



TI 2006-033/3

Tinbergen Institute Discussion Paper

Does Reporting Heterogeneity bias the Measurement of Health Disparities?

*Teresa Bago d'Uva*¹

Eddy van Doorslaer^{2,3,6}

Maarten Lindeboom^{3,4,6}

Owen O'Donnell^{5,2,6}

*Somnath Chatterji*⁷

¹ University of York;

² Erasmus University Rotterdam;

³ Tinbergen Institute;

⁴ Vrije Universiteit Amsterdam, HEB, IZA;

⁵ University of Macedonia

⁶Netspar;

⁷World Health Organization.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam

Roetersstraat 31

1018 WB Amsterdam

The Netherlands

Tel.: +31(0)20 551 3500

Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam

Burg. Oudlaan 50

3062 PA Rotterdam

The Netherlands

Tel.: +31(0)10 408 8900

Fax: +31(0)10 408 9031

Please send questions and/or remarks of non-scientific nature to driessen@tinbergen.nl.

Most TI discussion papers can be downloaded at <http://www.tinbergen.nl>.

Does reporting heterogeneity bias the measurement of health disparities?

Teresa Bago d'Uva

University of York

Eddy Van Doorslaer

Erasmus University, Netspar & Tinbergen Institute

Maarten Lindeboom

Free University of Amsterdam, HEB, IZA, Netspar & Tinbergen Institute

Owen O'Donnell

University of Macedonia, Erasmus University & Netspar

Somnath Chatterji

World Health Organization

March 2006

Abstract

Heterogeneity in reporting of health by socio-economic and demographic characteristics potentially biases the measurement of health disparities. We use anchoring vignettes to identify reporting heterogeneity in self reports on health for Indonesia, India and China. Correcting for reporting heterogeneity tends to reduce estimated disparities in health by age, sex (not Indonesia), urban/rural and education (not China) and to increase income disparities in health. Overall, while homogeneous reporting by socio-demographic group is significantly rejected, the results suggest that the size of the reporting bias in measures of health disparities is not large.

Keywords: health measurement, vignettes, self-reported health, reporting heterogeneity

JEL Classification: D30, D31, I10, I12

Acknowledgements

Teresa Bago d'Uva is funded by Fundação para a Ciência e Tecnologia, under PhD grant SFRH/BD/10551/2002. The authors thank the WHO for providing access to the MCS data. They acknowledge comments on earlier versions by Andrew Jones and Nigel Rice and by 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.

Corresponding author: Teresa Bago d'Uva, Centre for Health Economics, Alcuin A Block, University of York, York, YO10 5DD, United Kingdom. E-mail: tbd100@york.ac.uk.

1 Introduction

Self-reported health is a convenient and informative instrument, widely used in analyses of health determinants as well as the economic consequences of ill health. Inevitably, there is 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 random. Systematic differences in reporting behaviour are more problematic. For example, measurement of inequality in health will be biased if there are systematic differences in the way in which health is reported across the demographic and socio-economic characteristics against which inequality is being assessed. 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, Rustein 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 (HUI) 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.¹ While this evidence is 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 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, Stabile et al. forthcoming). 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, Murray et al. 2003; King, Murray et al. 2004; Salomon, Tandon et al. 2004). 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, Ozaltin 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. 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, Smith et al. (2004) 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 these groups. Our primary interest is in the degree to which measures of health inequality are biased by reporting heterogeneity in developing countries. 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 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.

2 Data – WHO Multi-Country Survey

The data used in this paper, as in Murray, Ozaltin 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. Ustun, Villanueva et al. (2003) provide a comprehensive report on the goals, design, instrument development and execution of this survey. 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 were 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 health measures.

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).² The dataset 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 dataset contains 7770 observations for Indonesia, 5129 for India and 7156 for China. Table A1 in the Appendix documents the number of observations lost due to item non-response.

2.1 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”, “None”. Table 1 presents the distributions of these self-reported health variables.

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 5 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 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 the Appendix.

TABLE 1

The distributions of the vignette evaluations are presented in Table1. 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 metres 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. Varying degrees of reporting heterogeneity can be seen across the vignettes for all domains and countries. This is the variation we exploit to test for systematic reporting heterogeneity in relation to demographic and socio-economic characteristics and to purge health disparities across such characteristics of reporting bias.

2.2 Socio-demographic variables

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: 15 to 29 years (reference category), 30 to 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).³ The variable $\log(\text{INCOME})$ is the log of monthly household earnings by equivalent adult (in national currencies).^{4 5} Table 2 presents descriptive statistics for the covariates by country.

TABLE 2

3 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.

3.1 Ordered Probit: Homogeneous reporting behaviour

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

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

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.⁶ The observed categorical response of individual i y_i relates to latent health in the following way:

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

$k=0, \dots, K$, $\tau^0 < \tau^1 < \dots < \tau^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. The parameters τ^k , $k=1, \dots, K-1$ 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 baseline threshold. If the covariate effect were the same on all thresholds, labelled ‘parallel shift’ (Hernandez-Quevedo, Jones 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, Jones 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.⁷ 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.

