. We are primarily interested in this issue in the context of the developing world since, as mentioned above, that is where the reporting heterogeneity problem is considered to be most severe. We apply the vignettes methodology to data from the three largest Asian countries – Indonesia, India and China – in order to test for systematic differences in reporting of health by sex, age, urban/rural location, education and income and to establish the extent to which estimated disparities in health by income and education change when reporting differences are purged from the health measures. In subsequent sections of the paper, the data, econometric models, results and conclusions are presented.

DATA – WHO MULTI-COUNTRY SURVEY

The data used in this paper, as in Murray *et al*. (2003), are from the WHO Multi-Country Survey Study on Health and Responsiveness 2000–2001 (WHO-MCS) that covered 71 adult populations in 61 countries. Üstün *et al*. (2003) provide a comprehensive report on the goals, design, instrument development and execution of this survey. The main goal of the WHO-MCS was the development of

[2] The vignette methodology was developed as a tool to permit valid cross-country comparisons of concepts elicited through self-reported categorical variables, such as health. The approach attempts to overcome the limitations of other approaches proposed previously to improve the comparability of self-reported data across countries (a summary of these can be found, e.g., in Iburg *et al*., 2002).

[3] Besides the main objective of comparing the health of different populations across countries, the WHO Multi-Country Survey Study also includes vignettes for several aspects of the health system responsiveness, which are intended to enable comparison of responsiveness across systems.

DOI: 10.1002/hec

'valid, reliable, and comparable instruments to describe individual health states and health system responsiveness on a core set of domains' (Üstün *et al.*, 2003, p. 764). Samples were drawn from up-to-date registries of all residents in each country, when available, or otherwise from registries providing postal coverage or post office listings. The data we use come from face-to-face household surveys, implemented in multistage stratified probability samples of between 5000 and 10 000 adults aged 18 years or above, non-institutionalised and living in private households (with a response rate of 84%, on average). In each stage of the sampling design, the WHO used probability methods to ensure that the resulting samples are representative of the target population. One respondent was randomly selected from the eligible individuals within each household.

Individuals were asked to report their health in each of six health domains (mobility, cognitive functioning, affective behaviour, pain or discomfort, self-care and usual activities). In addition, a sub-sample of individuals was asked to rate a set of anchoring vignettes describing fixed ability levels on each health domain. The general idea is to use the responses to these vignettes to identify reporting heterogeneity. Assessments of own health by domain can then be calibrated against the vignettes, purging reporting heterogeneity and giving interpersonally comparable ordinal health measures. The vignette descriptions are brief, simple and written in a culturally sensitive way. In spite of the difficulties involved, reliability tests for vignette ratings have shown overall good repeatability, which suggests that individuals generally understand the exercise and are thus able to rate the vignettes in a meaningful way (Üstün *et al.*, 2003). Reliability tests for the own health variables by domain have also returned satisfactory results (*ibid*).

We use the WHO-MCS data for Indonesia (excluding Papua, Aceh and Maluku), an Indian state (Andhra Pradesh) and three Chinese provinces (Gansu, Henan and Shan-dong).[4] The data set used here results from dropping individuals with missing data on own health, the socio-demographic variables used in the analysis and the vignettes. The resulting data set contains 7770 observations for Indonesia, 5129 for India and 7156 for China. Table AI in Appendix A documents the number of observations lost due to item non-response, showing that income is the covariate that contributes the most for the loss of observations, especially in the Indonesian and Chinese samples. We have checked and confirmed the robustness of the results for the remaining covariates to the exclusion of income from the analysis.[5]

Health variables: own health and vignettes

Health by domain is obtained from the questions: 'Overall in the last 30 days, how much...':

- difficulty did you have with moving around? (mobility);
- difficulty did you have with concentrating or remembering things? (cognition);
- pain or discomfort did you have? (pain);
- difficulty did you have with self-care, such as washing or dressing yourself? (self-care);
- difficulty did you have with work or household activities? (usual);
- distress, sadness or worry did you experience? (affect).

The five response categories are: 'extreme/cannot do', 'severe', 'moderate', 'mild' and 'none'.

For each domain, a random sub-sample of individuals is presented with a set of vignettes, describing levels of difficulty on that domain, and asked to evaluate these hypothetical cases in the same way as they evaluate their own health for that domain (i.e. using the same five response categories). Of course, there can be no reference to the experience of the vignettes over the last 30 days. One-half of the samples evaluate the vignettes in the domain of *mobility* and roughly one-quarter of the samples respond to the

[4] The Indonesian provinces were excluded from the sampling frame due to political and economic difficulties. Given the sizes of India and China and the multiplicity of languages, the surveys were limited to certain states/provinces.

[5] The direction of the correction for reporting bias in the health equation (see later section) only changes in five cases (among all covariates that enter the health equation significantly, all domains and both countries).

DOI: 10.1002/hec

vignettes in each of the other domains. Each respondent is asked to rate vignettes on two domains. Within a given domain, the set of vignettes is the same for all respondents. The vignette descriptions for all the domains are presented in Appendix B, and the distributions of the vignette evaluations can be found in Table AII in Appendix A.

Despite representing fixed levels of ability by domain, the vignette ratings show considerable variation, which can be attributed to reporting heterogeneity. For example, vignette 4 in the *mobility* domain describes a person who has chest pains and gets breathless after walking up to 200 m but is able to do so without assistance. In Indonesia, almost 35% of respondents categorise this as a moderate mobility problem, but 36% define it as severe and almost 19% as mild. There is even 2.8% with sufficiently high health expectations such that they consider this an extreme mobility problem. On the other hand, 7% do not consider this a problem at all. This is the type of variation we exploit to test for systematic reporting heterogeneity in relation to demographic and socio-economic characteristics.

Socio-demographic variables

Measurement of socio-economic disparities in health will be biased if the reporting of health varies directly with the socio-economic characteristic and/or with demographic factors correlated with that characteristic. Expectations for health and tolerance of illness may be influenced by an individual's socio-economic environment and demographic characteristics. The degree of functioning considered as good health might be expected to decline with age. Conceptions of good health may also differ by sex, although it is more difficult to predict the sign of the effect, which might differ across different health domains. Geographic and economic circumstances may mould health expectations through peer effects and access to medical care. Living within a community in which a large proportion of the population suffers poor health may lower the individual's expectations for her own health. Improved access to effective health care may lower tolerance of illness and disease. Reporting of health may vary with education not only because education acts as a proxy for permanent income but also through a direct effect. The latter will operate through conceptions of illness, understanding of disease and knowledge of the availability and effectiveness of health care. It is not immediately clear in which direction such effects will shift the reporting of health. One might expect the better educated to be less tolerant of poor health. On the other hand, the better educated should be better informed of the health of others and able to appreciate their relatively privileged position in the health distribution.

We test for reporting heterogeneity in relation to age, sex, urban/rural status, education and income. Age is represented by categories: 18–29 years (reference category), 30–44 (AGE3044), 45 and 59 (AGE4559) and more than 60 (AGE60). Sex is represented by the dummy variable FEMALE and location by the dummy URBAN. A flexible education effect is allowed for through a series of dummies indicating the highest level of education completed: less than primary (reference category), primary (EDUC2), secondary (EDUC3) and high school or above (EDUC4).[6] The variable log(INCOME) is the log of monthly household earnings by equivalent adult (in national currencies).[7,8] Table I presents descriptive statistics for the covariates by country.

[6] Dummies for education categories capture non-linearity in the relationship. Experimentation with years of education, which is highly right-skewed, gave broadly similar results apart from a rather implausible negative effect on 'true' health for some domains for China. Experimentation also revealed that the education effect on health for China was weakened if URBAN was excluded.

[7] The respondent has the option to report income for alternative reference periods. We used weekly household income (multiplied by 30.5/7), when available. When information on weekly income was not available, the information on monthly income was used. In the absence of either information on weekly or monthly income, we used annual income divided by 12. Finally, the resulting variable was divided by an equivalence scale (calculated as (number of adults in household $+ 0.5 \times$ number of children in household)$^{0.75}$).

[8] We experimented with an alternative specification in which income was entered through dummies for income quintiles. In general, the results were consistent with those obtained using log(INCOME). The latter is preferred for ease of presentation.

DOI: 10.1002/hec

Table I. Descriptive statistics of covariates

	Indonesia		India		China	
Variables	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
FEMALE	0.539	0.498	0.534	0.499	0.460	0.498
AGE3044	0.405	0.491	0.368	0.482	0.377	0.485
AGE4559	0.198	0.399	0.222	0.416	0.267	0.442
AGE60	0.120	0.325	0.151	0.358	0.111	0.315
EDUC2	0.191	0.393	0.094	0.291	0.154	0.361
EDUC3	0.203	0.402	0.077	0.267	0.331	0.471
EDUC4	0.134	0.340	0.207	0.405	0.416	0.493
Log(INCOME)	12.058	1.338	6.242	1.161	5.387	1.431
URBAN	0.481	0.500	0.337	0.473	0.366	0.482
N	7770		5129		7156	

ECONOMETRIC MODELS

Categorical data on health are typically modelled by assuming that the observed categorical variable is a discrete representation of an underlying unobserved true level of health, measured on a continuous scale. The categorical variable is defined as the result of a mapping between latent health and the response categories. Homogeneous reporting behaviour corresponds to the assumption that the mapping is constant across individuals. By contrast, reporting heterogeneity translates into different mappings between the latent variable and the observed categorical variable. Individuals might attach very different meanings to the labels used for each of the response categories, thus making the observed health variables incomparable, since they do not correspond to the same intervals in the latent health scale. After presenting the homogeneous case, we describe in detail below how vignette information can be used to identify reporting heterogeneity in self-reported health.

Ordered probit: homogeneous reporting behaviour

Let $y_i,\ i = 1, \ldots, N$, be a self-reported categorical health measure. It is assumed that y_i is generated by the latent health variable Y_i^*, specified as

$$Y_i^* = Z_i\beta + \varepsilon_i, \quad \varepsilon_i | Z_i \sim N(0,1) \tag{1}$$

where Z_i is a vector of covariates. Since the latent variable is unobserved and its observed counterpart is categorical, the variance of the error term, ε_i, conditional on Z_i, and the constant term are not identified and are usually set to 1 and 0, respectively.[9] The observed categorical response of individual i, y_i, relates to latent health in the following way:

$$y_i = k \ \Leftrightarrow \ \tau^{k-1} \leq Y_i^* < \tau^k \tag{2}$$

$k = 0, \ldots, K$, $\tau^0 < \tau^1 < \cdots < \tau^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. The parameters τ^k, $k = 1, \ldots, K-1$, are estimated along with the other parameters of the model (β). The assumption of homogeneous reporting that is inherent to the ordered probit model arises from the constant cut-points τ^k. If this assumption does not hold, in particular, if the cut-points vary according to some of the covariates Z_i, then imposing the restriction will lead to biased estimates of β, since they will reflect both health effects and reporting effects.

