

The measurement and comparison of health system responsiveness

Nigel Rice, Silvana Robone, Peter C. Smith
Centre for Health Economics, University of York
November 2007

Paper presented to HESG January 2008
Preliminary and not for quotation without the permission of authors

Abstract

Measuring the performance of health systems has become a key tool in aiding decision makers to describe, analyze, compare and ultimately improve the delivery and outcomes achieved by a system. The World Health Organization's (WHO) framework for assessing performance includes three intrinsic goals of health systems, namely health improvement, fairness in financial contribution and responsiveness to user preferences. Broadly speaking health system responsiveness can be defined as the way in which individuals are treated and the environment in which they are treated, encompassing the notion of patient experience. Perhaps the most ambitious attempt to implement a cross-country comparative instrument aimed at measuring health system performance is the World Health Survey (WHS). The modules on responsiveness and health ask respondents to rate their experiences using a 5-point categorical scale (e.g. "very good" to "very bad"). A common problem with self-reported data is that individuals, when faced with the instrument, are likely to interpret the meaning of the response categories in a way that systematically differs across populations or population sub-groups. In such cases the ordinal response categories will not be cross-population comparable since they will not imply the same underlying level of the construct. Recently the method of anchoring vignettes has been promoted as a means for controlling for systematic differences across socio-economic groups in preferences, expectations and norms when responding to survey questions.

This paper applies the method of anchoring vignettes to adjust survey reports of responsiveness for reporting heterogeneity. We present preliminary results for a selected number of domains and countries to illustrate the approach and find systematic reporting by income and education, but not by age and gender. Further analysis will extend the method to a larger set of countries and domains to investigate more fully the application of the approach for international comparative analysis of health system performance.

Acknowledgements

This research is funded by the Economic and Social Research Council through their Public Services Programme (RES-166-25-0038). We are grateful to the World Health Organization, and in particular, Tim Evans, Somnath Chatterji, Amit Prasad, Emese Verdes, for providing access to the World Health Survey and advice in understanding the survey module on health system responsiveness. We are further grateful to Teresa Bago d'Uva for providing the code to estimate the HOPIT model.

Introduction

Increasingly patients' views and opinions are being recognized as an essential means for assessing the provision of health services, to stimulate quality improvements and more recently, in measuring health systems performance. While traditionally, patients' views were sought on the quality of care provided and satisfaction with health services, in the context of performance assessment the concept of responsiveness has been promoted as a more desirable measure by which health systems can be judged. Responsiveness relates to a system's ability to respond to the legitimate expectations of potential users about non-health enhancing aspects of care and together with health and fairness of financial contribution has been identified as an intrinsic goal of health system performance (Murray and Frenk, 2000).

In broad terms, health system responsiveness has been defined as the way in which individuals are treated and the environment in which they are treated and importantly, encompasses the notion of an individual's experience of contact with the health system (Valentine et al, 2003a). These experiences are measured along a number of domains which can be classified into two broad dimensions, namely respect for human rights and client orientation. Human rights include concepts such as respecting patient autonomy and dignity, while client orientation focuses on aspects that are commonly expressed as hotel facilities, for example, the quality of basic amenities.

A clear purpose for outcome measurement is to enable institutions to compare and contrast their performance to that of others, including at a macro level, to performance obtained in other countries. The challenge of how appropriately to compare across institutional settings and populations is a central feature of comparative work for all public services. A fundamental problem, however, recognized by Blendon et al. (2003), is that studies aimed at comparative inference have rarely taken into consideration possible variations in cultural expectations that might impact on reporting behaviour. To this end, effort has been placed in producing more objective measures of responsiveness and developing instruments that are relevant across cultural settings (Murray et al, 2003). This is, in itself, however, unlikely to ensure response comparability if individuals, when faced with survey questions about the functioning of health systems, systematically interpret the meaning of the available response categories differentially across population sub-groups (Sadana et al., 2002). Where this is the case then a fixed level of underlying responsiveness is unlikely to be rated equally across sub-groups of interest (see Tandon et al., 2003).

The degree to which self-reported survey data are comparable across individuals, socio-economic groups or populations has been debated extensively, usually with regard to measures of health status (for example, Jürges, 2007, Kapteyn et al., 2007; Bago d'Uva et al., 2007; Lindeboom and van Doorslaer, 2004; Iburg et al., 2002; Manderbacka, 1998; Kempen et al., 1996; Kerkhofs and Lindeboom, 1995; Idler and Kasl, 1995). Similar concerns apply to self-reported data on health systems responsiveness where the characteristics of the systems and cultural norms regarding the use and experiences of public services are likely to predominate. The extent to which self-reported information on health system responsiveness is prone to reporting differences is an empirical question and one which this paper attempts to address.

The method of anchoring vignettes has been promoted as a means for controlling for systematic differences in preferences and norms when responding to survey questions (for example, see Salomon et al. (2004)). Vignettes represent descriptions of fixed levels of a latent construct – such as responsiveness - and accordingly any systematic variation across individuals in the rating

of the vignettes can be attributed to reporting behaviour. Systematic reporting behaviour results from individuals applying different response scales both to the hypothetical vignettes and to the reporting of their actual experiences of health services. Because individuals are asked to evaluate these hypothetical cases in the same way as they evaluate their own experience of the health care system, responses to the vignettes allow the researcher to model the response scales as a function of the characteristics of respondents. This information can then be used subsequently to adjust the self-reported data of a respondent's own experiences of health service contact. A number of studies have promoted the vignette approach and made use of what has been termed the HOPIT model to adjust self-reports. Again, these have predominantly been applied to self-reported data on health status (for example see, Iburg et al. (2002), Tandon et al (2003), Murray et al. (2003), King et al. (2004), Kapteyn et al. (2007), Bago d'Uva et al. (2007)). Recently, Valentine et al. (2003b) have considered the role of sex, age, years of education and reported health status on reporting behaviour applied to the WHO-MCSS responsiveness module while Puentes Rosas et al. (2006) consider age, sex, education and type of health care provider using a survey of user satisfaction in Mexico.

The purpose of this paper is to explore the utility of using information from vignettes to adjust self-reports of health system responsiveness to assess the level of reporting heterogeneity, its impact on responsiveness outcomes and the ability to adjust these to achieve cross-population comparability. We illustrate the use of the methods by exploring information on vignettes from the World Health Survey on the three countries: Mexico, India and the Philippines. While ultimately one would wish to adjust reports of responsiveness to aid cross-country comparison, our ambitions for this paper are more modest and we illustrate the methods across socio-economic groups within countries only.

Data – The World Health Survey (WHS)

The most ambitious attempt to date to measure and compare health systems responsiveness is the World Health Survey (WHS). The WHS is an initiative launched by the WHO in 2001 aimed at strengthening national capacity to monitor critical health outputs and outcomes through the fielding of a valid, reliable and comparable household survey instrument (see Üstün et al., 2003). Seventy countries participated in the WHS 2002-2003, consisting of a combination of 90-minute in-household interviews (53 countries), 30-minute face-to-face interviews (13 countries) and computer assisted telephone interviews (4 countries). All surveys were drawn from nationally representative frames with known probability resulting in sample sizes of between 600 and 10000 across the countries surveyed. Samples have undergone extensive quality assurance procedures, including the testing of the psychometric properties of the responsiveness instrument.