3.2 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 ($\tau_i^k = \tau(x_i \alpha^k)$, $k=1, \dots, K-1$). 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, Murray 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 *response consistency*: individuals classify 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. The second assumption necessary for identification of reporting behaviour via the vignettes is *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” (King, Murray 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 health for a given level of functioning in a single health domain.⁸

3.2.1 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 to represent a fixed level of ability, any association between the 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:

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

The observed vignette ratings y_{ij}^v relate to Y_{ij}^{v*} in the following way:

$$y_{ij}^v = k \Leftrightarrow \tau_i^{k-1} \leq Y_{ij}^{v*} < \tau_i^k \quad (4)$$

$k=0, \dots, K$, $\tau^0 < \tau_i^1 < \dots < \tau_i^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. The cut-points are defined as functions of covariates but are assumed not to vary across different vignettes j for a given health domain, for instance:

$$\tau_i^k = X_i \gamma^k \quad (5)$$

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.⁹

3.2.2 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 response consistency and the vignette equivalence 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:

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

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

$$y_i = k \Leftrightarrow \tau_i^{k-1} \leq Y_i^{s*} < \tau_i^k \quad (7)$$

$k=0, \dots, K$, $\tau^0 < \tau_i^1 < \dots < \tau_i^{K-1} < \tau^K$ 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 $i=1, \dots, N$ and $j=1, \dots, V$. 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.

3.2.3 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 (this is labelled as parallel cut-point shift). We return to this in the next section

4 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 2 models estimated for 6 domains and 3 countries, we do not present all the parameter estimates¹⁰. We first report results on tests for homogeneous reporting behaviour. Next we turn to the quantitative effects of reporting heterogeneity and to what degree reporting heterogeneity biases measures of inequality in health.

4.1 Tests of reporting homogeneity

Table 3 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 respective (groups of) covariates in the 4 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.

TABLE 3

In the final column of Table 3, 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.

4.2 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. lower probabilities of reporting better levels of health. Tables 4 and 5 present the cut-point coefficients of the income and education variables respectively. To save on space, we do not present the coefficients for sex, age and urban but illustrate the direction and magnitude of all effects graphically (Figure 1). 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.

TABLE 4

Reporting behaviour by education level is illustrated by the results presented in Table 5. There is less consistency than there is for income, with more cases of both positive and negative coefficients for a given domain and country and even cases of significant effects in opposite directions. For Indonesia and India, most of the coefficients for the uppermost cut-point are negative, although significance is reached only for *mobility* and *cognition*, indicating that more educated people are more likely to report no difficulties. The opposite is true for China, although there are significant positive effects on the uppermost cut-point only for *cognition*. The results for Indonesia and India are perhaps surprising. They do not support the contention that education raises health expectations. Rather, taken at face value, they suggest that the better educated are more likely to tolerate ill-health. Another possibility is that there is differential capacity by education level in comprehension of the vignette rating exercise. Indeed, Murray, Ozaltin et al. (2003) find that lower educated groups display greater inconsistencies between their ranking of the vignettes and the average ranking. But while this 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 notable that the Chinese sample has considerably higher levels of education than the others and that the direction of the education effect is consistent with what might be hypothesised *a priori*; that is, health expectations rising with education. It may be that the vignette exercise was more comprehensible for the Chinese sample.

TABLE 5

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.

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 4, 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 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*. As regards education (second row), there are again few quantitatively important effects. For Indonesia, the highest education category has a probability of reporting no pain that is 5% greater than that of the lowest group. For China, the direction of the education effect on the reporting of pain is reversed but is again about 5% in relative magnitude. Otherwise, the relative differences in reporting between high and low education groups barely exceed 2%.

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, Robinson et al. 1998). Generally, the youngest age group (15-29 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 may be explained by a greater capacity of elderly respondents, given their greater life experiences, to empathise with the vignette descriptions. If this were the case, it would imply violation of the assumption of *vignette equivalence*. Murray, Ozaltin et al. (2003) claim some support for this proposition in the WHO-MCS data, finding that the degree of correlation between an individual's ranking of vignettes and the average ranking is increasing with age.

The general picture that emerges from the figures is that the reporting effects of income and education are relatively small compared to the age, sex and urban/rural effects. This suggests that income and education related reporting heterogeneity may not have a large impact on measures of socio-economic inequality in health. We now turn to this issue.

4.3 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.

TABLES 6 & 7

The income and education coefficients in the health equations, (1) and (6), are shown in Tables 6 and 7 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 underestimated if reporting heterogeneity by income is not accounted for (in 14 of the 18 cases). In India, the income coefficient in the *pain* function becomes significant after the purging of reporting heterogeneity. But in Indonesia, significance is lost from the income coefficient in the *self-care* function.