It is possible to generalise the ordered probit model and allow the cut-points to depend on covariates, $\tau_i^k = \tau(x_i\alpha^k)$, (Terza, 1985). Normalising one threshold to a constant, the other threshold parameters are identified in the sense of showing how covariates shift the thresholds relative to their impact on the

[9] In this application, we fix these terms in a different way. See below.

DOI: 10.1002/hec

baseline threshold. If the covariate effect were the same on all thresholds, labelled 'parallel shift' (Hernandez-Quevedo *et al*., 2004), then the threshold coefficients would be zero and it is not possible to distinguish this case from an effect on the index function alone. Less attractively, identification could be achieved through a series of maintained assumptions that each covariate can be excluded from either the threshold or health index function (Pudney and Shields, 2000). While the generalised ordered probit allows thresholds to vary with covariates, in the present context it would be hazardous to interpret such effects as a reflection of reporting heterogeneity rather than heterogeneity in the latent health index itself (Hernandez-Quevedo *et al*., 2004). As specified in (1), it is assumed that there is a single latent health index that applies for all individuals. It is possible, however, that the relationship of true health with the covariates varies with the level of health itself. For example, income may have a weaker marginal impact on health at better levels of health and this may result in variation of the income coefficient across the categories of reported health. Interpretation of the varying thresholds of a generalised probit model as an indication of reporting heterogeneity would, therefore, rely strongly on the assumption that the latent health index was correctly specified as a homogeneous function of covariates.[10] With additional information provided by vignettes, it is possible to identify the separate effects of covariates on reporting behaviour and true health without relying on functional form and/or exclusion restrictions.

Hierarchical ordered probit: heterogeneous reporting behaviour

Suppose one has access to individuals' self-reports y_{ij} on specific health domains j and vignette ratings on these same domains y_{ij}^{v}. The vignettes describe the level of ability on each domain and ask individuals to rate these hypothetical cases in the same way as they evaluate their own health for that domain (i.e. using the same response scale). The health status of the hypothetical individual is exogenously varied across the vignettes and, therefore, individual variation in responses to these vignettes must be due to reporting heterogeneity. In the context of the generalised ordered probit, this means that we can use the external vignette information to separately identify the thresholds ($\tau_i^k = \tau(x_i\alpha^k),\ k = 1, \ldots, K-1$). These cut-offs can be imposed on the model for the self-reports with respect to the individual's own health, so that estimates of β now reflect true health differences rather than a mixture of health differences and reporting heterogeneity. This has been suggested by King *et al*. (2004), who label their model the hierarchical ordered probit (HOPIT).

The HOPIT model is specified in two parts: one reflecting reporting behaviour and another representing the relationship between the individual's own health and the observables. The use of vignettes to identify the cut-points and so systematic reporting heterogeneity relies on two assumptions. First, there must be *vignette equivalence*: 'the level of the variable represented by any one vignette is perceived by all respondents in the same way and on the same unidimensional scale, apart from random measurement error' (King *et al*., 2004, p. 194). If this did not hold, then one could not interpret variation in responses to a given vignette as reflecting differences in evaluations of a given level of functioning within a unidimensional health domain. Murray *et al*. (2003) conducted a partial test of vignette equivalence in the WHO-MCS data pooled across all countries, including the Chinese, Indian and Indonesian samples on which the present analysis is based, by checking whether there are systematic differences in the ranking of vignettes in relation to age, sex, education and questionnaire characteristics. They find statistically significant but small differences in rankings by each of these factors. Significance may simply be attributable to a sample size of almost half a million. With the exceptions of *mobility* and *pain*, the most highly educated display greatest consistency with the average rankings of the vignettes within each domain, but the differences are not large. It is greatest for *self-care*, but even in this case the difference between the highest and lowest education groups in their correlation coefficients of vignette rankings against the average ranking is only 0.05 (Murray *et al*., 2003,

[10] We are grateful to Andrew Jones, who pointed this to us in a private communication.

DOI: 10.1002/hec

Table 30.7). The elderly display greater consistency with the average rankings, but their rank correlation coefficient is only 0.02 greater than that of the youngest age group (Murray *et al.*, 2003, Table 30.8). Collectively, age, sex, education and questionnaire characteristics explain less than 0.1% of the sample variation in rankings (Murray *et al.*, 2003, Table 30.8), suggesting that almost all inconsistencies in rankings are attributable to random measurement error, which is permitted under the vignette equivalence assumption. Of course, more randomness in the vignette ratings means less information from which to identify reporting behaviour.

The second assumption necessary for identification, via vignettes, of the part of reported own health that is attributable to reporting heterogeneity is *response consistency*: individuals classify all the hypothetical cases represented by the vignettes in the same way as they rate their own health. That is, the mapping used to translate the perceived latent health of others to reported categories is the same as that governing the correspondence between own latent and reported health. This is essential if we are to learn about how individuals report their own health from how they rate others' health. The assumption is not indisputable. Strategic behaviour might influence reporting of own health but not that of others. For example, entitlement rules for disability transfers provide an incentive to understate own health but are irrelevant to the reporting on others' health. But such incentives are not present in the low-income countries studied here, where disability insurance programmes are not developed.

Note that vignette equivalence allows identification of systematic variation in the ordinal reporting of a given level of function, while response consistency allows the variation to be used to purge reporting differences from ordinal evaluations of own health.

Reporting behaviour. The first (vignette) component of the HOPIT uses information on the vignette ratings to model the cut-points as functions of covariates. For a given health domain, let Y_{ij}^{v*} be the latent health level of vignette j as perceived by individual i. Given that each vignette j is assumed, by vignette equivalence, to represent a fixed level of functioning perceptions of which vary only randomly, any association between the perceived latent level of health Y_{ij}^{v*} and individual characteristics is ruled out. $E[Y_{ij}^{v*}]$ is, therefore, assumed to depend solely on the corresponding vignette. Formally, it is assumed that Y_{ij}^{v*} is determined by

$$Y_{ij}^{v*} = \alpha_j + \varepsilon_{ij}^{v}, \quad \varepsilon_{ij}^{v} \sim N(0,1) \tag{3}$$

where the stochastic component reflects random measurement error in perceptions of the vignette level of functioning. The observed vignette ratings y_{ij}^{v} relate to Y_{ij}^{v*} in the following way:

$$y_{ij}^{v} = k \;\Leftrightarrow\; \tau_i^{k-1} \le Y_{ij}^{v*} < \tau_i^{k} \tag{4}$$

$k = 0, \ldots, K$, $\tau^0 < \tau_i^1 < \cdots < \tau_i^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. The cut-points are defined as functions of covariates but are constrained, due to the response consistency assumption, not to vary across different vignettes j for a given health domain, for instance,

$$\tau_i^k = X_i \gamma^k \tag{5}$$

Note that the individual's characteristics are included only in the cut-points, reflecting the assumption that all the systematic variation in the vignette ratings can be attributed to individual reporting behaviour, which follows from vignette equivalence.[11]

[11] Since each individual rates a number of vignettes within a given domain, it would be possible to allow for unobservable individual heterogeneity in the vignette ratings. To gauge the potential importance of this, we compared results from ordered probit and random effects ordered probit models of vignette responses within the mobility domain, with parallel cut-point shift imposed for computational feasibility. The results were very similar and hence, given the substantially greater computational cost, we decided not to allow for unobservable heterogeneity within the full HOPIT model.

DOI: 10.1002/hec

Health equation. Similar to the ordered probit, the second component of the HOPIT defines the latent level of individual own health, Y_i^{s*}, and the observation mechanism that relates this latent variable to the observed categorical variable, y_i. The difference is that the cut-points are no longer constant parameters but can vary across individuals, being determined by the vignette component of the model. Identification derives from the vignette equivalence and response consistency assumptions. The possibility of fixing the cut-points leads to the specification of the model for individual own health as an interval regression, enabling the identification of the constant term and the variance. The latent level of individual own health is specified as

$$Y_i^{s*} = Z_i\beta + \varepsilon_i^s, \quad \varepsilon_i^s|Z_i \sim N(0,\sigma^2) \tag{6}$$

where Z_i is a vector of covariates including a constant. The observed categorical variable y_i is determined by

$$y_i = k \;\Leftrightarrow\; \tau_i^{k-1} \le Y_i^{s*} < \tau_i^k \tag{7}$$

$k = 0,\ldots,K$, $\tau^0 < \tau_i^1 < \cdots < \tau_i^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$ and where τ_i^k are as defined as in (5).

It is assumed that the error terms in the vignette and own latent health equations, ε_{ij}^v and ε_i^s, respectively, are independent for all $i = 1,\ldots,N$ and $j = 1,\ldots,V$. The likelihood function depends on the probabilities of observing particular vignette responses and the probability of a particular own health category being reported. Although the errors in the two components of the model are assumed independent, the likelihood does not factorise into two independent parts, since the two components of the model are linked through parameter restrictions. The vignette component identifies the threshold parameters, which are imposed in the estimation of the latent health function.

Test of homogeneous reporting behaviour. This framework offers the possibility of testing for heterogeneous reporting behaviour in relation to individual characteristics. This is done by means of log-likelihood ratio tests of significance of (groups of) covariates in the cut-points of model (4)–(5). If a set of coefficients relating to some factors is found to be jointly significant, then the null hypothesis of homogeneity of reporting behaviour with respect to these factors is rejected. We also use likelihood ratio tests to test whether the effect of a covariate is equal across all thresholds, which is labelled 'parallel cut-point shift'. We return to this in the next section.

RESULTS

For each of the six health domains, we estimate ordered probit models, Equations (1) and (2), and HOPIT models, Equations (3)–(7), separately for each of the three countries. The index function and the cut-points are specified as functions of the same covariates: FEMALE, AGE3044, AGE4559, AGE60, EDUC2, EDUC3, EDUC4, Log(INCOME) and URBAN. The mean health function in the vignette component of the HOPIT includes only dummies indicating the respective vignettes. With two models estimated for 6 domains and 3 countries, we do not present all the parameter estimates.[12] We first report results on tests for homogeneous reporting behaviour. Next we turn to estimates of the magnitude of reporting heterogeneity and finally to the degree to which reporting heterogeneity biases measures of socio-economic inequality in health.

Tests of reporting homogeneity

Table II presents the results of tests of homogeneous reporting behaviour and parallel cut-point shift. For homogeneity, each column gives the *p*-values of likelihood ratio tests of joint significance of the

[12] These are available from the authors upon request.