The measurement of responsiveness is obtained by asking respondents to rate their most recent experience of contact with the health system within each of eight domains. The domains cover aspects of responsiveness valued highly by individuals in their contact with health systems, including autonomy, choice, clarity of communication, confidentiality of personal information, dignity, prompt attention, quality of basic amenities and access to family and community support. Definitions of these domains together with examples of the questions asked of respondents are provided in Appendix 1. For each domain respondents were asked up to 2 questions about their experiences of contact with health systems. The response categories available to respondents are “very good”, “good”, “moderate”, “bad” and “very bad”. Accordingly, responsiveness is viewed as a multidimensional concept, with each domain measured as a categorical variable for which there is an assumed underlying latent scale.

The WHS further contains a number of vignettes describing the experiences of hypothetical individuals within each of the eight domains. The vignettes have been divided into four sets (Set A-D), each set containing 5 vignettes for each item present across two domains (all domains contain two items, with the exception of “Choice”, which contains only one item).¹ Due to the constraints of interview length, each respondent in the survey rated the vignettes present in only one of the sets. Therefore, each vignette has been rated by approximately 25% of survey respondents. Survey respondents are asked to rate the responsiveness of a sample of vignettes using the same response scale as the one used for the rating of their own experiences. Examples of the vignettes are provided in Appendix 2.

Variables available in the WHS on individual characteristics include age, gender, level of education and income. Level of education is measured as both a categorical variable containing 7 categories representing, for example, ‘primary school completed’, ‘secondary school completed’ to ‘post graduate degree completed’ and a continuous variable measuring the number of years in education. Gender is a dummy variable that assumes value equal to 1 if the respondent is a woman, 0 if a man. Income is a categorical variable indicating the quintiles of the distribution of household permanent income, where 1 represents respondents in the lowest income quintile and 5 those in the highest income quintile. Permanent income is a derived variable based on household assets (Ferguson et al., 2003). The approach to its measurement uses information on the physical assets owned by the households to estimate permanent income. Some assets are measured by discrete variables (i.e. the number of chairs or tables in the home), others by dichotomous variables which take value 1 if the household have access to some services (i.e. electricity) or goods (i.e. bicycle or refrigerator). In cross country comparisons, the use of the simple physical assets as proxy of the permanent income could be problematic since the same bundle of goods may map to different levels of permanent income in different countries. The approach used to define permanent income uses a variant of the HOPIT model described in our paper in an attempt to enhance comparability of the income measure (Ferguson et al., 2003). Descriptive statistics for the set of explanatory variables are provided in Table 1.

For the analysis that follows we concentrate on three countries within the WHS; Mexico, India and the Philippines. These countries have been selected to illustrate the extent of reporting heterogeneity, the methodological approach and the ability of vignette data to control for differential reporting by socio-economic group. The countries are taken from three of the four macro-geographical areas as defined by the WHO and all satisfy well the set of psychometric properties for the response module of the WHS and hence have desirable survey properties. Further, the sample size for Mexico is far greater than for other countries increasing the scope and precision of analysis.

Empirical approach

Reporting heterogeneity

The reporting of responsiveness provides an ordered categorical variable which is assumed to be a discrete representation of some underlying latent responsiveness scale. If it is assumed that individuals map the latent scale to the response categories in a consistent way, irrespective of their characteristics or circumstances, then we observe homogeneous reporting behaviour and the

¹ Set A contains vignettes for “Dignity” and “Prompt Attention”, set B for “Communication” and “Quality of Basic Amenities”, set C for “Confidentiality” and “Choice”, and set D for “Social Support” and “Autonomy”.

standard ordered probit estimator would provide an appropriate method to model such data. In contrast, reporting heterogeneity arises when individuals differ in their mapping of the latent construct to the response categories. It is natural to think of good or poor system performance to mean different things to different people and accordingly individuals might attach very different interpretations to the response categories. Systematic variation in reporting heterogeneity can be examined in relation to measured attributes of individuals such as their socio-economic status. For example, income has been shown to be a determinant of reporting heterogeneity in self-reported general health status such that more wealthy individuals have higher expectations of health and hence down report health status compared to less wealthy counterparts (Bago d'Uva, 2007).

The Hierarchical Ordered Probit Model (HOPIT)

The ordered probit model assumes homogeneous reporting across individuals that is reflected in the constant cut-points that determine the mapping from the underlying construct to the response categories observed. If this assumption does not hold, and we observe heterogeneous reporting, the use of the ordered probit model will lead to biased estimates of the impact of explanatory regressors on responsiveness. This is due to the estimated effect of regressors comprising both their true impact on health system responsiveness together with their effect on reporting behaviour. For example, high income earners may have access to better services and may be treated with greater consideration but may also have higher expectations of the quality of service they receive compared to their lower income counterparts.

Heterogeneous reporting behaviour can be accounted for using the hierarchical ordered probit model (HOPIT) developed by Tandon et al. (2003) (also see Terza, 1985). The method draws on the use of the anchoring vignettes to provide a source of external information that enables the identification of the cut-points as functions of covariates. Under the assumption that variation across respondents in vignette rating is fully attributable to reporting bias, the impact of covariates on the cut-points can be identified.

The model is specified in two parts: the first to identify the cut-points as a function of relevant covariates (*reporting behaviour equation*), and the second to map individual socio-economic and other characteristics to underlying health system responsiveness while controlling for reporting heterogeneity (*responsiveness equation*). Details of the method are provided below.

1. Reporting behaviour equation

To identify the cut-points as a function of respondent covariates, let R_{ik}^{v*} represent the underlying health system responsiveness for vignette, k , rated by individual i . Given that each vignette is fixed and unrelated to a respondent's characteristics, it is assumed that the expected value of the underlying latent scale depends solely on the corresponding vignette, such that:

$$R_{ik}^{v*} = K_i \eta + \varepsilon_{ik}^v, \quad \varepsilon_{ik}^* | K_i \sim N(0,1) \quad (1)$$

where K_i is the vector of vignettes, η is a conformably dimensioned vector of parameters and ε_{ik}^* is an idiosyncratic error term. R_{ik}^{v*} is unobservable to the researcher and instead we observe the vignette rating, r_{ik}^v on a five point scale ranging from 'Very bad' to 'Very good'. We assume the observed category of r_{ik}^v is related to R_{ik}^{v*} through the following mechanism:

$$r_{ik}^v = j \quad \text{if} \quad \mu_i^{j-1} \leq R_{ik}^{v*} < \mu_i^j \quad (2)$$

for $\mu_i^0 = -\infty, \mu_i^5 = \infty, \forall i, k; \quad j = 1, \dots, 5$

If the cut-points represent fixed constants (to be estimated) then the above mapping is common to the ordered probit model. For the HOPIT model, however, we allow the cut-points to be functions of covariates, X such that:

$$\mu_i^j = X_i \gamma^j \quad (3)$$

where $\mu_i^j, j = 1, \dots, 5$ are parameters to be estimated along with η and $\mu_i^1 < \mu_i^2 < \dots < \mu_i^5$. Inference on the contribution of individual characteristics to explaining reporting behaviour can be made with reference to likelihood ratio tests.