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 7). The vignette adjustment for reporting bias leads to a decrease in 13 of the 18 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 one characteristic, 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 of each characteristic holding the others 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 only that for which the health disparity is computed. The results for income, education, sex, age and the urban-rural differences are depicted in Figure 2

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 noticeable 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 but not substantially so. For Indonesia and India there is an apparent urban health advantage that is reduced after purging reporting bias. The adjustment is greatest for *pain*, particularly in Indonesia, where we find an 8% point difference between the two bars. For China, we always find the probability of very good health to be lower in urban areas and this disparity is increased in four of the six domains after purging reporting bias. The gender gradient varies most across the different health dimensions and less so across countries. Where there is a clear male advantage (*cognition* and *pain*), purging reporting differences reduces the disparity in India and China and raises it in Indonesia. Finally, as

expected, we find large and positive age effects for all three countries. The correction for reporting heterogeneity generally reduces the age gradient and in some cases the changes are substantial.

Figure 2 illustrates the effect of purging reporting heterogeneity from partial associations between covariates 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, Wagstaff et al. 1997) for the predicted probability of reporting 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). But the effect of purging reporting heterogeneity from both the partial and total correlations is similar. There is a slight upward adjustment to the disparities in most cases and a marked increase in health inequality by income only in the domains of *pain* and *affect* in China.

5 Conclusion

In this paper we have investigated whether there is heterogeneity in health reporting and whether and how this affects the measurement of socio-demographic 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 domains, include their assessments of vignette descriptions of the health domains. Such data allow for the estimation of hierarchical ordered probit models which 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 socio-economic and demographic characteristics.

The hypothesis of homogeneous reporting across all socio-demographics is rejected for all countries and health domains. There is variation across countries and health domains in the rejection of homogeneous reporting with respect to individual socio-demographic characteristics. Homogeneity tends to be most consistently and decisively rejected across urban/rural, income and age differences and less consistently across sex and 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 and so obscuring this variation, 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. Reporting of health varies most across age, sex and urban/rural differences and less by education and income. Correcting for reporting heterogeneity tends to reduce disparities in health by age, sex (not Indonesia), urban/rural and education (not China) and to increase income disparities in health. Overall, while homogeneous reporting is significantly rejected, our results suggest that the size of the reporting bias in measures of health disparities is not large.

Some of these results might be considered surprising. Previous evidence suggests that the elderly have lower expectations of health and on this basis one would expect age disparities in health to increase after purging reporting heterogeneity. As mentioned above, a possible explanation for our contradictory result is that the assumption of *vignette equivalence* does not hold with respect to age. Older respondents may comprehend the level of functioning/pain/distress that a vignette is intended to describe differently from younger respondents. Recognising conditions described in the vignettes from their own experience, older respondents may be more appreciative of the consequent constraints on health while younger respondents, lacking exposure to such conditions, may be more dismissive. If this were true, it would invalidate the vignette approach. We have no evidence to support such a conclusion. In fact, it is mainly for *pain* that we find the elderly to have lower thresholds and, as noted above, the evidence on whether perceptions of pain vary with age is ambiguous (Gibson and Helme 2001). In the domain of *mobility*, Murray,

Ozaltin et al. (2003) find reporting behaviour consistent with health expectations falling with age for six countries covered by the WHO-MCS. But they also find that ranking of vignettes varies systematically with age and education, suggesting that comprehension of the described levels of health varies with these characteristics. Further evaluation of the validity of the vignette approach is clearly required in the form of experiments designed to directly test the assumption of *vignette equivalence* and that of *response consistency*.

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 interpersonally comparable. Any attempt to measure health inequality must assume that it is. 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 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 removed. There is no heterogeneity deriving from differential weighting of each dimension of 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.

References

- Baker, J. L. and J. Van der Gaag (1993). Equity in health care and health care financing: Evidence from five developing countries. Equity in the finance and

- delivery of health care. E. Van Doorslaer, A. Wagstaff and F. Rutten. Oxford, Oxford University Press.
- Baker, M., M. Stabile and C. Deri (forthcoming). “What do self-reported, objective measures of health measure?” Journal of Human Resources.
- Benitez-Silva, H., M. Buschinski, H. M. Chan, S. Cheidvasser and J. Rust (1999). “How large is the bias in self-reported disability?” Journal of Applied Econometrics **19**(6): 649-670.
- Bound, J. (1991). “Self reported versus objective measures of health in retirement models.” Journal of Human Resources **26**: 107-137.
- Disney, R., C. Emerson and M. Wakefield (forthcoming). “Ill-health and retirement in Britain: A panel data based analysis.” Journal of Health Economics.
- Gibson, S. J. and R. D. Helme (2001). “Age-related differences in pain perception and report.” Clinical Geriatric Medicine **17**: 433-456.
- Gwatkin, D. R., S. Rustein, K. Johnson, R. Pande and A. Wagstaff (2000). Socioeconomic differences in health, nutrition and population. World Bank Health, Nutrition and Population Discussion Paper. Washington DC.
- Hernandez-Quevedo, C., A. M. Jones and N. Rice (2004). Reporting bias and heterogeneity in self-assessed health. Evidence from the British Household Panel Survey. ECuity III Working Papers. York.
- Idler, E. L. (1993). “Age differences in self-assessments of health: age changes, cohort differences, or survivorship?” Journal of Gerontology **48**(6): S289-300.
- Kakwani, N., A. Wagstaff and E. van Doorslaer (1997). “Socioeconomic inequalities in health: measurement, computation and statistical inference.” Journal of Econometrics **77**(1): 87-104.
- Kapteyn, A., J. Smith and A. van Soest (2004). Self-reported work disability in the US and the Netherlands. RAND Working Paper. Santa Monica.
- Kerkhofs, M. J. M. and M. Lindeboom (1995). “Subjective health measures and state dependent reporting errors.” Health Economics **4**: 221-235.
- King, G., C. J. L. Murray, J. Salomon and A. Tandon (2004). “Enhancing the validity and cross-cultural comparability of measurement in survey research.” American Political Science Review **98**(1): 184-91.
- Kreider, B. (1999). “Latent work disability and reporting bias.” Journal of Human Resources **34**(4): 734-769.