DOI: 10.1002/hec

Table II. Log-likelihood ratio tests of homogeneity and parallel cut-point shift: *p*-values

	Homogeneity						Parallel cut-point shift
	All	Female	Age	Educ	Log(Inc)	Urban	All
Indonesia							
Mobility	0.000	0.121	0.000	0.000	0.323	0.000	0.000
Cognition	0.000	0.758	0.077	0.001	0.000	0.000	0.000
Pain	0.000	0.028	0.000	0.462	0.058	0.000	0.000
Self-care	0.000	0.051	0.000	0.166	0.000	0.000	0.000
Usual	0.000	0.698	0.867	0.045	0.009	0.000	0.000
Affect	0.000	0.024	0.210	0.071	0.001	0.001	0.000
India							
Mobility	0.000	0.000	0.076	0.171	0.000	0.000	0.000
Cognition	0.000	0.000	0.355	0.364	0.000	0.049	0.000
Pain	0.000	0.000	0.279	0.016	0.109	0.233	0.000
Self-care	0.000	0.000	0.003	0.909	0.009	0.707	0.000
Usual	0.000	0.000	0.002	0.020	0.000	0.179	0.000
Affect	0.000	0.000	0.742	0.535	0.006	0.005	0.000
China							
Mobility	0.000	0.000	0.000	0.557	0.369	0.069	0.043
Cognition	0.000	0.471	0.014	0.058	0.003	0.000	0.002
Pain	0.000	0.005	0.000	0.000	0.000	0.002	0.000
Self-care	0.000	0.768	0.000	0.050	0.000	0.000	0.000
Usual	0.000	0.170	0.085	0.288	0.006	0.000	0.000
Affect	0.027	0.935	0.180	0.516	0.000	0.100	0.663

respective (groups of) covariates in the four cut-points. For each country, the first column shows evidence of cut-point heterogeneity according to at least one of the characteristics for all health domains. For the specific characteristics, the tests indicate some variation in the presence of reporting heterogeneity across domains and countries. Homogeneity of reporting by sex is rejected (5% or less) for all domains in the case of India but only for two domains in each of Indonesia and China. Homogeneity by age is rejected for four domains in China, three domains in Indonesia and two domains in India. The null hypothesis that the cut-points are invariant with respect to education is rejected for three domains in Indonesia and two in each of India and China, but there is relatively little consistency across the countries in the domains for which there is evidence of reporting heterogeneity. The evidence for reporting heterogeneity by income is stronger. The null is rejected for all but one domain in each of India and China and for all but two domains in Indonesia. There is also strong evidence for differences in reporting behaviour across urban and rural locations.

In the final column of Table II, we report tests of whether the covariates affect all cut-points by the same magnitude, i.e. whether there is parallel cut-point shift. The null is decisively rejected in all cases but for affective behaviour in China. This suggests that covariates do not simply alter the overall conception of health but that reporting behaviour is stronger at some levels of health than others and that the effect need not even be monotonic. The nature of the reporting differences can be better understood through examination of the cut-point coefficients themselves. We now turn to this.

Reporting behaviour

The response categories for the degree of difficulty/pain/distress within any domain range from 'extreme/cannot do' to 'none' and so higher health standards or expectations are represented in the HOPIT model by positive shifts in the cut-points. If a certain covariate has positive coefficients across all the cut-points, then higher values of the covariate are associated with higher health standards, i.e.

DOI: 10.1002/hec

Table III. Estimated coefficients of LOG(INCOME) in the cut-points

	Indonesia				India				China			
	ctpt1	ctpt2	ctpt3	ctpt4	ctpt1	ctpt2	ctpt3	ctpt4	ctpt1	ctpt2	ctpt3	ctpt4
Mobility	0.022	−0.003	−0.005	0.000	−0.011	**−0.058**	−0.017	0.029	0.009	0.003	0.011	−0.006
	(1.829)	(−0.385)	(−0.684)	(−0.014)	(−0.691)	(−4.569)	(−1.364)	(1.921)	(0.806)	(0.379)	(1.374)	(−0.734)
Cognition	0.007	0.007	**0.034**	**0.029**	0.007	−0.020	**0.041**	0.011	0.024	0.018	**0.038**	0.017
	(0.499)	(0.769)	(4.226)	(3.135)	(0.287)	(−1.366)	(2.848)	(0.586)	(1.482)	(1.692)	(3.848)	(1.466)
Pain	**−0.033**	−0.001	−0.003	−0.008	0.000	−0.014	0.006	**0.059**	**0.038**	**0.042**	**0.032**	**0.047**
	(−2.956)	(−0.158)	(−0.290)	(−0.594)	(0.020)	(−1.062)	(0.428)	(2.247)	(2.740)	(4.204)	(3.228)	(2.976)
Self-care	**−0.063**	−0.014	0.004	−0.012	0.020	0.014	**0.045**	**0.052**	**0.071**	**0.080**	**0.065**	**0.035**
	(−5.670)	(−1.502)	(0.413)	(−0.974)	(1.002)	(1.076)	(3.256)	(2.866)	(4.295)	(6.928)	(6.228)	(2.611)
Usual	−0.016	0.010	**0.026**	0.018	**0.043**	−0.003	**0.054**	**0.049**	−0.015	−0.011	**0.026**	−0.005
	(−1.346)	(1.231)	(3.248)	(1.851)	(2.000)	(−0.254)	(3.734)	(2.531)	(−0.944)	(−0.952)	(2.590)	(−0.409)
Affect	−0.012	−0.017	0.013	**0.037**	0.030	−0.017	**0.037**	0.000	0.019	**0.028**	**0.037**	**0.044**
	(−0.868)	(−1.753)	(1.309)	(3.449)	(1.385)	(−1.025)	(2.208)	(−0.009)	(1.415)	(2.530)	(3.821)	(4.064)

Note: *t*-ratios in parentheses. Bold indicates significance at 5%. ctpt, cut-point. Ctpt1 is the lowest cut-point determining probability of extreme difficulty/pain/distress, Ctpt4 is the highest cut-point determining probability of no difficulty/pain/distress.

lower probabilities of reporting better levels of health. The cut-point coefficients for income are presented in Table III. To save on space, we do not present the coefficients for the other variables but illustrate the direction and magnitude of all effects graphically in Figure 1.[13] For income, consistent with the LR tests, there is a significant effect on at least one cut-point for all domains and countries except for *mobility* in Indonesia and China. The significant coefficients are mostly positive, with the exceptions of *pain* and *self-care* for Indonesia and *mobility* for India. There are no significant negative effects on the uppermost cut-point (4) and many significant positive effects, the latter implying that the better-off have a lower probability rating a vignette as corresponding to no difficulty/pain/distress. They have a higher standard regarding what it means to have very good health.

Since it is difficult to directly assess the relative importance of the reporting effects from the coefficients alone, we use the parameter estimates of the reporting model (3)–(5) to calculate, per domain, the probability that an individual with given characteristics will rate a hypothetical individual (vignette) as being without difficulty/pain/distress, which we will refer to as very good health. As the reference individual we use a male from the youngest age group, without even primary level education, living in a rural area and with income at the threshold of the poorest quintile. To assess the effect of reporting differences by income alone, we redo the calculation for the same individual, but now with income at the threshold of the richest quintile. The ratio of the two probabilities is used as a measure of the relative magnitude of the reporting effect. We repeat these calculations changing in turn sex from male to female, age from the youngest to oldest group, location from rural to urban and education from the lowest to highest level. The results are depicted in Figure 1.

The top row of the figure presents the income effects. A ratio smaller than one implies that an individual with income at the threshold of the top quintile has a lower probability of reporting very good health than an otherwise identical individual at the threshold of the bottom income quintile. For Indonesia, as was evident from the coefficients in Table III, there are no differences in reporting by income for four of the six domains. Even the two significant effects for *cognition* and *affect* are quantitatively unimportant, with the probability of the highest quintile reporting very good health being only 1% smaller than that of the bottom quintile. For India, the significant effects of income for *pain*, *self* and *usual* are quantitatively slightly more important. In particular, at the highest quintile, the

[13] The education coefficients are given in the Appendix Table AIII in Appendix A.

DOI: 10.1002/hec

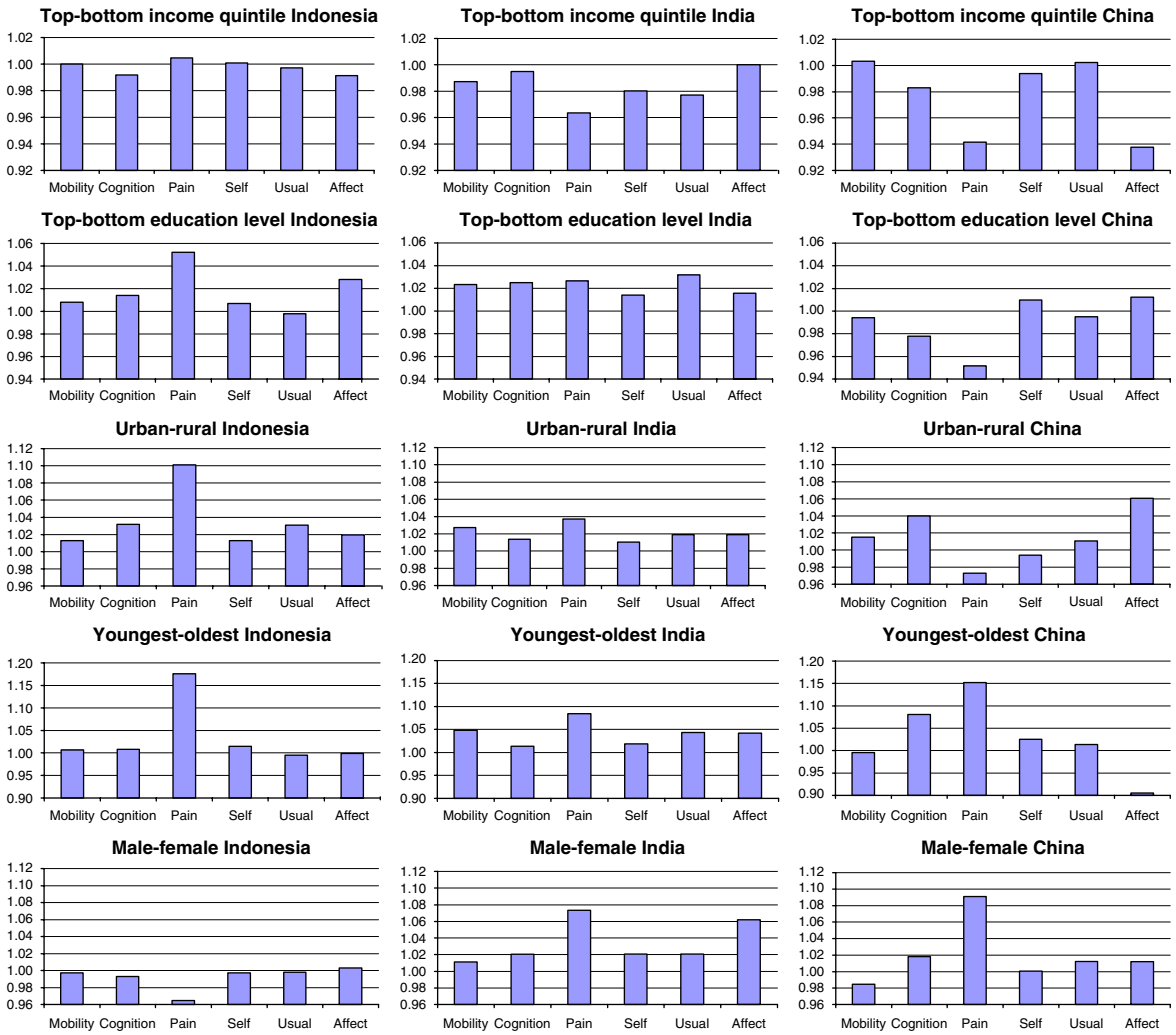

Figure 1. Relative probabilities of reporting very good health by socio-economic group

probability of reporting no pain is 4% less than that of the lowest quintile. For China, the relative difference in probabilities reaches about 6% for both *pain* and *affect*.