2. Responsiveness equation

Underlying health system responsiveness faced by individual i can be expressed as:

$$R_i^{s*} = Z_i \beta + \varepsilon_i^s, \quad \varepsilon_i^s | Z_i \sim N(0, \sigma^2) \quad (4)$$

where Z_i represents a set of regressors correlated with responsiveness (X_i might be a sub-set of Z_i). As with the vignettes R_i^{s*} represents an unobserved latent variable and we assume that the observed categorical response, r_i^s , relates to R_i^{s*} in the following way:

$$r_i^s = j \quad \text{if} \quad \mu_i^{j-1} \leq R_i^{s*} < \mu_i^j \quad (5)$$

for $\mu_i^0 = -\infty, \mu_i^5 = \infty, \forall i; \quad j = 1, \dots, 5$

where μ_i^j are defined by (3) with γ^j fixed and it is assumed that R_{ik}^{v*} and R_i^{s*} are independent for all $i = 1, \dots, N$ and $k = 1, \dots, V$.

It follows that the probabilities associated with each of the 5 categories are given by:

$$\Pr(r_i = j) = \Phi(\mu_i^j - X_i \beta) - \Phi(\mu_i^{j-1} - X_i \beta), \quad j = 1, \dots, 5 \quad (6)$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution.

The use of vignettes to identify reporting heterogeneity relies on two assumptions. First, it is assumed that individuals classify the vignettes in a way that is consistent with the rating of their own experiences of the service provided. This assumption is termed response consistency. Secondly, conditional on the socio-economic characteristics that determine reporting behaviour, the level of responsiveness faced by an individual does not influence the way s/he reports the responsiveness of the hypothetical scenarios. This assumption is termed the irrelevance of own provider responsiveness or vignette equivalence.

Results

Descriptive analysis of vignettes

To conserve space we present descriptive results for Mexico only.² Corresponding results for India and the Philippines are available upon request. Table 2 presents the proportion of respondents reporting each of the five categories of responsiveness. This is presented for each of the eight domains, the two items per domain and the five vignettes for each item. We also present the frequencies of respondents' valuations of their own experiences in the first column for each item.

A comparison of own versus vignette ratings clearly indicates that individuals are much more polarised in the reporting of own experiences compared to the hypothetical cases provided through the vignettes. While, approximately 70% of respondents rate their own experiences of contact with health systems as 'Good', the vignette ratings are far more dispersed across the available response categories and are infrequently observed to be above 50%. The observed differences in own ratings will consist of a combination of variation in actual experiences of contact with health systems leading to exposures to different levels of responsiveness, and reporting heterogeneity, such that individuals faced with the same level and experience of health services will use different response categories to report the experience. For India and the Philippines, in general, the distribution of scores for ratings of own experiences are in line with the distributions observed for the various vignettes.³

In contrast the difference in ratings of the vignettes is assumed to be due only to reporting heterogeneity. It can be seen from Table 2 that vignette ratings clearly exhibit heterogeneity across the response categories. This holds for all three countries and is more pronounced for India and the Philippines. For example, the first vignette for the domain of travelling time for India attracts near equal (approximately 30%) ratings across three of the five response categories. Given the fixed and exogenous nature of vignettes, such variation in respondents' ratings provides a clear indication of reporting heterogeneity within the three countries.

Table 3 to Table 6 investigate response heterogeneity by socio-economic position, age and gender. Again, results are presented for Mexico only, due both to a desire to keep the number of tables to a minimum and because of the comparative richness of data available for Mexico. Each table presents vignette ratings for the three domains of dignity, communication and confidentiality. These represent domains rated (by respondents) as most relevant across the countries of the WHS. For each domain results for the first two vignettes for the first item are provided stratified by one of educational status, income quintile, gender or age group. Accordingly, since not all domains, items or vignettes are represented, the results are illustrative only and do not present a full analysis of how vignette ratings varying with respect to respondent characteristics.

For each of the available response categories (for example, 'Very good') a gradient across the variable of interest provides evidence of reporting varying by that variable. For example, the second vignette for the domain of dignity in Table 3 shows a clear gradient across education. In

² The larger sample size afforded to Mexico ensures that the frequencies are estimated with greater precision than corresponding frequencies for India or the Philippines.

³ While the location of the distributions might differ when comparing own ratings to vignette ratings, the distribution of scores across the response categories across own and vignette scores are more similarly spread compared to those observed for Mexico.

general, more educated respondents are more likely to rate this particular vignette as ‘very good’ compared to less well educated respondents.⁴ Similar relationships hold across the other domains and vignettes.

Gradients in responses are also apparent across the income quintiles in Table 4. For example, the responses to vignette 2 for the domain of communication show that individuals further along the income distribution are more likely to report ‘very good’ and less likely to report ‘moderate’ or ‘bad’ levels of responsiveness compared to individuals at the lower end of the distribution. Similar gradients can be observed for other vignettes.

Table 5 presents responses across the vignettes by gender. In general, reporting behaviour does not appear to be influenced by the gender of the respondent with a maximum difference between women and men of four percentage points observed for the first vignette for the domain of dignity. Similarly, while some indications of a gradient across age groups in reporting behaviour can be observed for particular vignettes, on the whole, the differences across age are small (see Table 6).

Homogeneity in reporting behaviour

Table 7 presents results of formal tests for homogeneous reporting behaviour and parallel cut-point shift⁵. For each domain and item the first five columns report p-values for a Wald test under the null hypothesis of homogenous reporting. For the individual socio-demographic characteristics, this is simply a test of the joint significance of the respective estimated coefficients in the four cut-points. The first column reports the corresponding test across all four socio-demographic characteristics.

For each country there is evidence of cut point heterogeneity according to at least one of the variables for all response items. For individual characteristics, the results show variation across both country and domain items. Homogeneity in reporting is clearly rejected for income in Mexico and India, whilst for the Philippines the test fails to reject the null for the domain of Communication and only just rejects homogeneity for the domain of Dignity. Similar results are found for education status. Homogeneity is rejected across all domains and items for Mexico, for the domain of Dignity and the first item of Communication for India and for the domain of Communication only for the Philippines. For only one item is homogeneity in reporting rejected for age for Mexico, compared to three of the four items for India and all domains and items for the Philippines. Reporting behaviour by gender is more variable across the three countries. Whilst homogeneity can be rejected in three of the four domain and item combinations for Mexico, it is only rejected in one item for India and is not rejected at all for the Philippines. Overall, and consistent with the descriptive analysis, the models indicates greater reporting heterogeneity by income, followed by education, age and finally gender.

The final five columns of Table 7 report results investigating the existence of parallel cut-point shift. Again, tests are presented for all variables and by income, education, age and gender alone. In general, a similar pattern emerges for the tests for parallel cut-point shift as for the tests for

⁴ Some caution is required when interpreting these results as the education category ‘Post Graduate’ is sparsely populated and hence the response ratings are not estimated with same level of precision as for other education groups.

⁵ Parallel cut-point shift relates to a shift that is a function of covariates but, importantly, is an equal shift across all cut-points.

homogeneity. The only notable exception is for India where parallel cut-point shift is not rejected for three of the four domain items for education.

Due to reporting homogeneity being rejected most emphatically for income and education, we investigate reporting behaviour in more detail for these variables only.

Determinants of reporting behaviour

The effect of income and education on reporting behaviour is illustrated in Table 9. The table presents the coefficients of income and education on the four cut-points together with their associated standard errors. As the measures used are increasing in income and education positive coefficients across all the cut-points indicate higher responsiveness expectations and a lower probability of reporting better levels of responsiveness. This scenario can be observed for education in the domain and items of Dignity for India. Better educated individuals appear to have higher expectations of how they ought to be treated compared to lower educated individuals and accordingly are less likely to report high levels of responsiveness.