- Lindeboom, M. and E. van Doorslaer (2004). "Cut-point shift and index shift in self-reported health." Journal of Health Economics **23**(6): 1083-1099.
- Mathers, C. D. and R. M. Douglas (1998). Measuring progress in population health and well-being. Measuring progress: is life getting better? R. Eckersley. Collingwood, CSIRO Publishing: 125-155.
- Murray, C. J. L. (1996). Epidemiology and morbidity transitions in India. Health, poverty and development in India. M. Dasgupta, C. L.C. and T. N. Krishnan. Delhi, Oxford University Press: 122-147.
- Murray, C. J. L., E. Ozaltin, A. Tandon, J. Salomon, R. Sadana and S. Chatterji (2003). Empirical evaluation of the anchoring vignettes approach in health surveys. Health systems performance assessment: debates, methods and empiricism. C. J. L. Murray and D. B. Evans. Geneva, World Health Organization.
- Pudney, S. and M. Shields (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, J. L., M. E. Robinson, E. A. Wise, C. D. Myers and R. B. Fillingim (1998). "Sex differences in the perception of noxious experimental stimuli: a meta-analysis." Pain **74**: 181-187.
- Salomon, J., A. Tandon, C. J. L. Murray and W. H. S. P. S. C. Group (2004). "Comparability of self-rated health: Cross sectional multi-country survey using anchoring vignettes." British Medical Journal **328**: 258.
- Sen, A. (2002). "Health: perception versus observation." British Medical Journal **324**: 860-1.
- Stern, S. (1989). "Measuring the effect of disability on labor force participation." Journal of Human Resources **24**: 361-395.
- Tandon, A., C. J. L. Murray, J. A. Salomon and G. King (2003). Statistical models for enhancing cross-population comparability. Health systems performance assessment: debates, methods and empiricisms. C. J. L. Murray and D. B. Evans. Geneva, World Health Organization: 727-746.
- Terza, J. V. (1985). "Ordinal probit: a generalization." Communications in Statistics **14**(1): 1-11.
- Unruh, A. M. (1996). "Gender variations in clinical pain experience." Pain **65**: 123-167.

- Ustun, T. B., M. Villanueva, L. Benib, C. Celik, R. Sadana, N. B. Valentine, J. P. Ortiz, A. Tandon, J. Salomon, C. Yang, W. J. Xie, E. Ozaltin, C. D. Mathers and C. J. L. Murray (2003). WHO Multi-Country Survey Study on Health and Responsiveness 2001-2. Health systems performance assessment: debates, methods and empiricisms. C. J. L. Murray and D. B. Evans. Geneva, World Health Organization: 762-796.
- Van Doorslaer, E. and U.-G. Gerdtham (2003). “Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data.” Social Science & Medicine **57**(9): 1621-1629.
- Wagstaff, A. (2002). “Poverty and health sector inequalities.” Bulletin of the World Health Organization **80**(2): 97-105.

Notes

¹ 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, Buschinski et al. 1999; Kreider 1999; Disney, Emerson et al. forthcoming). 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.

² 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.

³ 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.

⁴ 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 $(\text{number of adults in household} + 0.5 \times \text{number of children in household})^{0.75}$).

⁵ 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 $\log(\text{INCOME})$. The latter is preferred for ease of presentation.

⁶ In this application, we fix these terms in a different way. See below.

⁷ We are grateful to Andrew M. Jones who made this point to us in private communication.

⁸ Murray, Ozaltin et al. (2003) attempt to test vignette equivalence in the WHO-MCS data by testing for systematic differences in the ranking of vignettes in relation to age, sex, education and questionnaire characteristics.

⁹ 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 so, given the substantially greater computational cost, we decided not to allow for unobservable heterogeneity within the full HOPIT model.

¹⁰ These are available from the authors upon request.

Appendix

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.

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.

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.