In Indonesia and India there are either no differences in reporting by education or the better educated are more likely to report very good health, whereas in China the opposite is true for two of the domains (*cognition* and *pain*). But the differences are again generally small in magnitude. The greatest differences are in the domain of *pain*, with the most educated having a probability of reporting no pain that is 5% greater than that of the least educated in Indonesia but 5% smaller than the least educated in China. Otherwise, the relative differences in reporting between high and low education groups barely exceed 2%. The results for Indonesia and India are perhaps surprising. They do not support the contention that education raises health expectations. It might be that the better educated are more capable of conceptualising the health consequences of a given level of functioning described in a vignette. Or there may be differential capacity by education level in comprehension of the vignette rating exercise. As mentioned above, Murray *et al.* (2003) find that lower educated groups display greater inconsistencies

DOI: 10.1002/hec

between their ranking of the vignettes and the average ranking. But this education effect is not large and anyway while it suggests that there is more noise in the ratings of the less well educated, it does not explain why the better educated are more likely to give more positive evaluations. It is, however, notable that the Chinese sample has considerably higher levels of education than the others. It may be that the vignette exercise was more comprehensible for the Chinese sample.

For Indonesia and India, people from urban areas are more likely to report very good health than those living in rural areas (Figure 1, row 3). The effect is largest for *pain*, with about a 10% relative difference in probabilities for Indonesia and a difference of almost 4% for India. In China, the urban–rural effect is in the opposite direction for *pain*. For India and China but not for Indonesia, men are more likely than women to report very good health. The relative difference in probabilities is largest for *pain*, reaching 7 and 9% for India and China, respectively. This is consistent with a large body of epidemiological and experimental evidence showing that women are more likely to report negative responses to (their own) pain (Unruh, 1996; Riley *et al*., 1998). Generally, the youngest age group (18–29 years) is more likely to rate a vignette as corresponding to no difficulty/pain/distress than the oldest group (60+ years). The differences are again greatest for *pain*. The direction of these age effects is perhaps surprising given evidence that, conditional on some objective health indicator, the elderly are more likely to assess their health positively (Idler, 1993; Van Doorslaer and Gerdtham, 2003; Lindeboom and Van Doorslaer, 2004). The evidence on whether perceptions of pain vary with age is, however, ambiguous (Gibson and Helme, 2001). Our results are consistent with the elderly empathising more with the vignette descriptions, which might be expected given their greater life experiences. We now turn to the question of how the reporting heterogeneity by socio-economic and demographic characteristics identified in the figure affects measured health disparities by income and education.

Purging reporting bias from measures of health disparities

The parameter estimates of the index function of the standard ordered probit model (1) will reflect true health effects and the effects of reporting heterogeneity. Therefore, in the presence of reporting heterogeneity, inequality measures based on these parameter estimates will be biased. With estimates from the HOPIT model, we can separate the reporting heterogeneity (parameters γ from Equation (5)) from the true health effects (parameters β from Equation (6)). In order to gauge the degree of bias generated by reporting heterogeneity, we compare inequality measures based on the ordered probit model with those obtained from the appropriate parameters of the HOPIT model. Given that the scale of the latent variable is not identifiable in the ordered probit model, the constant term and the variance are usually set equal to 0 and 1, respectively. Here, in order to make the estimated effects from the two models comparable, we fix the scale of the ordered probit model by setting the constant term and the variance equal to those estimated by the HOPIT model.

The income and education coefficients in the health equations, (1) and (6), are shown in Tables IV and V, respectively. For all three countries, the ordered probit results indicate significant positive relationships between income and all health domains, except for *pain* in Indonesia and India, even without any adjustment for reporting heterogeneity. For 14 of the 18 cases (6 health domains by 3 countries), the HOPIT adjustment increases the magnitude of the income coefficient. Indeed, regarding reporting behaviour, as we saw in Section 4.2, better-off individuals generally have higher expectations (standards) for health. The HOPIT model can separate this effect from the health effects and, therefore, gives greater income gradients than the ordered probit model. A first conclusion is, therefore, that the positive association between income and health is generally underestimated if reporting heterogeneity by income is not accounted for.

The ordered probit education coefficients are significantly positive for all three countries in almost every domain confirming a positive association between health and education, before the correction for reporting bias (Table V). The vignette adjustment for reporting bias leads to a decrease in 13 of the 18

DOI: 10.1002/hec

Table IV. Estimated coefficients of LOG(INCOME) before and after adjustment

	Indonesia		India		China	
	Before	After	Before	After	Before	After
Mobility	**0.054**	**0.054**	**0.055**	**0.065**	**0.137**	**0.134**
	(2.678)	(2.524)	(2.166)	(2.333)	(8.363)	(7.365)
Cognition	**0.034**	**0.063**	**0.108**	**0.121**	**0.089**	**0.109**
	(2.280)	(3.699)	(3.987)	(3.923)	(6.929)	(6.710)
Pain	0.013	0.007	0.042	**0.074**	**0.099**	**0.141**
	(1.155)	(0.452)	(1.524)	(2.249)	(7.223)	(7.630)
Self-care	**0.062**	0.051	**0.074**	**0.124**	**0.185**	**0.225**
	(2.356)	(1.787)	(2.968)	(4.227)	(6.229)	(6.899)
Usual activities	**0.066**	**0.085**	**0.110**	**0.158**	**0.137**	**0.138**
	(3.297)	(3.886)	(4.142)	(5.168)	(7.394)	(6.523)
Affect	**0.037**	**0.068**	**0.179**	**0.183**	**0.107**	**0.148**
	(2.065)	(3.321)	(5.560)	(4.996)	(8.327)	(9.449)

Note: *t*-ratios in parentheses. Bold indicates significance at 5%.

education coefficients for Indonesia and 15 of the 18 for India. For these two countries, in general, more educated people appear to over report their health (in particular, they are more likely to report no difficulties/pain/distress in a given domain) and this means that the estimated effects of education on health are overstated when reporting bias is not accounted for. The direction of the adjustment is in the opposite direction in the case of China. Purging reporting bias raises 12 of the 18 education coefficients.

Again we performed some calculations with the model in order to quantify the effects of correcting for reporting heterogeneity on a measure of inequality. We calculate per domain the probability of having no difficulty/pain/distress for the reference individual as defined above, i.e., male, youngest age group, lowest education, rural dweller and income at threshold of poorest quintile. Changing either income or education, re-computing the probability and expressing this as a ratio of that for the reference individual gives a measure of relative inequality that reflects the health gradient by income/education holding the other characteristics constant. We calculate the ratio using the standard ordered probit and the HOPIT model. For the HOPIT calculations, we fix the cut-points to the characteristics of the reference individual. The calculated ratios based on the HOPIT model now reflect purely health effects. The difference between the ordered probit and the HOPIT results gives an indication of the extent of the bias induced by reporting heterogeneity. Note that this procedure purges reporting heterogeneity deriving from all covariates and not only that from income or education for which the health disparity is computed. The results are depicted in Figure 2.

From the first row of the figure it is immediately clear that income gradients in health are strongest in China and are negligible in Indonesia. The correction for reporting heterogeneity does not give rise to any noticeable gradient in Indonesia. For India, the income gradients are modest and while purging reporting bias (indicated by the difference between the light and dark bars) consistently shifts them upward, substantially so in relative terms for *pain*, *self-care* and *usual*, they remain modest after the adjustment. For China, we generally see a strong positive income gradient that is increased, most noticeably for *pain* and *affect*, after correction for reporting heterogeneity. There is a positive education gradient for all countries, which is shifted slightly downward for India and Indonesia and, in some cases, slightly upward for China. But all adjustments are small.

Figure 2 illustrates the effect of purging reporting heterogeneity from the partial associations between income/education and health. Measurement of socio-economic inequalities in health usually focuses on the total association between health and some measure of socio-economic rank, possibly standardised for demographics like age and sex. To check on the effect of reporting bias on a measure of total socio-economic inequality in health, we compute the concentration index (Kakwani *et al*., 1997) for the