For income, there is a significant effect on at least one of the cut-points for each of the domains and items and countries except for the item of 'privacy respected during examination' for the Philippines. More effects are significant for Mexico, followed by India and finally the Philippines. This is consistent with the tests of homogeneity reported in Table 7. Ten of the thirteen significant coefficients for Mexico are positive whilst for India all significant coefficients are negative. Two of the three significant coefficients for the Philippines are positive. For India, the negative and significant coefficients imply that the better-off are more likely to report a vignette in the domains of Dignity and Communication more favourably than their lower income counterparts. For Mexico we also observe significant negative coefficients on income for the fourth cut-point, implying a greater probability of reporting excellent responsiveness among the better-off compared to the less wealthy. However, the better-off also appear more likely to report lower levels of responsiveness. This can be observed by the positive and significant coefficients for income across the first two cut-points. These results taken together appear to imply that in Mexico, the wealthy are more likely to rate vignettes more extremely than their less wealthy counterparts.

The estimated coefficients on education are presented in the lower half of Table 8. For India, all significant effects are positive, implying that the better educated have higher expectations and are, accordingly, less likely to rate the vignettes favourably compared to their less wealthy counterparts. This is particularly the case for the domain of Dignity where all effects are significant. For Mexico, the results again are inconsistent across the cut-points. Positive and significant coefficients are observed for the first cut-point, whilst negative and significant coefficients are observed for the fourth cut-point. As with income, this appears to imply that the better educated are more discerning in their ratings of the vignettes and are willing to use the extremes of the available response categories more often than their less wealthy counterparts. In keeping with the results for income, few coefficients of education are significant for the Philippines.

Adjusting for reporting heterogeneity

To assess further the impact of socio-economic characteristics on reporting behaviour we investigate the effect that adjusting for income and education has on the reporting of own experiences of health service contact. In the presence of reporting bias, the coefficients of the

index function obtained using an ordered probit model will reflect both the effect of the regressors on responsiveness and their impact on reporting behaviour. The HOPIT model attempts to separate reporting heterogeneity from impacts on underlying responsiveness by adjusting the cut-points using relevant information from the responses to the vignettes. Since the scale of the latent variable is not identified in the ordered probit model it is customary to fix the constant term and the variance to 0 and 1 respectively. In order to obtain comparability in the coefficients of the ordered probit model and the HOPIT model, we instead fixed the constant and variance parameters of the ordered probit model to those of the HOPIT model. Results are reported in Table 9 where coefficients and standard errors using an ordered probit model are compared to the corresponding coefficients and standard errors for the HOPIT model where the cut-points have been modelled as a function of income, educational status, age and gender.

The coefficients on income are notably changed when using the HOPIT approach. For Mexico and the Philippines, the coefficients from the HOPIT model are up to a third of the corresponding values from an ordered probit model, while for India the coefficients are up to a half the value of the ordered probit results. For all domains and items, the ordered probit results indicate a positive and significant income effect on responsiveness implying higher responsiveness is enjoyed by wealthy individuals compared to their less wealthy counterparts. After adjusting for reporting heterogeneity, we still observe a positive and significant effect of income, albeit much reduced for the majority of domain and item combinations. The item “greeted and talked to respectfully” in the domain of dignity for Mexico and the Philippines is, however, no longer significant and, in addition, the item of “privacy respected during examination” is no longer significant for the Philippines. Overall, the results indicate that the positive relationship between income and responsiveness is over-estimated if reporting heterogeneity by income is not accounted for. An exception to this is the domain of communication for the Philippines where the magnitude of the coefficients for the two items from the HOPIT model are greater than the corresponding coefficients from the ordered probit model. Accordingly, the ordered probit model appears to underestimate the effects of income on responsiveness. This is consistent with the sign of the cut points in Table 8, which denotes higher expectations for higher incomes.

The reported coefficients for education are perhaps more striking. The ordered probit results are all positive with effects significant for Mexico and the Philippines. While all coefficients are positive for India, none are significant. Adjusting for reporting heterogeneity results in a reduction in the absolute value of the estimated effects for education for both Mexico and the Philippines. Further, the results for Mexico become non-significant. The largest effects of education on responsiveness are observed for the Philippines. When adjusting for reporting bias, these effects are approximately halved, but remain significant, for the domain of dignity but become non-significant for the domain of communication. For India, none of the estimated coefficients from the ordered probit model are significant, while the results for the HOPIT model are larger and significant for the domain of dignity. Accordingly, for this country and domain the ordered probit model appears to underestimate the effect of education on responsiveness.

As a further indication of the effects of adjusting for reporting heterogeneity Table 10 presents the ex-ante and ex-post frequencies of reporting each of the five response categories available to respondents. Ex-ante results simply report the frequencies observed in the data for each of the domains and items. The ex-post results are based on predictions from the HOPIT model after adjusting for reporting behaviour. For Mexico the ex-ante and ex-post results are very similar. The largest differences are observed for the communication domain where there is a 4% increase in the frequency of respondents reporting moderate responsiveness for the first item and a 5%

increase on the second. Other domain and item combinations differ, in general, by up to 2%. Larger differences are, however, observed for both India and the Philippines. For India, differences of up to 12% are observed. Differences can be seen across both the domains of dignity and communication. For the Philippines, small differences are observed for the domain of dignity, while large differences are apparent for the domain of communication, where, for example, a difference of 15% is observed for the response category of ‘Good’ for the first item on how well health care providers explain things.

Conclusions and discussion

The performance of health systems is becoming the subject of increasing scrutiny. Both health and non-health benefits have been promoted as appropriate outcomes on which health systems can be judged. The concept of responsiveness has been promoted by the World Health Organization as a credible method of measuring non-health benefits and how these impact directly on patient well-being. While much has been written about the concept and measurement of responsiveness, to date little use has been made of the instrument in empirical work. This is a potentially prolific area of research where there is much to be learned about the behaviour of the instrument in practice, and in particular, its use as a comparative instrument for cross-country analysis of health systems performance.

To operationalize a survey on the scale of the WHS requires a reliance on self-reported data. However, the extent to which self-reports are comparable across socio-economic, demographic and cultural groups has been debated extensively. Debate has mainly focused on measures of health status and their use as measures of health system outcomes and equality. Similar concerns are also relevant to surveys of health systems responsiveness, where again reliance is placed on self-reported measures on a categorical scale. Systematic variation in the way individuals interpret the response categories leads to reporting heterogeneity which bias estimates of underlying service responsiveness. This paper provides some preliminary results on the extent of reporting bias, using data from three countries contained within the WHS. The analysis shows clearly that heterogeneity in reporting behaviour exists, that it is a function of income and education, but does not appear to be strongly related to age and gender. Adjusting for reporting bias impacts on the estimated coefficients of the responsiveness mean function when results from the HOPIT model are compared to those from an ordered probit regression.

A comparison of the ex-post and ex-ante frequencies for the response categories shows that for Mexico little difference is observed once reporting behaviour has been taken into account. For India and the Philippines, however, the differences in ex-post and ex-ante reporting is much greater. For India, differences of up to 12% are observed, while for the Philippines a difference of 15% is observed in the domain of communication. Overall, the results indicate that reporting behaviour is more prominent in some countries compared to others, varies across domains and can be explained, to a greater or lesser extent, by socio-economic group and in particular, for income and education.

The framework used in this paper has the potential to be extended to compare performance across countries. Factors other than income and education would need to be considered, for example the organization of the health service together with wider cultural influences. The framework described above, however, offers a means to search for characteristics of individuals, the health care systems they face and country-specific influences on those systems to control for differential reporting and enable more coherent cross-country analyses. This will be the focus of future research.