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



Figure 2: Relative probabilities of being in very good health by socio-economic group, before and after adjustment

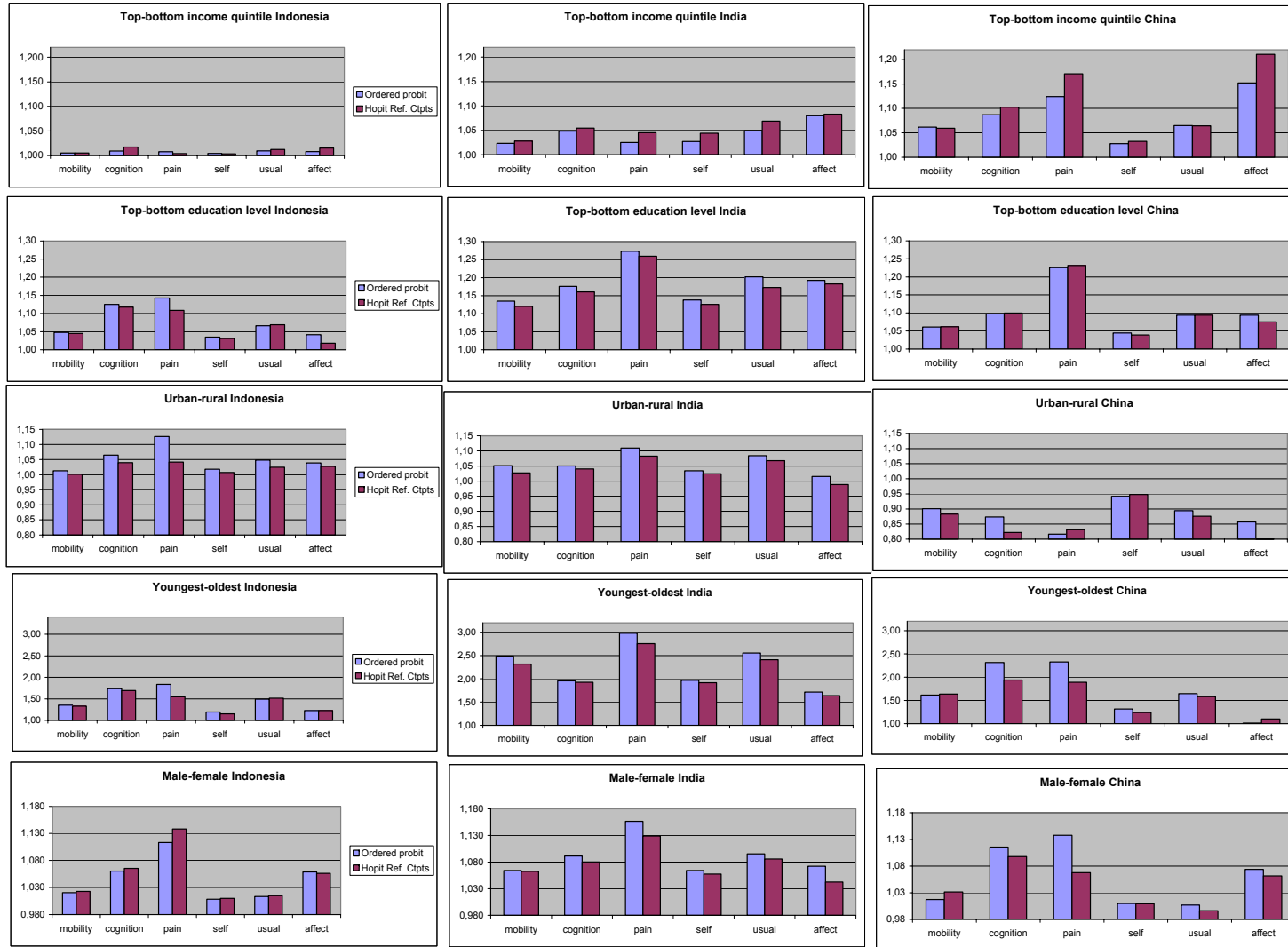


Figure 3: Age-Sex standardised concentration indices of predicted probability of reporting very good health, before and after adjustment