DOI: 10.1002/hec

Table V. Estimated coefficients of education dummies before and after adjustment

		Indonesia		India		China	
		Before	After	Before	After	Before	After
Mobility	EDUC2	**0.282**	**0.249**	**0.268**	0.162	**0.206**	**0.241**
		(3.154)	(2.656)	(2.796)	(1.520)	(2.423)	(2.542)
	EDUC3	**0.344**	**0.251**	**0.321**	0.235	**0.334**	**0.355**
		(3.498)	(2.437)	(2.884)	(1.902)	(4.093)	(3.914)
	EDUC4	**0.627**	**0.563**	**0.530**	**0.467**	**0.288**	**0.299**
		(4.955)	(4.248)	(6.108)	(4.839)	(3.258)	(3.041)
Cognition	EDUC2	**0.157**	**0.136**	**0.282**	**0.331**	**0.221**	**0.281**
		(2.695)	(2.032)	(2.724)	(2.842)	(3.360)	(3.428)
	EDUC3	**0.269**	**0.149**	**0.674**	**0.554**	**0.343**	**0.439**
		(4.236)	(2.042)	(5.351)	(3.970)	(5.477)	(5.581)
	EDUC4	**0.514**	**0.474**	**0.658**	**0.586**	**0.210**	**0.224**
		(6.567)	(5.345)	(6.959)	(5.513)	(3.112)	(2.652)
Pain	EDUC2	0.001	0.004	0.167	−0.020	**0.246**	**0.257**
		(0.032)	(0.066)	(1.626)	(−0.161)	(3.460)	(2.698)
	EDUC3	**0.127**	**0.141**	**0.515**	**0.395**	**0.402**	**0.286**
		(2.810)	(2.322)	(4.290)	(2.672)	(5.948)	(3.155)
	EDUC4	**0.236**	**0.181**	**0.731**	**0.670**	**0.384**	**0.409**
		(4.488)	(2.623)	(7.938)	(5.871)	(5.244)	(4.100)
Self-care	EDUC2	**0.330**	**0.345**	**0.260**	0.140	**0.404**	**0.378**
		(2.775)	(2.747)	(2.709)	(1.222)	(2.687)	(2.299)
	EDUC3	**0.436**	**0.353**	0.211	0.177	**0.460**	**0.375**
		(3.290)	(2.529)	(1.918)	(1.352)	(3.183)	(2.375)
	EDUC4	**0.602**	**0.517**	**0.704**	**0.632**	**0.710**	**0.600**
		(3.682)	(3.009)	(7.617)	(5.817)	(4.407)	(3.411)
Usual activities	EDUC2	**0.195**	**0.204**	**0.222**	**0.320**	**0.329**	**0.359**
		(2.331)	(2.234)	(2.219)	(2.789)	(3.455)	(3.336)
	EDUC3	**0.453**	**0.406**	**0.367**	0.217	**0.515**	**0.545**
		(4.765)	(3.938)	(3.119)	(1.641)	(5.639)	(5.258)
	EDUC4	**0.508**	**0.526**	**0.778**	**0.659**	**0.440**	**0.447**
		(4.542)	(4.365)	(8.182)	(6.081)	(4.441)	(3.979)
Affect	EDUC2	0.068	0.014	**0.246**	0.190	**0.185**	**0.187**
		(0.991)	(0.181)	(2.021)	(1.377)	(2.687)	(2.110)
	EDUC3	**0.154**	0.028	**0.603**	**0.469**	**0.227**	0.146
		(2.097)	(0.337)	(4.243)	(2.930)	(3.494)	(1.749)
	EDUC4	**0.187**	0.078	**0.686**	**0.632**	**0.139**	0.111
		(2.203)	(0.803)	(6.344)	(5.115)	(1.975)	(1.226)

Note: *t*-ratios in parentheses. Bold indicates significance at 5%.

predicted probability of being in very good health (as defined in Figure 2) against income. Probabilities are obtained both from the ordered probit and from the HOPIT with cut-points set equal to those of the reference individual as defined above. Control is made for differences in demographic composition by income level by setting age and sex to the values of the reference individual in predicting the health index from both models. Results are presented in Figure 3 and show that *total* income-related health inequality is generally largest in India, whereas the *partial* correlations show greatest disparities in China (Figure 2, row 1). In general, the effect of purging reporting heterogeneity from the total correlations is even smaller than that observed for the partial correlations in Figure 2 and there are even some cases in which the adjustment reduces the inequality. This is because while purging differences in health reporting by income increases measured inequality, this is offset by the effect of purging

DOI: 10.1002/hec

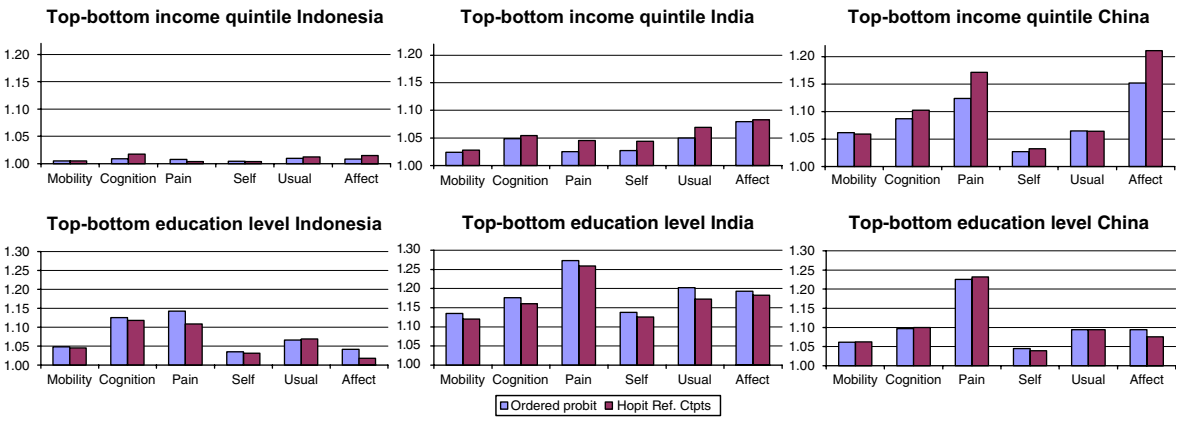

Figure 2. Relative probabilities of being in very good health by income quintile and education level, before and after adjustment

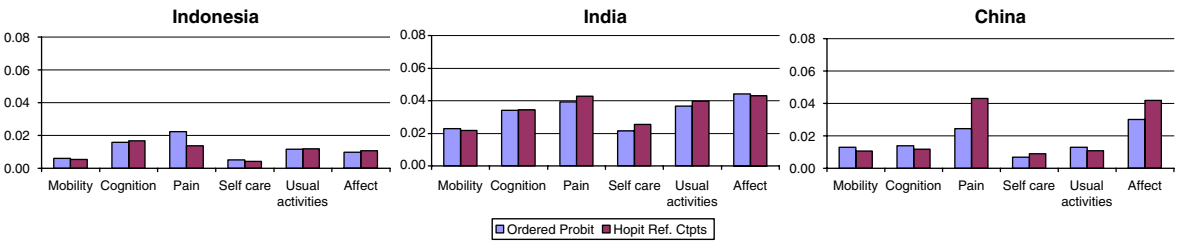

Figure 3. Age–sex standardised concentration indices of predicted probability of being in very good health, before and after adjustment

reporting differences by some other characteristics that are correlated with income. Only in China in the domains of *pain* and *affect* are there marked increases in income-related health inequality.

CONCLUSION AND DISCUSSION

In this paper we have investigated whether there is heterogeneity in health reporting and whether and how this affects the measurement of socio-economic disparities in six domains of self-reported health. We have done this for three low-/middle-income Asian countries (India, Indonesia and China) using WHO-MCS data that, in addition to respondents' assessments of their own health, include their assessments of vignette descriptions of health functioning. Such data allow for the estimation of hierarchical ordered probit models that consist of two simultaneously estimated parts: the vignette ratings are used to estimate the effects of socio-demographics on thresholds for reporting levels of health, while respondents' own health ratings are used to estimate socio-demographic effects on own health. We then use these estimates to test reporting homogeneity and to examine the impact of correcting for heterogeneity on disparities in health by income and education.

The hypothesis of homogeneous reporting across all socio-demographics is rejected for all countries and health domains. Homogeneity tends to be most consistently (across countries and health domains) and decisively rejected across urban/rural, income and age groups and less consistently across sex and

DOI: 10.1002/hec

education groups. Parallel shift of reporting thresholds is rejected in all but one case, indicating that socio-demographics do not simply shift the thresholds by the same magnitude and in the same direction. There is variation in the direction and strength of the reporting differences across countries and domains. Generalising somewhat, younger, male (not Indonesia), better educated (not China), low-income and urban respondents display lower health expectations. These groups are more likely to assess a health condition positively. Correcting for reporting heterogeneity tends to slightly reduce disparities in health by education in India and Indonesia and to increase income disparities in health in all three countries. But the most important finding of the study is that while systematic reporting heterogeneity is significant, it is relatively small in magnitude and does not have a substantial effect on measured socio-economic inequality in health. For most countries and health domains, correction for reporting differences changes the concentration index by no more than 0.005. In China, however, the index rises by 0.019 (76%) and by 0.012 (39%) for the domains of pain and affect, respectively.

While this result is consistent with studies of high-income countries that have found only limited evidence of reporting differences by income and education (Van Doorslaer and Gerdtham, 2003; Lindeboom and Van Doorslaer, 2004), in a low-income setting one would expect the purging of reporting differences to raise income gradients in health to a greater degree, given evidence of substantial inequalities in objective, but not subjective, indicators of health. This suggests that the vignette method has not identified the full scale of reporting heterogeneity, which is a disappointing conclusion for researchers searching for an instrument to measure socio-economic inequality in health in developing countries. It is possible that the identifying assumptions – vignette equivalence and response consistency – are not valid in these data. As mentioned above, Murray *et al.* (2003) do find some variation in the ranking of the vignettes by age, sex and education, which suggests some systematic differences in perceptions of the level of functioning described by any one vignette, possibly arising from multidimensionality of the health domain. While this casts some doubt on the validity of vignette equivalence, the systematic variation is very small and certainly does not amount to a decisive rejection of the assumption. It is perhaps relevant that Murray *et al.* (2003) also find the degree of consistency with the global average of the vignette rankings is generally higher in high-income than in low-income countries. This suggests that there is more noise in the vignette data for low-income countries and consequently it will be more difficult to identify reporting behaviour from them. It may also suggest that the vignette exercise is more feasible in more highly educated populations. Note that the vignette correction to income-related health inequality was largest for the Chinese sample, which has a much higher level of education than the other two. The feasibility of applying the vignette method in poorly educated, low-income populations should be further examined. There is also a need for evaluation of the validity of the vignette approach in the form of experiments designed to directly test the assumptions of vignette equivalence and response consistency. The appropriate design of such experiments represents an important challenge for future research concerning the measurement of health in large-scale household surveys.

Future applications of the vignette approach should also give consideration to what variation in reporting it is appropriate to remove from a health measure. Arguably, perceptions of health are more important to quality of life experiences than are objective health conditions. This raises the difficult question of whether health is ever interpersonally comparable. Any attempt to measure health inequality must assume that it is, at least to some degree. In this context, we argue that an appropriate measure of socio-economic inequality in health should correct for any tendency of better-off individuals to report their health more negatively for a given condition. But it may not be considered appropriate to remove differences in the reporting of health by sex, for example. The tendency for women to report pain more negatively, confirmed here for India and China, presumably does indicate that the real experience of pain is greater for women and this should be reflected in a health measure.

Finally, our general finding that, while significant, reporting heterogeneity does not appear to have a large quantitative impact on measured socio-economic disparities in health may be contingent upon the

DOI: 10.1002/hec

measurement of health separately in each of six domains rather than through a single indicator of general health. By separating health into six dimensions, much of the heterogeneity in the reporting of the standard self-assessed health question is possibly removed. There is no heterogeneity deriving from differential weighting of each dimension of health. For example, manual and non-manual workers may differ little in how they report that a given condition constrains their mobility but the former may place more weight on this mobility constraint in assessing their general health. It remains to be seen whether the vignette approach can be extended to the measurement of general health and if so what will be the impact on disparities in general health.