Appendix 1: Domains of responsiveness

The eight domains of responsiveness defined by the WHO are as follows (see Valentine et al., 2003a for a full exposition of these domains):

- *Autonomy*: patient autonomy implies that providers of health services must respect patients' views of what is appropriate and allow the patient to make informed choices;
- *Choice*: this reflects an individual's right or opportunity to choose a health care institution and health provider. It also relates to a patient having the ability to secure a second opinion and access specialist services when required;
- *Clarity of communication*: this domain refers to providers clearly explaining to patients and family the nature of the illness, details of treatment and available options;
- *Confidentiality of Personal Information*: this relates to providers providing privacy in the environment in which consultations are conducted and the concept of privileged communication and confidentiality of medical records;
- *Dignity*: this domain refers to the ability of patients to receiving care in a respectful, caring, and non-discriminatory setting;
- *Prompt attention*: this refers to the ability of people to access care rapidly in the case of emergencies, or readily with short waiting times for non-emergencies. It further applies to modes of access to curative and public health interventions;
- *Quality of basic amenities*: this domain refers to the physical environment and services often referred to as "hotel facilities". It incorporates the extent to which a health facility offers a welcoming and pleasant environment, including clean surroundings, regular maintenance, adequate furniture, sufficient ventilation, enough space in waiting rooms etc;
- *Access to family and community support*: this domain reflects the extent to which patients have access to their family and friends when receiving care. It also includes the right to receive food and other consumables from family and where deemed appropriate, the opportunity to carry out religious and cultural practices including practicing alternative therapies.

Example questions used in the WHS to measure responsiveness include:

- *Autonomy*: How would you rate your experience of being involved in making decisions about your health care or treatment?
- *Choice*: How would you rate the freedom you had to choose the health care providers that attended to you?
- *Communication*: How would you rate your experience of how clearly health care providers explained things to you?
- *Confidentiality*: How would you rate the way your personal information was kept confidential?
- *Dignity*: How would you rate the way your privacy was respected during physical examinations and treatments?
- *Quality of basic amenities*: How would you rate the cleanliness of the rooms inside the facility, including toilets?
- *Prompt attention*: How would you rate the amount of time you waited before being attended to?
- *Access to family and friends*: How would you rate the ease of having family and friends visit you?

The above provide examples only and not an exhaustive list of questions for each domain. The response categories available to respondents were "very good", "good", "moderate", "bad" and "very bad".

Appendix 2: Examples of vignette questions used in the WHS

Prompt attention

Vignette 1:

[Niels] woke up with a sore back so he decided to go to the clinic. It took 30 minutes to travel to the clinic and he was seen within 5 minutes.

Q1: How would you rate his travelling time?

Q2: How would you rate the amount of time he waited before being attended to?

Communication and Quality of Basic Amenities

[Wing] had his own room in the hospital and shared a bathroom with two others. The room and bathroom were cleaned frequently and had fresh air.

Q1: How would you rate the cleanliness of the rooms inside the facility, including toilets?

Q2: How would you rate the amount of space [Wing] had?

[Rose] cannot write or read. She went to the doctor because she was feeling dizzy. The doctor didn't have time to answer her questions or to explain anything. He sent her away with a piece of paper without telling her what it said.

Q1: How would you rate her experience of how clearly health care providers explained things to her?

Q2: How would you rate her experience of getting enough time to ask questions about her health problem of treatment?

Confidentiality, Choice and Involvement

[Simon] was speaking to his doctor about an embarrassing problem. There was a friend and a neighbour of his in the crowded waiting room and because of the noise the doctor had to shout when telling [Simon] the treatment he needed.

Q1: How would you rate the way the health services ensured [Simon] could talk privately to health care providers?

Q2: How would you rate the way [Simon's] personal information was kept confidential?

In [William's] town there is a large day clinic where there are several doctors and nurses. When [William] has a sensitive health problem he can see a male rather than a female doctor or nurse.

Q1: How would you rate [William's] freedom to choose his health care provider?

Social Support to Patient and Autonomy

When [Joseph] was in hospital he could have no visitors nor could he receive any presents from friends or relatives. The hospital had no telephones and he could not get any news from outside.

Q1: For [Joseph's] last hospital stay, how would you rate the ease of having family and friends visit him?

Q2: For [Joseph's] last hospital stay, how would you rate his experience of staying in contact with the outside world when he was in hospital?

[Mark] had a serious health problem. The doctor prescribed the best treatment for Mark but without telling him the implications on his quality of life or the cost. [Mark] felt powerless and was not given any information to help him to feel more in control.

Q1: How would you rate [Mark's] experience of getting information about other types of treatments or tests?

Q2: How would you rate [Mark's] experience of being involved in making decisions about his health care or treatment?

Note that the above provide examples only and not an exhaustive list of possible vignettes for each domain. The response categories available to respondents were "very good", "good", "moderate", "bad" and "very bad".

References

- Bago d'Uva, T., E. van Doorslaer, M. Lindeboom, O. O'Donnell. Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 2007; In Press.
- Blendon R. J., C. Schoen, C. DesRoches, R. Osborn, K. Zapert, Common Concerns Amid Diverse Systems: Health Care Experiences In Five Countries. *Health Affairs*, 2003: 106-121.
- Ferguson B.D., Tandon A., Gakidou E., Murray C.J.L., (2003), "Estimating Permanent Income using Indicator Variables", In: Murray CJL., Evans, DB. Eds. *Health systems performance assessment: debates, methods and empiricism*. Geneva, World Health Organisation, 2003: 748-760, chp 56
- Iburg, K. M., J. Salomon, A. Tandon and C. J. L. Murray. (2002) Cross-country comparability of physician-assessed and self-reported measures of health. In: *Summary measures of population health: concepts, ethics, measurement and applications*. C. J. Murray, J.A., Salomon, C.D. Mathers and Lopez, A.D. The World Health Organization, Geneva. 2002: 433-448.
- Idler, E. L. and S. V. Kasl. Self-ratings of health: do they also predict change in functional ability? *Journal of Gerontology*, 1995; 50B: 344-353
- Jürges, H. True health versus response styles: exploring cross-country differences in self-reported health. *Health Economics*, 2007; 16(2): 163-178.
- Kapteyn, A., J. Smith and A. van Soest. Vignettes and self-reports of work disability in the US and the Netherlands. *American Economic Review*, 2007; 97(1): 461-473
- Kempen, G.I. Steverink N., Ormel J., Deeg D.J. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. *Journal of Gerontology*, 1996; 51B(5): 254-260
- Kerkhofs, M. J. M. and M. Lindeboom Subjective health measures and state dependent reporting errors. *Health Economics*, 1995; 4: 221-235
- King, G., C. J. L. Murray, J. Salomon and A. Tandon. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 2004; 98(1): 184-91
- Lindeboom, M. and E. van Doorslaer. Cut-point shift and index shift in self-reported health. *Journal of Health Economics*, 2004; 23(6): 1083-1099
- Manderbacka, K. Examining what self-rated health question is understood to mean by respondents. *Scandinavian Journal of Social Medicine*, 1998; 26(2), 145-153
- Murray, CJL., Frenk, J., A framework for assessing the performance of health systems. *Bulletin of the World Health Organization*, 2000; 78: 717-731.
- Murray, C. J. L., Ozaltin, E., Tandon, A.J., Salomon, et al. Empirical evaluation of the anchoring vignettes approach in health surveys. In: *Health systems performance assessment: debates, methods and empiricism*. C. J. L. Murray and D. B. Evans (eds). Geneva, World Health Organization, 2003.
- Puentes Rosas E., Gómez Dantés O, Garrido Latorre F. Trato a los usuarios en los servicios públicos de salud en México. *Rev Panam Salud Publica*, 2006;19(6):394-402.
- Sadana R, Mathers C.D., Lopez A.D., Murray C.J.L., Iburg K.M. (2002). Comparative analyses of more than 50 household surveys on health status. In: *Summary measures of population health: concepts, ethics, measurement and applications*. Murray C.J.L., Salomon J.A., Mathers C.D. and Lopez A.D (eds). Geneva, World Health Organization, 2002: 369-386