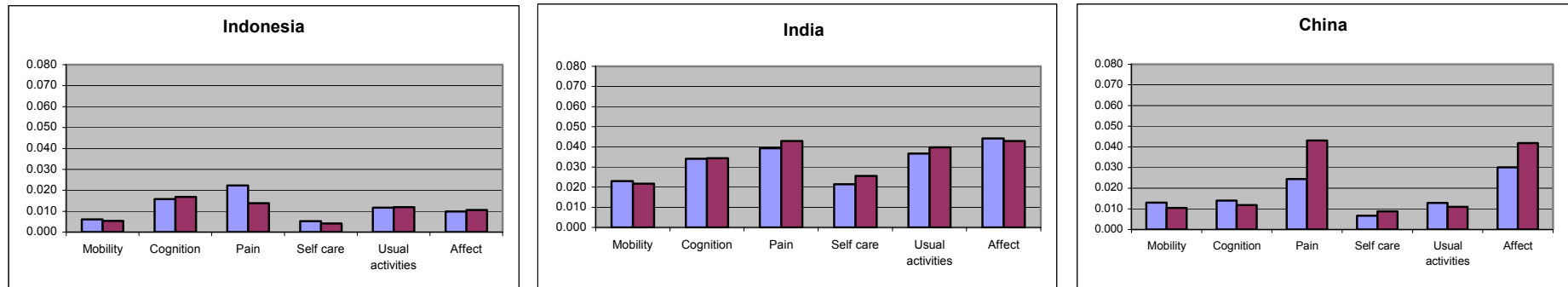


Table 2: Descriptive statistics of covariates

Variables	Indonesia		India		China	
	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(Inc)	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	

Table 3: Log-Likelihood ratio tests of homogeneity and parallel cut-point shift: p-values

Indonesia							
	Homogeneity						Parallel cut-point shift
	All	Female	Age	Educ	Log(Inc)	Urban	All
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							
	Homogeneity						Parallel cut-point shift
	All	Female	Age	Educ	Log(Inc)	Urban	All
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							
	Homogeneity						Parallel cut-point shift
	All	Female	Age	Educ	Log(Inc)	Urban	All
Mobility	0.000	0.000	0.000	0.557	0.369	0.069	0.000
Cognition	0.000	0.471	0.014	0.058	0.003	0.000	0.000
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	0.000

Table 4: 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 (1.829)	-0.003 (-0.385)	-0.005 (-0.684)	0.000 (-0.014)	-0.011 (-0.691)	-0.058 (-4.569)	-0.017 (-1.364)	0.029 (1.921)	0.009 (0.806)	0.003 (0.379)	0.011 (1.374)	-0.006 (-0.734)
Cognition	0.007 (0.499)	0.007 (0.769)	0.034 (4.226)	0.029 (3.135)	0.007 (0.287)	-0.020 (-1.366)	0.041 (2.848)	0.011 (0.586)	0.024 (1.482)	0.018 (1.692)	0.038 (3.848)	0.017 (1.466)
Pain	-0.033 (-2.956)	-0.001 (-0.158)	-0.003 (-0.29)	-0.008 (-0.594)	0.000 (0.02)	-0.014 (-1.062)	0.006 (0.428)	0.059 (2.247)	0.038 (2.74)	0.042 (4.204)	0.032 (3.228)	0.047 (2.976)
Self-care	-0.063 (-5.67)	-0.014 (-1.502)	0.004 (0.413)	-0.012 (-0.974)	0.020 (1.002)	0.014 (1.076)	0.045 (3.256)	0.052 (2.866)	0.071 (4.295)	0.080 (6.928)	0.065 (6.228)	0.035 (2.611)
Usual	-0.016 (-1.346)	0.010 (1.231)	0.026 (3.248)	0.018 (1.851)	0.043 (2)	-0.003 (-0.254)	0.054 (3.734)	0.049 (2.531)	-0.015 (-0.944)	-0.011 (-0.952)	0.026 (2.59)	-0.005 (-0.409)
Affect	-0.012 (-0.868)	-0.017 (-1.753)	0.013 (1.309)	0.037 (3.449)	0.030 (1.385)	-0.017 (-1.025)	0.037 (2.208)	0.000 (-0.009)	0.019 (1.415)	0.028 (2.53)	0.037 (3.821)	0.044 (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.