ACKNOWLEDGEMENTS

Teresa Bago d'Uva was funded by Fundação para a Ciência e Tecnologia, under PhD grant SFRH/BD/10551/2002. All four authors are funded by the Netspar project: 'Health, income and work across the life cycle'. The authors thank the WHO for providing access to the MCS data. In particular, we are very grateful to Somnath Chatterji for supplying the data set and advising on its use. We are grateful for comments received from two anonymous referees, Andrew Jones, Nigel Rice and seminar participants at the University of Melbourne, the University of New South Wales, the University of York, Erasmus University Rotterdam and the ECuity Project meeting at the IZA in Bonn. The usual disclaimer applies.

APPENDIX A

WHO-MCS data for Indonesia (excluding Papua, Aceh and Maluku), an Indian state (Andhra Pradesh) and three Chinese provinces (Gansu, Henan and Shan-dong) were used. Table AI documents the number of observations lost due to item non-response. The distributions of the vignette evaluations are given in Table AII. The estimated coefficients of education dummies in the cut-points are given Table AIII.

Table AI. Sample sizes and item non-response

	Indonesia	India	China
Full sample	9994	5196	9486
Observations lost due to item non-response			
Own health domains	81	52	83
Income	2091	41	2223
All covariates	2187	43	2304
Final sample	7770	5129	7156
Additional observations lost from vignette component of HOPIT due to non-response to at least one vignette			
Mobility	2	3	3
Self-care	2	3	1
Usual	6	10	102
Pain	8	9	26
Affect	2	17	9
Cognition	7	13	106

DOI: 10.1002/hec

Table AII. Frequencies of own health and vignettes by domain and country

	Indonesia									India									China								
	Own	vig1	vig2	vig3	vig4	vig5	vig6	vig7	vig8	Own	vig1	vig2	vig3	vig4	vig5	vig6	vig7	vig8	Own	vig1	vig2	vig3	vig4	vig5	vig6	vig7	vig8
Mobility																											
Extreme	0.21	0.36	0.39	0.98	2.88	5.83	46.11			0.76	0.04	0.12	0.75	2.25	9.79	63.77			0.20	0.22	0.41	1.62	1.71	6.71	61.68		
Severe	0.98	4.08	3.03	10.27	36.06	50.50	39.97			5.58	0.91	1.58	8.48	30.35	54.14	29.32			0.81	1.18	1.13	4.84	24.65	44.26	26.66		
Moderate	2.64	5.73	6.42	27.20	34.96	24.51	5.99			7.35	0.43	1.97	26.05	41.20	24.98	3.91			3.63	2.15	4.87	20.77	47.59	34.11	5.42		
Mild	5.28	11.64	16.18	34.09	18.57	12.61	4.08			22.19	2.80	4.78	50.00	24.03	10.54	1.93			15.24	5.89	17.39	59.12	23.41	12.68	4.10		
None	90.90	78.19	73.98	27.46	7.53	6.55	3.85			64.13	95.82	91.55	14.72	2.17	0.55	1.07			80.12	90.56	76.20	13.65	2.64	2.23	2.15		
N	7770	3893	3893	3893	3893	3893	3893			5129	2534	2534	2534	2534	2534	2534			7156	3635	3635	3635	3635	3635	3635		
Cognition																											
Extreme	0.15	0.37	1.42	0.68	1.11	6.37	5.74	11.47	20.16	0.58		0.16	0.16	0.88	2.00	6.31	17.11	17.19	0.10	0.53	1.35	0.94	0.53	6.69	3.17	5.87	38.73
Severe	1.31	3.89	14.21	8.95	19.26	46.89	38.79	51.42	57.21	4.04	2.00	5.36	20.06	17.91	35.09	48.68	58.75	66.27	1.20	0.41	13.97	4.64	3.58	31.98	20.77	35.74	40.73
Moderate	5.66	6.58	28.16	34.74	33.53	32.79	37.79	25.58	14.68	8.34	3.04	21.34	34.53	29.82	38.13	26.14	19.50	11.27	5.64	2.23	30.99	16.49	20.54	37.38	41.67	39.67	13.97
Mild	14.04	16.58	33.16	42.84	37.00	11.32	14.58	9.11	4.89	19.83	9.03	45.64	38.77	42.37	23.66	17.67	4.16	4.40	25.61	6.87	40.02	49.71	57.45	19.84	29.23	16.02	4.87
None	78.83	72.58	23.05	12.79	9.11	2.63	3.11	2.42	3.05	67.21	85.93	27.50	6.47	9.03	1.12	1.20	0.48	0.88	67.44	89.96	13.67	28.23	17.90	4.11	5.16	2.7	1.7
N	7770	1900	1900	1900	1900	1900	1900	1900	1900	5129	1251	1251	1251	1251	1251	1251	1251	1251	7156	1704	1704	1704	1704	1704	1704	1704	1704
Pain																											
Extreme	0.26	0.98	1.44	9.26	4.73	6.33	9.21	46.58		1.07	0.55	3.40	4.98	8.77	4.82	12.01	48.42		0.14	3.04	0.72	8.52	3.87	1.31	7.81	61.50	
Severe	2.41	6.85	21.00	45.08	44.62	50.03	52.08	39.48		8.21	11.06	33.97	52.92	68.64	59.32	58.69	47.87		1.47	16.75	6.79	35.10	37.19	19.25	34.09	27.65	
Moderate	11.49	26.56	44.98	30.06	30.88	28.98	26.92	7.46		13.10	25.67	38.31	30.17	16.67	27.41	18.40	2.61		7.60	38.68	35.88	41.60	44.87	42.19	40.35	6.91	
Mild	30.01	55.84	26.87	12.45	16.37	10.29	9.73	3.86		27.43	57.03	23.30	11.77	5.45	8.29	10.66	1.11		36.94	39.63	51.79	13.41	12.93	35.04	16.57	3.16	
None	55.84	9.78	5.71	3.14	3.40	4.37	2.06	2.62		50.19	5.69	1.03	0.16	0.47	0.16	0.24			53.85	1.91	4.83	1.37	1.13	2.21	1.19	0.77	
N	7770	1943	1943	1943	1943	1943	1943	1943		5129	1266	1266	1266	1266	1266	1266	1266		7156	1678	1678	1678	1678	1678	1678	1678	
Self-care																											
Extreme	0.21	0.36	1.54	2.62	3.44	6.05	4.26	25.91		0.66	0.24	1.73	3.14	2.04	15.40	7.07	37.16		0.10	0.41	0.41	1.24	1.36	7.08	2.83	45.04	
Severe	0.49	3.59	13.65	29.66	37.51	42.79	40.43	51.72		3.06	1.96	37.23	38.81	34.88	44.38	49.65	51.37		0.43	1.12	6.43	10.80	18.00	44.21	15.76	39.91	
Moderate	1.47	6.88	43.92	40.69	36.12	21.70	33.71	11.24		4.70	2.99	41.95	33.39	45.48	14.22	28.59	8.09		1.47	3.13	30.64	31.88	48.17	30.81	40.91	10.04	
Mild	4.18	18.47	31.20	21.96	15.96	16.11	15.08	7.13		16.79	3.14	17.44	23.57	15.79	9.51	13.35	3.14		6.69	11.33	53.13	49.41	30.76	14.29	35.42	4.19	
None	93.66	70.70	9.70	5.08	6.98	13.34	6.52	4.00		74.79	91.67	1.65	1.10	1.81	16.50	1.34	0.24		91.31	84.00	9.39	6.67	1.71	3.60	5.08	0.83	
N	7770	1949	1949	1949	1949	1949	1949	1949		5129	1273	1273	1273	1273	1273	1273	1273		7156	1694	1694	1694	1694	1694	1694	1694	
Usual activities																											
Extreme	0.37	0.47	1.52	3.94	4.20	5.25	4.05	4.94	16.40	1.42	0.32	1.44	3.68	3.44	4.80	3.04	4.40	22.14	0.50	0.59	0.88	2.05	0.41	5.87	2.35	4.81	40.38
Severe	1.65	3.89	21.39	28.48	35.68	39.46	37.20	46.14	56.86	4.91	6.87	21.66	20.70	45.32	47.40	51.48	66.67	52.28	0.92	0.59	5.34	9.74	2.17	26.82	14.85	25.59	38.44
Moderate	3.45	7.83	44.93	34.79	33.68	32.84	41.25	31.74	17.45	6.75	5.36	37.81	32.29	31.41	36.13	35.81	22.30	15.43	2.95	2.76	24.59	28.76	11.03	41.20	40.49	46.71	13.67
Mild	9.07	17.87	25.07	27.64	20.97	16.34	14.24	11.46	6.20	21.19	6.00	25.34	41.65	18.39	9.91	8.79	6.24	8.55	15.73	12.44	50.35	49.30	47.83	20.48	36.33	19.37	5.63
None	85.46	69.94	7.09	5.15	5.47	6.10	3.26	5.73	3.10	65.72	81.45	13.75	1.68	1.44	1.76	0.88	0.40	1.60	79.89	83.63	18.84	10.15	38.56	5.63	5.99	3.52	1.88
N	7770	1903	1903	1903	1903	1903	1903	1903	1903	5129	1251	1251	1251	1251	1251	1251	1251	1251	7156	1704	1704	1704	1704	1704	1704	1704	1704
Affect																											
Extreme	0.31	0.83	0.98	1.92	1.40	14.09	39.12			1.15	0.16	0.16	0.96	2.01	31.94	39.00			0.25	1.57	0.22	0.32	0.70	32.47	50.16		
Severe	1.06	3.63	10.62	21.87	17.72	56.01	44.82			7.92	1.44	9.55	17.74	20.87	60.19	54.74			1.36	1.19	2.27	12.88	8.17	46.59	34.90		
Moderate	5.21	5.18	38.08	36.17	42.64	19.90	7.67			9.50	1.12	30.50	36.36	40.93	5.22	3.69			6.25	2.11	14.07	43.02	32.68	13.53	6.60		
Mild	12.22	13.83	41.97	31.87	32.49	7.46	3.78			22.62	5.14	54.65	42.70	34.67	2.65	1.20			33.34	8.87	69.97	38.47	53.41	5.03	4.06		
None	81.20	76.53	8.34	8.19	5.75	2.54	4.61			58.82	92.13	5.14	2.25	1.52		1.36			58.81	86.26	13.47	5.30	5.03	2.38	4.27		
N	7770	1930	1930	1930	1930	1930	1930			5129	1246	1246	1246	1246	1246	1246			7156	1848	1848	1848	1848	1848	1848		