- Salomon, J., A. Tandon, C. J. L. Murray and World Health Survey Pilot Study Collaborating Group. Comparability of self-rated health: Cross sectional multi-country survey using anchoring vignettes. *British Medical Journal*, 2004; 328: 258
- Tandon, A., C. J. L. Murray, J. A. Salomon and G. King. Statistical models for enhancing cross-population comparability. *In: Health systems performance assessment: debates, methods and empiricisms*. C. J. L. Murray and D. B. Evans (eds). Geneva, World Health Organization, 2003: 727-746.
- Terza, J. V. Ordinal probit: a generalization. *Communications in Statistics*, 1985: 14(1): 1-11.
- Üstün, T. B., S. Chatterji, A. Mechbal, C. Murray, et al. The World Health Surveys. *In: Health systems performance assessment: debates, methods and empiricisms*. C. J. L. Murray and D. B. Evans (eds). Geneva, World Health Organization, 2003b: 762-796.
- Valentine, N.B., De Silva, A., Kawabata, K., Darby, C., Murray, C.J.L., Evans, D. Health system responsiveness: concepts, domains and operationalization. *In: Health systems performance assessment: debates, methods and empiricisms*. C. J. L. Murray and D. B. Evans (eds). Geneva, World Health Organization, 2003a: 573-596.
- Valentine, N.B., Salomon, J.A. Weights for responsiveness domains: analysis of country variation in 65 national sample surveys. *In: Health system performance assessment: debates, methods and empiricism*. Murray, C.L.J., Evans, D.B. (eds). Geneva, World Health Organization, 2003.
- Valentine N. B., Ortiz J.P., Tandon A., Kawabata K., Evans DB., Murray C.J.L. Patient Experiences with Health Services: Population Surveys from 16 OECD Countries. *In: Health systems performance assessment: debates, methods and empiricisms*. C. J. L. Murray and D. B. Evans (eds). Geneva, World Health Organization, 2003b: 643 – 652

Mexico					
	n	mean	s.d.	min	Max
Women	38745	0.577	0.494	0	1
Age	38745	41.0	16.74	18	106
Education categories	38745	4.09	0.628	3	7
Education in years	38745	7.20	4.97	0	27
Income	38745	3.00	1.41	1	5
India					
	n	mean	s.d.	min	Max
Women	8356	0.504	0.500	0	1
Age	8356	38.59	15.26	15	101
Education categories	8356	2.91	1.80	1	7
Education in years	8356	5.17	5.09	0	30
Income	8356	3.05	1.42	1	5
The Philippines					
	n	mean	s.d.	min	Max
Women	10053	0.538	0.499	0	1
Age	10053	38.92	14.57	18	99
Education categories	10053	3.57	1.23	3	7
Education in years	10053	8.53	3.70	0	24
Income	10053	3.00	1.41	1	5

Table 1: Descriptive statistics

Prompt Attention												
	Travelling time						Waiting times					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	11.8	11.8	2.6	5.4	35.6	3.2	14.4	25.0	2.7	5.9	35.8	2.7
Good	68.1	43.4	8.0	8.7	46.8	16.2	68.0	53.2	6.3	8.5	46.1	16.5
Moderate	13.1	34.1	13.0	9.0	9.9	43.2	10.8	15.4	10.5	8.4	10.2	40.9
Bad	6.0	9.1	48.8	37.5	5.8	30.5	5.4	5.0	40.2	31.7	5.7	30.7
Very bad	1.1	1.7	27.7	39.1	1.9	6.9	1.5	1.5	40.3	45.5	2.2	9.2
N	8442	9629	9620	9609	9624	9613	8443	9628	9622	9624	9623	9612
Dignity												
	Greeted and talked to respectfully						Privacy respected during examination and treatment					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	16.3	3.4	24.4	33.6	15.3	2.3	16.5	1.9	29.8	36.5	18.2	2.2
Good	71.2	13.2	50.6	47.6	51.8	6.1	73.5	8.4	46.7	46.3	47.7	5.9
Moderate	8.7	23.9	12.8	9.2	23.0	9.7	7.4	16.2	11.3	8.9	23.4	9.1
Bad	3.1	47.6	8.4	6.5	7.6	30.9	2.3	50.2	8.0	5.7	8.5	27.2
Very bad	0.7	11.9	3.8	3.2	2.3	51.0	0.4	23.3	4.2	2.6	2.2	55.6
N	8443	9618	9612	9618	9618	9620	8443	9627	9623	9627	9613	9583
Communication												
	How clear health care providers explain things						Getting enough time to ask questions					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	16.3	21.8	31.9	5.4	2.2	2.2	16.1	21.9	31.9	4.7	2.3	2.4
Good	71.8	62.0	50.2	23.4	6.4	11.5	71.0	60.1	49.2	19.9	6.3	11.7
Moderate	8.0	10.5	10.2	36.9	9.7	28.9	8.5	12.5	11.1	35.0	10.9	25.9
Bad	3.2	4.7	5.9	26.9	45.2	42.8	3.7	4.5	5.7	29.2	44.4	44.2
Very bad	0.7	1.0	1.8	7.5	36.6	14.7	0.7	1.1	2.1	11.2	36.1	15.9
N	8443	9655	9642	9644	9648	9643	8443	9656	9655	9651	9651	9644
Quality of Basic Amenities												
	Cleanliness of room						Enough Space provided					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	15.9	20.0	48.8	2.0	1.8	1.7	14.6	17.3	44.2	1.9	1.9	1.6
Good	70.1	51.9	37.1	6.3	6.6	8.6	68.1	48.9	40.3	5.9	7.0	7.0
Moderate	10.0	20.6	8.1	9.3	13.7	16.9	11.8	24.5	9.1	10.3	13.2	17.9
Bad	3.3	6.4	5.0	35.0	45.9	48.0	4.7	7.9	5.1	35.5	47.2	47.7
Very bad	0.7	1.1	1.1	47.5	32.0	24.9	0.8	1.4	1.4	46.5	30.7	25.8
N	8443	9648	9644	9638	9646	9646	8443	9653	9648	9649	9643	9562
Confidentiality												
	Talk privately to health care providers						Information kept confidential					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	14.0	2.0	2.9	4.0	2.9	24.6	14.3	2.0	2.8	3.5	3.1	27.2
Good	70.1	11.2	19.3	12.8	13.4	44.1	71.9	10.5	16.1	12.1	12.0	46.3
Moderate	10.6	15.6	23.8	15.1	18.7	18.4	10.0	15.2	20.6	14.7	18.7	16.3
Bad	4.2	54.1	39.9	52.1	52.2	10.6	2.7	52.9	42.2	48.4	51.6	7.9
Very bad	0.6	17.1	14.0	16.1	12.8	2.3	1.2	19.5	18.2	21.3	14.6	2.3
N	8443	9680	9666	9669	9665	9662	8416	9681	9666	9663	9663	9657
Access to family and community support												
	Family and friends able to visit						Contact with the outside world					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	15.3	2.8	1.5	4.6	35.0	25.2	15.0	2.8	1.9	5.1	34.5	23.8
Good	68.2	14.1	7.2	29.8	41.9	47.0	67.4	13.8	6.8	30.1	42.4	47.6
Moderate	10.6	29.3	10.9	42.1	14.3	20.0	11.1	30.0	11.3	41.7	14.5	20.0
Bad	4.8	46.2	53.3	19.6	6.8	6.3	5.4	44.1	50.5	19.5	6.7	6.8
Very bad	1.1	7.6	27.1	4.0	2.0	1.6	1.2	9.3	29.5	3.6	1.9	1.8
N	8442	9675	9659	9669	9670	9660	8443	9683	9677	9671	9670	9661
Autonomy												
	Information about other services/ treatment						Involved decision making					
	Own	Vig1	Vig2	Vig3	Vig4	Vig5	Own	Vig1	Vig2	Vig3	Vig4	Vig5
Very Good	13.6	2.8	2.6	11.3	29.1	22.0	13.6	1.6	2.9	11.2	29.4	20.9
Good	68.3	10.4	10.9	37.7	44.4	48.5	69.2	8.4	9.8	36.7	42.6	45.2
Moderate	11.5	21.2	24.1	28.5	17.1	17.5	11.5	18.0	24.8	29.0	18.0	21.2
Bad	5.9	52.8	47.9	18.4	7.5	8.8	5.0	56.7	47.2	18.6	8.2	9.3
Very bad	0.8	12.8	14.6	4.1	1.9	3.2	0.6	15.2	15.2	4.4	1.9	3.4
N	8441	9678	9669	9665	9662	9662	8443	9680	9666	9671	9668	9606