Table 5: 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 (0.407)	0.000 (-0.014)	0.023 (0.85)	-0.045 (-1.559)	-0.074 (-1.168)	-0.059 (-1.189)	-0.082 (-1.67)	-0.121 (-2.122)	0.020 (0.324)	-0.007 (-0.149)	-0.001 (-0.03)	0.043 (0.874)
	Educ3	-0.033 (-0.803)	0.034 (1.183)	0.017 (0.614)	-0.109 (-3.775)	-0.091 (-1.257)	0.059 (1.074)	-0.037 (-0.664)	-0.118 (-1.826)	-0.031 (-0.547)	-0.076 (-1.699)	-0.029 (-0.685)	0.035 (0.75)
	Educ4	0.101 (2.153)	0.079 (2.313)	0.019 (0.592)	-0.078 (-2.301)	-0.015 (-0.285)	0.038 (0.891)	0.012 (0.279)	-0.083 (-1.687)	-0.080 (-1.276)	-0.096 (-1.977)	-0.045 (-0.988)	0.025 (0.492)
Cognition	Educ2	0.026 (0.523)	0.000 (0)	0.020 (0.624)	-0.030 (-0.785)	0.052 (0.653)	0.081 (1.582)	0.004 (0.077)	0.054 (0.814)	0.063 (0.786)	0.064 (1.173)	0.027 (0.529)	0.082 (1.382)
	Educ3	-0.041 (-0.772)	-0.010 (-0.305)	-0.091 (-2.742)	-0.131 (-3.302)	0.092 (1.071)	-0.010 (-0.178)	-0.075 (-1.358)	-0.136 (-1.966)	0.011 (0.146)	0.006 (0.118)	0.010 (0.203)	0.136 (2.4)
	Educ4	0.128 (2.337)	0.097 (2.638)	0.003 (0.072)	-0.049 (-1.094)	-0.045 (-0.652)	-0.008 (-0.187)	-0.025 (-0.575)	-0.083 (-1.499)	-0.049 (-0.587)	-0.066 (-1.178)	-0.063 (-1.209)	0.047 (0.77)
Pain	Educ2	-0.025 (-0.559)	0.001 (0.022)	0.040 (1.179)	-0.016 (-0.359)	-0.023 (-0.329)	-0.065 (-1.228)	-0.160 (-2.685)	-0.241 (-2.418)	0.060 (0.813)	-0.013 (-0.257)	0.010 (0.194)	0.025 (0.288)
	Educ3	-0.007 (-0.142)	0.052 (1.549)	0.029 (0.807)	0.004 (0.078)	-0.055 (-0.699)	0.029 (0.509)	-0.155 (-2.394)	-0.159 (-1.353)	0.040 (0.575)	-0.134 (-2.722)	-0.085 (-1.665)	-0.102 (-1.231)
	Educ4	0.028 (0.55)	0.013 (0.347)	0.023 (0.574)	-0.086 (-1.615)	0.106 (1.903)	-0.005 (-0.118)	-0.105 (-2.146)	-0.066 (-0.739)	0.031 (0.412)	-0.167 (-3.147)	-0.108 (-1.993)	0.083 (0.91)
Self-care	Educ2	0.075 (1.455)	0.042 (1.274)	0.014 (0.421)	0.013 (0.307)	0.014 (0.201)	-0.066 (-1.264)	-0.073 (-1.271)	-0.134 (-1.844)	0.134 (1.558)	-0.044 (-0.743)	0.044 (0.803)	-0.041 (-0.542)
	Educ3	-0.018 (-0.326)	0.011 (0.309)	-0.029 (-0.806)	-0.085 (-1.989)	-0.019 (-0.234)	-0.027 (-0.481)	0.012 (0.185)	-0.023 (-0.275)	0.062 (0.752)	-0.097 (-1.727)	-0.065 (-1.242)	-0.094 (-1.307)
	Educ4	0.103 (1.713)	0.014 (0.348)	0.003 (0.066)	-0.098 (-2.072)	-0.046 (-0.79)	-0.052 (-1.238)	-0.032 (-0.674)	-0.058 (-0.929)	-0.056 (-0.627)	-0.123 (-2.034)	-0.051 (-0.907)	-0.126 (-1.631)
Usual	Educ2	-0.021 (-0.416)	0.033 (1.103)	0.073 (2.315)	-0.005 (-0.117)	0.138 (1.937)	0.071 (1.527)	0.109 (2.072)	0.103 (1.454)	-0.098 (-1.22)	-0.029 (-0.51)	0.030 (0.595)	0.035 (0.579)
	Educ3	0.047 (0.916)	0.069 (2.198)	0.001 (0.03)	-0.056 (-1.321)	0.095 (1.193)	-0.086 (-1.705)	-0.096 (-1.757)	-0.163 (-2.255)	-0.059 (-0.785)	-0.039 (-0.733)	-0.005 (-0.099)	0.043 (0.739)
	Educ4	-0.041 (-0.7)	0.075 (2.158)	0.038 (1.027)	0.014 (0.287)	-0.029 (-0.46)	-0.072 (-1.792)	-0.066 (-1.516)	-0.108 (-1.841)	-0.122 (-1.48)	-0.138 (-2.363)	-0.061 (-1.158)	0.021 (0.345)
Affect	Educ2	-0.022 (-0.425)	0.025 (0.645)	-0.002 (-0.055)	-0.067 (-1.515)	-0.010 (-0.116)	0.005 (0.078)	-0.052 (-0.819)	-0.080 (-0.93)	-0.057 (-0.747)	0.000 (0.002)	-0.009 (-0.16)	0.009 (0.14)
	Educ3	-0.115 (-2.097)	-0.052 (-1.314)	-0.065 (-1.722)	-0.139 (-3.102)	-0.015 (-0.17)	-0.067 (-0.924)	-0.166 (-2.436)	-0.142 (-1.509)	0.027 (0.376)	-0.009 (-0.156)	-0.087 (-1.632)	-0.082 (-1.327)
	Educ4	-0.012 (-0.198)	0.029 (0.658)	-0.060 (-1.378)	-0.122 (-2.356)	0.049 (0.678)	-0.088 (-1.498)	-0.037 (-0.679)	-0.052 (-0.694)	-0.006 (-0.082)	0.021 (0.327)	-0.063 (-1.081)	-0.018 (-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.