DOI: 10.1002/hec

Table AIII. Estimated coefficients of education dummies in the cut-points

		Indonesia				India				China			
		ctpt1	ctpt2	ctpt3	ctpt4	ctpt1	ctpt2	ctpt3	ctpt4	ctpt1	ctpt2	ctpt3	ctpt4
Mobility	EDUC2	0.016	0.000	0.023	−0.045	−0.074	−0.059	−0.082	**−0.121**	0.020	−0.007	−0.001	0.043
		(0.407)	(−0.014)	(0.850)	(−1.559)	(−1.168)	(−1.189)	(−1.67)	(−2.122)	(0.324)	(−0.149)	(−0.030)	(0.874)
	EDUC3	−0.033	0.034	0.017	**−0.109**	−0.091	0.059	−0.037	−0.118	−0.031	−0.076	−0.029	0.035
		(−0.803)	(1.183)	(0.614)	(−3.775)	(−1.257)	(1.074)	(−0.664)	(−1.826)	(−0.547)	(−1.699)	(−0.685)	(0.750)
	EDUC4	**0.101**	**0.079**	0.019	**−0.078**	−0.015	0.038	0.012	−0.083	−0.080	**−0.096**	−0.045	0.025
		(2.153)	(2.313)	(0.592)	(−2.301)	(−0.285)	(0.891)	(0.279)	(−1.687)	(−1.276)	(−1.977)	(−0.988)	(0.492)
Cognition	EDUC2	0.026	0.000	0.020	−0.030	0.052	0.081	0.004	0.054	0.063	0.064	0.027	0.082
		(0.523)	(0.000)	(0.624)	(−0.785)	(0.653)	(1.582)	(0.077)	(0.814)	(0.786)	(1.173)	(0.529)	(1.382)
	EDUC3	−0.041	−0.010	**−0.091**	**−0.131**	0.092	−0.010	−0.075	**−0.136**	0.011	0.006	0.010	**0.136**
		(−0.772)	(−0.305)	(−2.742)	(−3.302)	(1.071)	(−0.178)	(−1.358)	(−1.966)	(0.146)	(0.118)	(0.203)	(2.400)
	EDUC4	**0.128**	**0.097**	0.003	−0.049	−0.045	−0.008	−0.025	−0.083	−0.049	−0.066	−0.063	0.047
		(2.337)	(2.638)	(0.072)	(−1.094)	(−0.652)	(−0.187)	(−0.575)	(−1.499)	(−0.587)	(−1.178)	(−1.209)	(0.770)
Pain	EDUC2	−0.025	0.001	0.040	−0.016	−0.023	−0.065	**−0.160**	**−0.241**	0.060	−0.013	0.010	0.025
		(−0.559)	(0.022)	(1.179)	(−0.359)	(−0.329)	(−1.228)	(−2.685)	(−2.418)	(0.813)	(−0.257)	(0.194)	(0.288)
	EDUC3	−0.007	0.052	0.029	0.004	−0.055	0.029	**−0.155**	−0.159	0.040	**−0.134**	−0.085	−0.102
		(−0.142)	(1.549)	(0.807)	(0.078)	(−0.699)	(0.509)	(−2.394)	(−1.353)	(0.575)	(−2.722)	(−1.665)	(−1.231)
	EDUC4	0.028	0.013	0.023	−0.086	0.106	−0.005	**−0.105**	−0.066	0.031	**−0.167**	**−0.108**	0.083
		(0.550)	(0.347)	(0.574)	(−1.615)	(1.903)	(−0.118)	(−2.146)	(−0.739)	(0.412)	(−3.147)	(−1.993)	(0.910)
Self-care	EDUC2	0.075	0.042	0.014	0.013	0.014	−0.066	−0.073	−0.134	0.134	−0.044	0.044	−0.041
		(1.455)	(1.274)	(0.421)	(0.307)	(0.201)	(−1.264)	(−1.271)	(−1.844)	(1.558)	(−0.743)	(0.803)	(−0.542)
	EDUC3	−0.018	0.011	−0.029	**−0.085**	−0.019	−0.027	0.012	−0.023	0.062	−0.097	−0.065	−0.094
		(−0.326)	(0.309)	(−0.806)	(−1.989)	(−0.234)	(−0.481)	(0.185)	(−0.275)	(0.752)	(−1.727)	(−1.242)	(−1.307)
	EDUC4	0.103	0.014	0.003	**−0.098**	−0.046	−0.052	−0.032	−0.058	−0.056	**−0.123**	−0.051	−0.126
		(1.713)	(0.348)	(0.066)	(−2.072)	(−0.790)	(−1.238)	(−0.674)	(−0.929)	(−0.627)	(−2.034)	(−0.907)	(−1.631)
Usual	EDUC2	−0.021	0.033	**0.073**	−0.005	0.138	0.071	**0.109**	0.103	−0.098	−0.029	0.030	0.035
		(−0.416)	(1.103)	(2.315)	(−0.117)	(1.937)	(1.527)	(2.072)	(1.454)	(−1.22)	(−0.51)	(0.595)	(0.579)
	EDUC3	0.047	**0.069**	0.001	−0.056	0.095	−0.086	−0.096	**−0.163**	−0.059	−0.039	−0.005	0.043
		(0.916)	(2.198)	(0.030)	(−1.321)	(1.193)	(−1.705)	(−1.757)	(−2.255)	(−0.785)	(−0.733)	(−0.099)	(0.739)
	EDUC4	−0.041	**0.075**	0.038	0.014	−0.029	−0.072	−0.066	−0.108	−0.122	**−0.138**	−0.061	0.021
		(−0.700)	(2.158)	(1.027)	(0.287)	(−0.460)	(−1.792)	(−1.516)	(−1.841)	(−1.48)	(−2.363)	(−1.158)	(0.345)
Affect	EDUC2	−0.022	0.025	−0.002	−0.067	−0.010	0.005	−0.052	−0.080	−0.057	0.000	−0.009	0.009
		(−0.425)	(0.645)	(−0.055)	(−1.515)	(−0.116)	(0.078)	(−0.819)	(−0.93)	(−0.747)	(0.002)	(−0.160)	(0.140)
	EDUC3	**−0.115**	−0.052	−0.065	**−0.139**	−0.015	−0.067	**−0.166**	−0.142	0.027	−0.009	−0.087	−0.082
		(−2.097)	(−1.314)	(−1.722)	(−3.102)	(−0.17)	(−0.924)	(−2.436)	(−1.509)	(0.376)	(−0.156)	(−1.632)	(−1.327)
	EDUC4	−0.012	0.029	−0.060	**−0.122**	0.049	−0.088	−0.037	−0.052	−0.006	0.021	−0.063	−0.018
		(−0.198)	(0.658)	(−1.378)	(−2.356)	(0.678)	(−1.498)	(−0.679)	(−0.694)	(−0.082)	(0.327)	(−1.081)	(−0.272)

Note: *t*-ratios in parentheses. Bold indicates significance at 5%. ctpt, cut-point. Ctpt1 is the lowest cut-point determining probability of extreme difficulty/pain/distress. Ctpt4 is the highest cut-point determining probability of no difficulty/pain/distress.

DOI: 10.1002/hec

APPENDIX B

Vignette descriptions

MOBILITY

1 - [Paul] is an active athlete who runs long distance races of 20 kilometres twice a week and engages in soccer with no problems.

2 - [Mary] has no problems with moving around or using her hands, arms and legs. She jogs 4 kilometres twice a week without any problems.

3 - [Rob] is able to walk distances of up to 200 metres without any problems but feels breathless after walking one kilometre or climbing up more than one flight of stairs. He has no problems with day-to-day physical activities, such as carrying food from the market.

4 - [Margaret] feels chest pain and gets breathless after walking distances of up to 200 metres, but is able to do so without assistance. Bending and lifting objects such as groceries produces pain.

5 - [Louis] is able to move his arms and legs, but requires assistance in standing up from a chair or walking around the house. Any bending is painful and lifting is impossible.

6 - [David] is paralysed from the neck down. He is confined to bed and must be fed and bathed by somebody else.

COGNITION

1 - [Rob] can do complex mathematical problems in his mind. He can pay attention to the task at hand for long uninterrupted periods of time. He can remember names of people, addresses, phone numbers and such details that go back several years.

2 - [Sue] can only count money and bring back the correct change after shopping. Mental arithmetic is otherwise a problem. She can find her way around the neighbourhood and know where her own belongings are kept.

3 - [Henriette] can pay attention to the task at hand for periods of up to one hour, with occasional distractions and can quickly return to the task. She can remember names of people she meets often, their addresses and important numbers, but occasionally has to remind herself of the names of distant relatives or acquaintances.

4 - [Helena] can remember details of events that have taken place or names of people she has met many years ago, She can do everyday calculations in her mind. During periods of anxiety lasting a few hours, she becomes confused and cannot think very clearly.

5 - [Tom] finds it difficult to concentrate on reading newspaper articles, or watching television programmes. He is forgetful and once a week or so, he misplaces important things, such as keys or money, and spends a considerable amount of time looking for them, but is able to find them eventually.

6 - [Julian] is easily distracted, and within 10 minutes of beginning a task, his attention shifts to something else happening around him. He can remember important facts when he tries, but several times a week finds that he has to struggle to recollect what people have said or events that have taken place recently.

7 - [Christian] is very forgetful and often loses his way around places which are not very familiar. He needs to be prompted about names of close relatives and loses important things such as keys and money, as he cannot recollect where they have been kept. He has to make notes to remind himself to do even very important tasks.

8 - [Peter] does not recognize even close relatives and cannot be trusted to leave the house unaccompanied for fear of getting lost. Even when prompted, he shows no recollection of events or recognition of relatives.

DOI: 10.1002/hec

PAIN

1 - [Laura] has a headache once a month that is relieved one hour after taking a pill. During the headache she can carry on with her day to day affairs.

2 - [Phil] has pain in the hip that causes discomfort while going to sleep. The pain is there throughout the day but does not stop him from walking around.

3 - [Patricia] has a headache once a week that is relieved 3–4 hours after taking a pill. During the headache she has to lie down, and cannot do any other tasks.

4 - [Mark] has joint pains that are present almost all the time. They are at their worst in the first half of the day. Taking medication reduces the pain though it does not go away completely. The pain makes moving around, holding and lifting things, quite uncomfortable.

5 - [Jim] has back pain that makes changes in body position very uncomfortable. He is unable to stand or sit for more than half an hour. Medicines decrease the pain a little, but it is there all the time and interferes with his ability to carry out even day to day tasks.

6 - [Tom] has a toothache for about 10 minutes, several times a day. The pain is so intense that Tom finds it difficult to concentrate on work.

7 - [Steve] has excruciating pain in the neck radiating to the arms that is very minimally relieved by any medicines or other treatment. The pain is sharp at all times and often wakes him from sleep. It has necessitated complete confinement to the bed and often makes him think of ending his life.

SELF CARE

1 - [Helena] keeps herself neat and tidy. She requires no assistance with cleanliness, dressing and eating.

2 - [Anne] takes twice as long as others to put on and take off clothes, but needs no help with this. She is able to bathe and groom herself, though that requires effort and leads to reducing the frequency of bathing to half as often as before. She has no problems with feeding.

3 - [Paul] has no problems with cleanliness, dressing and eating. However, he has to wear clothes with special fasteners as joint problems prevent him from buttoning and unbuttoning clothes.