Table 2: Summary frequencies for the reporting of vignettes: Mexico. Domain of Choice omitted.

Dignity: Greeted and talked to respectfully										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Primary school	3.1	16.7	28.2	41.9	10.0	22.2	50.5	14.4	9.4	3.5
Edu secondary	3.3	12.9	23.7	48.2	11.9	23.4	51.2	12.9	8.6	3.9
High school	3.8	11.9	22.0	49.4	12.9	28.3	48.9	11.6	7.4	3.7
Post grad	2.6	13.2	21.0	42.1	21.1	36.8	42.1	5.3	7.9	7.8
Communication: How clear health care providers explained things										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Primary school	20.5	61.6	11.0	5.7	1.2	27.2	51.6	11.5	7.5	2.2
Edu secondary	21.3	62.5	10.5	4.8	1.0	30.7	51.3	10.5	5.8	1.6
High school	23.6	61.0	10.6	4.0	0.8	37.9	46.4	8.8	5.2	1.9
Post grad	34.0	54.0	2.0	6.0	4.0	48.0	40.0	6.0	4.0	2.0
Confidentiality: Talk privately to health care providers										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Primary school	2.1	13.1	15.8	51.5	17.5	3.3	22.1	25.6	38.2	10.8
Edu secondary	1.9	11.5	15.7	54.6	16.4	2.6	19.4	23.4	40.5	14.1
High school	2.1	9.1	15.5	54.6	18.8	3.4	17.4	23.7	39.5	16.0
Post grad	2.6	7.7	7.7	46.2	35.9	5.1	23.1	15.4	46.2	10.3

Table 3: Vignette ratings by Education: Mexico

Dignity: Greeted and talked to respectfully										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
1	3.4	15.4	27.3	42.7	11.1	22.4	50.1	14.8	8.7	4.1
2	3.8	13.7	23.5	48.9	10.1	22.7	52.5	12.6	8.9	3.3
3	3.0	12.9	23.7	47.5	13.0	23.5	51.0	12.2	9.2	4.2
4	3.3	12.3	22.4	50.1	11.9	24.5	51.6	12.4	7.7	3.9
5	3.5	11.6	22.4	49.2	13.4	29.0	47.6	11.7	7.8	3.9
Communication: How clear health care providers explained things										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
1	19.1	62.5	10.8	6.4	1.2	26.5	52.8	12.1	6.6	2.0
2	19.4	64.4	10.6	4.7	0.9	28.2	53.1	10.6	6.4	1.7
3	22.5	62.0	10.3	4.4	0.9	32.8	50.1	9.7	5.7	1.7
4	22.7	61.4	10.6	4.4	0.9	33.3	49.9	10.2	5.2	1.4
5	25.7	59.2	10.5	3.8	0.9	39.3	45.0	8.4	5.2	2.1
Confidentiality: Talk privately to health care providers										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
1	2.4	13.0	15.9	53.2	15.6	3.1	23.8	25.6	36.6	11.0
2	2.1	12.0	15.9	52.7	17.3	2.9	20.5	23.3	40.3	13.0
3	1.5	11.7	15.9	53.9	16.9	2.9	18.4	24.6	39.8	14.3
4	2.2	10.1	14.9	55.6	17.2	3.3	17.1	23.1	41.1	15.4
5	1.8	8.8	15.7	55.2	18.5	2.6	17.2	22.0	41.6	16.6

Table 4: Vignette ratings by Income quintiles: Mexico

Dignity: Greeted and talked to respectfully										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Women	3.2	12.2	23.5	49.2	11.9	24.2	51.9	12.3	7.9	3.8
Men	3.7	14.6	24.6	45.2	11.9	24.7	48.6	13.5	9.3	3.9
Communication: How clear health care providers explained things										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Women	22.2	61.7	10.4	4.7	1.1	32.2	50.0	10.1	5.8	1.9
Men	21.2	62.4	10.7	4.8	0.9	31.5	50.4	10.5	6.0	1.6
Confidentiality: Talk privately to health care providers										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
Women	1.9	10.7	15.2	55.2	17.0	2.9	19.3	23.0	40.5	14.3
Men	2.1	11.8	16.1	52.7	17.3	2.9	19.4	24.8	39.2	13.8

Table 5: Vignette ratings by Gender: Mexico

Dignity: Greeted and talked to respectfully										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
10 – 35	3.4	12.7	23.9	47.1	13.0	25.6	49.8	12.5	8.5	3.6
36 - 50	3.3	12.7	23.4	49.7	10.9	23.9	50.7	13.0	8.7	3.7
51 - 65	3.6	14.2	23.8	47.5	10.9	23.2	51.2	12.6	8.8	4.3
66 +	3.2	15.0	26.0	44.3	11.5	22.2	52.8	13.4	7.1	4.5
Communication: How clear health care providers explained things										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
10 – 35	21.3	62.8	10.2	4.6	1.1	32.7	50.0	9.7	5.9	1.7
36 - 50	23.9	60.9	10.2	4.1	0.8	32.9	49.2	11.0	5.2	1.6
51 - 65	20.5	60.1	12.4	6.0	1.0	30.5	50.4	10.6	6.8	1.7
66 +	19.9	63.9	10.0	5.2	1.0	27.8	53.4	9.8	6.5	2.5
Confidentiality: talk privately to health care providers										
	Vignette 1					Vignette 2				
	Very Good	Good	Moderate	Bad	Very Bad	Very Good	Good	Moderate	Bad	Very Bad
10 – 35	2.2	11.7	16.0	53.2	16.9	3.0	19.1	24.0	39.5	14.3
36 - 50	2.1	10.6	15.5	54.5	17.3	3.0	19.4	23.5	40.0	14.1
51 - 65	1.6	9.5	15.1	56.1	17.8	2.9	19.5	23.4	40.8	13.4
66 +	1.5	12.8	15.1	53.9	16.7	2.5	19.6	24.0	40.1	13.8