Table 6: Estimated coefficients of LOG(INCOME) before and after adjustment

	Indonesia		India		China	
	Before	After	Before	After	Before	After
SAH	0.019 (1.926)	- -	0.052 (3.649)	- -	0.085 (8.586)	- -
Mobility	0.054 (2.678)	0.054 (2.524)	0.055 (2.166)	0.065 (2.333)	0.137 (8.363)	0.134 (7.365)
Cognition	0.034 (2.28)	0.063 (3.699)	0.108 (3.987)	0.121 (3.923)	0.089 (6.929)	0.109 (6.71)
Pain	0.013 (1.155)	0.007 (0.452)	0.042 (1.524)	0.074 (2.249)	0.099 (7.223)	0.141 (7.63)
Self-care	0.062 (2.356)	0.051 (1.787)	0.074 (2.968)	0.124 (4.227)	0.185 (6.229)	0.225 (6.899)
Usual activities	0.066 (3.297)	0.085 (3.886)	0.110 (4.142)	0.158 (5.168)	0.137 (7.394)	0.138 (6.523)
Affect	0.037 (2.065)	0.068 (3.321)	0.179 (5.56)	0.183 (4.996)	0.107 (8.327)	0.148 (9.449)

Note: t-ratios in parentheses

Table 7: Estimated coefficients of Education dummies before and after adjustment

		Indonesia		India		China	
		Before	After	Before	After	Before	After
SAH	Educ2	0.145 (4.095)	-	0.092 (1.691)	-	0.166 (3.124)	-
	Educ3	0.252 (6.813)	-	0.293 (4.807)	-	0.148 (2.938)	-
	Educ4	0.282 (6.673)	-	0.378 (8.164)	-	0.207 (3.804)	-
Mobility	Educ2	0.282 (3.154)	0.249 (2.656)	0.268 (2.796)	0.162 (1.52)	0.206 (2.423)	0.241 (2.542)
	Educ3	0.344 (3.498)	0.251 (2.437)	0.321 (2.884)	0.235 (1.902)	0.334 (4.093)	0.355 (3.914)
	Educ4	0.627 (4.955)	0.563 (4.248)	0.530 (6.108)	0.467 (4.839)	0.288 (3.258)	0.299 (3.041)
Cognition	Educ2	0.157 (2.695)	0.136 (2.032)	0.282 (2.724)	0.331 (2.842)	0.221 (3.36)	0.281 (3.428)
	Educ3	0.269 (4.236)	0.149 (2.042)	0.674 (5.351)	0.554 (3.97)	0.343 (5.477)	0.439 (5.581)
	Educ4	0.514 (6.567)	0.474 (5.345)	0.658 (6.959)	0.586 (5.513)	0.210 (3.112)	0.224 (2.652)
Pain	Educ2	0.001 (0.032)	0.004 (0.066)	0.167 (1.626)	-0.020 (-0.161)	0.246 (3.46)	0.257 (2.698)
	Educ3	0.127 (2.81)	0.141 (2.322)	0.515 (4.29)	0.395 (2.672)	0.402 (5.948)	0.286 (3.155)
	Educ4	0.236 (4.488)	0.181 (2.623)	0.731 (7.938)	0.670 (5.871)	0.384 (5.244)	0.409 (4.1)
Self-care	Educ2	0.330 (2.775)	0.345 (2.747)	0.260 (2.709)	0.140 (1.222)	0.404 (2.687)	0.378 (2.299)
	Educ3	0.436 (3.29)	0.353 (2.529)	0.211 (1.918)	0.177 (1.352)	0.460 (3.183)	0.375 (2.375)
	Educ4	0.602 (3.682)	0.517 (3.009)	0.704 (7.617)	0.632 (5.817)	0.710 (4.407)	0.600 (3.411)
Usual activities	Educ2	0.195 (2.331)	0.204 (2.234)	0.222 (2.219)	0.320 (2.789)	0.329 (3.455)	0.359 (3.336)
	Educ3	0.453 (4.765)	0.406 (3.938)	0.367 (3.119)	0.217 (1.641)	0.515 (5.639)	0.545 (5.258)
	Educ4	0.508 (4.542)	0.526 (4.365)	0.778 (8.182)	0.659 (6.081)	0.440 (4.441)	0.447 (3.979)
Affect	Educ2	0.068 (0.991)	0.014 (0.181)	0.246 (2.021)	0.190 (1.377)	0.185 (2.687)	0.187 (2.11)
	Educ3	0.154 (2.097)	0.028 (0.337)	0.603 (4.243)	0.469 (2.93)	0.227 (3.494)	0.146 (1.749)
	Educ4	0.187 (2.203)	0.078 (0.803)	0.686 (6.344)	0.632 (5.115)	0.139 (1.975)	0.111 (1.226)

Note: t-ratios in parentheses

Appendix Tables

Table A1: 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:

<i>mobility</i>	2	3	3
<i>self-care</i>	2	3	1
<i>usual</i>	6	10	102
<i>pain</i>	8	9	26
<i>affect</i>	2	17	9
<i>cognition</i>	7	13	106