4 - [Peter] can wash his face and comb his hair, but cannot wash his whole body without help. He needs assistance with putting clothes on over his head, but can put garments on the lower half of his body. He has no problems with feeding.

5 - [John] cannot wash, groom or dress himself without personal help. He has no problems with feeding.

6 - [Rachel] feels pain and discomfort while washing, and in combing her hair. As a result, she neglects her personal appearance. She needs assistance with putting on and taking off clothes. She has no problems with feeding.

7 - [Sue] requires the constant help of a person to wash and groom herself and has to be dressed and fed.

USUAL

1 - [John] is a teacher and goes to work regularly. He teaches the senior grades and takes classes for 6 hours each day. He prepares lessons and corrects exam papers. Students come to him for advice.

2 - [Dan] is a mason in a building firm. Three to four times per week, he is noticed to leave his bricklaying tasks incomplete. With help and supervision, he is able to use his skills to finish the walls of the buildings well.

3 - [Mathew] is a clerk in the local government office. He maintains ledgers with no errors and keeps them up to date. However, he ends up not doing any work for a day once every 2 weeks or so because of a migraine headache.

DOI: 10.1002/hec

4 - [Maria] is an accountant in the local bank. She is regularly at work. However, she makes minor errors in the accounts and tends to postpone tasks. She delays producing account statements and is late on deadlines.
5 - [Carol] is a housewife who leaves most chores around the house half done. Even with domestic help she cannot complete important tasks in time, such as getting her son ready for school. Her husband has had to take over the cooking.
6 - [Doris] is a housewife and does most of the cooking and cleaning around the house. About once a week she leaves tasks half done. Her cooking has deteriorated and the house is not as clean as it used to be. She also takes about twice as long to do the chores.
7 - [Karen] is a teacher and has had to miss work for 2 weeks in the past month. Even now she feels tired and exhausted, and cannot stand for long periods in the classroom. Colleagues notice that she is making serious mistakes in correcting answer papers.
8 - [Jack] is a clerk at the local post office. He just sits around all day and cannot engage in any work. He cannot sort letters, manage the counter or interact with customers. His employers are considering replacing him.

AFFECT
1 - [Ken] remains happy and cheerful almost all the time. He is very enthusiastic and enjoys life.
2 - [Henriette] remains happy and cheerful most of the time, but once a week feels worried about things at work. She gets depressed once a month and loses interest but is able to come out of this mood within a few hours.
3 - [Jan] feels nervous and anxious. He is depressed nearly every day for 3–4 hours thinking negatively about the future, but feels better in the company of people or when doing something that really interests him.
4 - [Eva] feels worried all the time about things at work and home, and feels that they will go wrong. She gets depressed once a week for a day, thinking negatively about the future, but is able to come out of this mood within a few hours.
5 - [John] feels tense and on edge all the time. He is depressed nearly everyday and feels hopeless. He also has a low self esteem, is unable to enjoy life, and feels that he has become a burden.
6 - [Roberta] feels depressed all the time, weeps frequently and feels completely hopeless. She feels she has become a burden, feels it is better to be dead than alive, and often plans suicide.

REFERENCES

Adams P, Hurd MD, McFadden D, Merrill A *et al*. 2003. Healthy, wealthy and wise? Tests for direct causal paths between health and socio-economic status. *Journal of Econometrics* **112**: 3–56.
Baker JL, Van der Gaag J. 1993. Equity in health care and health care financing: evidence from five developing countries. In *Equity in the Finance and Delivery of Health Care*, Van Doorslaer E, Wagstaff A, Rutten F (eds). Oxford University Press: Oxford.
Baker M, Stabile M, Deri C. 2004. What do self-reported, objective measures of health measure? *Journal of Human Resources* **39**(4): 1067–1093.
Benitez-Silva H, Buschinski M, Chan HM, Cheidvasser S *et al*. 1999. How large is the bias in self-reported disability? *Journal of Applied Econometrics* **19**(6): 649–670.
Benzeval M, Taylor J, Judge K. 2000. Evidence on the relationship between low income and poor health: is the Government doing enough? *Fiscal Studies* **21**(3): 375–399.
Bound J. 1991. Self reported versus objective measures of health in retirement models. *Journal of Human Resources* **26**: 107–137.
Deaton A, Paxson C. 1998. Aging and inequality in income and health. Demographic trends and economic consequences. *American Economic Review Papers and Proceedings* **88**(2): 248–253.
Disney R, Emerson C, Wakefield M. 2006. Ill-health and retirement in Britain: a panel data based analysis. *Journal of Health Economics* **25**(4): 621–649.

DOI: 10.1002/hec

Etilé F, Milcent C. 2006. Income-related reporting heterogeneity in self-assessed health: evidence from France. *Health Economics* **15**(9): 965–981.
Ettner SL. 1996. New evidence on the relationship between income and health. *Journal of Health Economics* **15**(1): 67–85.
Frijters P, Haisken-DeNew JP, Shields MA. 2005. The causal effect of income on health: evidence from German reunification. *Journal of Health Economics* **24**(5): 997–1017. DOI: 10.1016/j.jhealeco.2005.01.004.
Gibson SJ, Helme RD. 2001. Age-related differences in pain perception and report. *Clinical Geriatric Medicine* **17**: 433–456.
Gwatkin DR, Rustein S, Johnson K, Pande R *et al*. 2000. Socioeconomic differences in health, nutrition and population. *World Bank Health*, *Nutrition and Population Discussion Paper*, Washington, DC.
Hernandez-Quevedo C, Jones AM, Rice N. 2004. Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey. *ECuity III Working Papers*, York.
Iburg KM, Salomon J, Tandon A, Murray CJL. 2002. Cross-country comparability of physician-assessed and self-reported measures of health. In *Summary Measures of Population Health*: *Concepts*, *Ethics*, *Measurement and Applications*, Murray CJL, Salomon JA, Mathers CD, Lopez AD (eds). WHO: Geneva, 433–448.
Idler E, Benyamini Y. 1997. Self-rated health and mortality: a review of twenty-seven community studies. *Journal of Health and Social Behavior* **38**(1): 21–37.
Idler EL. 1993. Age differences in self-assessments of health: age changes, cohort differences, or survivorship? *Journal of Gerontology* **48**(6): S289–S300.
Idler EL, Kasl SV. 1995. Self-ratings of health: do they also predict change in functional ability? *Journal of Gerontology* **50B**: 344–353.
Jürges H. 2007. True health versus response styles: exploring cross-country differences in self-reported health. *Health Economics* **16**(2): 163–178. DOI: 10.1002/hec.1207.
Kakwani NC, Wagstaff A, Van Doorslaer E. 1997. Socioeconomic inequalities in health: measurement, computation and statistical inference. *Journal of Econometrics* **77**(1): 87–104.
Kapteyn A, Smith J, van Soest A. Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review* **97**(1): 461–473.
Kerkhofs MJM, Lindeboom M. 1995. Subjective health measures and state dependent reporting errors. *Health Economics* **4**: 221–235.
King G, Murray CJL, Salomon J, Tandon A. 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* **98**(1): 184–191.
Kreider B. 1999. Latent work disability and reporting bias. *Journal of Human Resources* **34**(4): 734–769.
Lindeboom M, Van Doorslaer E. 2004. Cut-point shift and index shift in self-reported health. *Journal of Health Economics* **23**(6): 1083–1099.
Mathers CD, Douglas RM. 1998. Measuring progress in population health and well-being. In *Measuring Progress*: *Is Life Getting Better?* Eckersley R (ed.). CSIRO Publishing: Collingwood, 125–155.
Murray CJL. 1996. Epidemiology and morbidity transitions in India. In *Health*, *Poverty and Development in India*, Dasgupta M, Chen CL, Krishnan TN (eds). Oxford University Press: Delhi, 122–147.
Murray CJL, Ozaltin E, Tandon A, Salomon J *et al*. 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). World Health Organization: Geneva.
Pudney S, Shields M. 2000. Gender, race, pay and promotion in the British nursing profession: estimation of a generalized probit model. *Journal of Applied Econometrics* **15**: 367–399.
Riley JL, Robinson ME, Wise EA, Myers CD *et al*. 1998. Sex differences in the perception of noxious experimental stimuli: a meta-analysis. *Pain* **74**: 181–187.
Salomon J, Tandon A, Murray CJL, World Health Survey Pilot Study Collaborating Group. 2004. Comparability of self-rated health: cross sectional multi-country survey using anchoring vignettes. *British Medical Journal* **328**: 258–263.
Sen A. 2002. Health: perception versus observation. *British Medical Journal* **324**: 860–861.
Smith J. 1999. Healthy bodies and thick wallets: the dual relation between health and socioeconomic status. *Journal of Economic Perspectives* **13**: 145–166.
Stern S. 1989. Measuring the effect of disability on labor force participation. *Journal of Human Resources* **24**: 361–395.
Tandon A, Murray CJL, Salomon JA, King G. 2003. Statistical models for enhancing cross-population comparability. In *Health Systems Performance Assessment*: *Debates*, *Methods and Empiricisms*, Murray CJL, Evans DB (eds). World Health Organization: Geneva, 727–746.
Terza JV. 1985. Ordinal probit: a generalization. *Communications in Statistics* **14**(1): 1–11.

Unruh AM. 1996. Gender variations in clinical pain experience. *Pain* **65**: 123–167.
Üstün TB, Chatterji S, Villanueva M, Benib L *et al*. 2003. WHO Multi-Country Survey Study on health and responsiveness 2001–2002. In *Health Systems Performance Assessment*: *Debates*, *Methods and Empiricisms*. Murray CJL, Evans DB (eds). World Health Organization: Geneva, 762–796.
Van Doorslaer E, Gerdtham U-G. 2003. Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science & Medicine* **57**(9): 1621–1629.
Van Doorslaer E, Koolman X. 2004. Explaining the differences in income-related health inequalities across European countries. *Health Economics* **13**(7): 609–628.
Van Doorslaer E, Koolman X, Jones AM. 2004. Explaining income-related inequalities in doctor utilization in Europe. *Health Economics* **13**(7): 629–647.
Van Doorslaer E, Wagstaff A, Bleichrodt H, Calonge S *et al*. 1997. Income-related inequalities in health: some international comparisons. *Journal of Health Economics* **16**(1): 93–112.
Van Doorslaer E, Wagstaff A, van der Burg H, Christiansen T *et al*. 2000. Equity in the delivery of health care in Europe and the US. *Journal of Health Economics* **19**(5): 553–584.
Van Ourti T. 2003. Socio-economic inequality in ill-health amongst the elderly. Should one use current income or permanent income? *Journal of Health Economics* **22**(2): 187–217.
Wagstaff A. 2002. Poverty and health sector inequalities. *Bulletin of the World Health Organization* **80**(2): 97–105.

DOI: 10.1002/hec