Table 6: Vignette ratings by Age: Mexico

		Mexico									
		Heterogeneity					Parallel cut-point shift				
		All	Inc.	Women	Age	Educ.	All	Inc.	Women	Age	Educ.
<i>Dignity</i>											
Greeted and talked to respectfully		.000	.000	.000	.988	.000	.000	.000	.000	.978	.000
Privacy respected		.000	.000	.000	.851	.000	.000	.000	.000	.738	.000
<i>Communication</i>											
Provides explained things		.000	.000	.210	.010	.000	.000	.000	.119	.005	.000
Enough time for questions		.000	.000	.012	.083	.000	.000	.000	.005	.053	.000
		India									
		Heterogeneity					Parallel cut-point shift				
		All	Inc.	Women	Age	Educ.	All	Inc.	Women	Age	Educ.
<i>Dignity</i>											
Greeted and talked to respectfully		.002	.000	.209	.024	.030	.044	.007	.362	.389	.862
Privacy respected		.000	.005	.007	.001	.003	.000	.012	.044	.094	.710
<i>Communication</i>											
Provides explained things		.000	.000	.314	.009	.002	.000	.002	.199	.036	.003
Enough time for questions		.010	.005	.557	.160	.183	.005	.003	.392	.189	.101
		Philippines									
		Heterogeneity					Parallel cut-point shift				
		All	Inc.	Women	Age	Educ.	All	Inc.	Women	Age	Educ.
<i>Dignity</i>											
Greeted and talked to respectfully		.003	.092	.716	.011	.250	.004	.205	.560	.017	.159
Privacy respected		.005	.619	.505	.016	.138	.009	.714	.362	.069	.073
<i>Communication</i>											
Provides explained things		.000	.044	.521	.027	.005	.000	.042	.565	.017	.002
Enough time for questions		.000	.050	.232	.008	.000	.000	.038	.140	.003	.000

Table 7: Tests of homogenous reporting and parallel cut-point shift

		Income											
		Mexico				India				Philippines			
		Mu1	Mu2	Mu3	Mu4	Mu1	Mu2	Mu3	Mu4	Mu1	Mu2	Mu3	Mu4
<i>Dignity</i>													
Talked Respectfully		.020 (.007)	.014 (.006)	.000 (.005)	-.030 (.006)	-.006 (.018)	.007 (.013)	-.028 (.012)	-.062 (.014)	-.036 (.015)	-.009 (.011)	-.016 (.011)	.001 (.014)
Privacy respected		.020 (.007)	.014 (.006)	.000 (.005)	-.017 (.006)	.021 (.017)	-.009 (.014)	-.031 (.013)	-.042 (.014)	-.011 (.014)	-.001 (.011)	-.011 (.011)	-.018 (.014)
<i>Communication</i>													
Pro explained things		.039 (.007)	.016 (.005)	.011 (.005)	-.024 (.007)	-.056 (.021)	-.013 (.014)	-.018 (.013)	-.065 (.014)	-.021 (.017)	.000 (.012)	.015 (.011)	.038 (.014)
Time for questions		.036 (.007)	.020 (.005)	.011 (.005)	-.008 (.007)	-.004 (.019)	.011 (.014)	.001 (.013)	-.048 (.015)	-.010 (.017)	-.009 (.012)	.028 (.011)	.019 (.014)
		Education											
		Mexico				India				Philippines			
		Mu1	Mu2	Mu3	Mu4	Mu1	Mu2	Mu3	Mu4	Mu1	Mu1	Mu1	Mu1
<i>Dignity</i>													
Talked Respectfully		.006 (.002)	.000 (.002)	-.002 (.002)	-.008 (.002)	.010 (.005)	.008 (.004)	.008 (.004)	.011 (.004)	.001 (.006)	.004 (.004)	-.002 (.004)	-.007 (.006)
Privacy respected		.006 (.002)	.001 (.002)	.000 (.002)	-.010 (.002)	.015 (.005)	.013 (.004)	.013 (.004)	.009 (.004)	.009 (.006)	.004 (.004)	-.007 (.004)	-.006 (.006)
<i>Communication</i>													
Pro explained things		.039 (.007)	.016 (.005)	.011 (.005)	-.024 (.007)	.020 (.006)	.007 (.004)	-.002 (.004)	-.002 (.004)	.011 (.007)	.012 (.005)	-.007 (.005)	-.009 (.005)
Time for questions		.011 (.002)	.002 (.005)	-.001 (.002)	-.010 (.002)	.008 (.006)	-.003 (.004)	-.005 (.004)	-.005 (.004)	.006 (.007)	.015 (.005)	-.009 (.005)	-.005 (.005)

Table 8: Estimated coefficients and standard errors (in parentheses) of permanent income and education of the cut-points.

			Income					
			Mexico		India		Philippines	
			oprobit	hopit	oprobit	hopit	oprobit	hopit
<i>Domain: Dignity</i>								
Greeted and talked to respectfully	Coef		.043	.012	.158	.075	.045	.020
	SE		.011	.009	.022	.019	.021	.018
Privacy respected during examination	Coef		.063	.025	.162	.082	.045	.017
	SE		.011	.008	.022	.019	.021	.017
<i>Domain: Communication</i>								
Health care providers explained things	Coef		.041	.019	.142	.074	.067	.071
	SE		.011	.009	.022	.020	.021	.019
Enough time for questions	Coef		.035	.023	.165	.116	.048	.058
	SE		.011	.009	.025	.021	.021	.019
			Education					
			Mexico		India		Philippines	
			oprobit	hopit	oprobit	hopit	oprobit	hopit
<i>Domain: Dignity</i>								
Greeted and talked to respectfully	Coef		.011	.002	.008	.015	.040	.025
	SE		.003	.003	.006	.006	.008	.007
Privacy respected during examination	Coef		.010	.000	.002	.012	.041	.021
	SE		.003	.003	.006	.006	.008	.007
<i>Domain: Communication</i>								
Health care providers explained things	Coef		.009	-.001	.006	.008	.025	.014
	SE		.003	.003	.006	.006	.008	.008
Enough time for questions	Coef		.009	-.001	.007	.004	.036	.002
	SE		.003	.003	.006	.006	.008	.007

Table 9: Coefficients of permanent income and education in the OPROBIT and HOPIT model

			Mexico		India		Philippines	
Response category			Ex-ante	Ex post	Ex ante	Ex post	Ex ante	Ex post
<i>Domain: Dignity</i>								
Greeted and talked to respectfully	Very Good		16.3	14.6	19.3	14.9	7.6	4.6
	Good		71.2	73.1	58.7	67.6	57.1	61.7
	Moderate		8.7	11.1	17.8	15.8	33.3	32.4
	Bad		3.1	1.6	3.8	1.7	1.7	1.3
	Very Bad		0.7	0.0	0.5	0.0	0.3	0.0
Privacy respected during examination	Very Good		16.5	16.2	20.9	16.5	7.4	4.9
	Good		73.5	72.6	55.3	65.1	56.3	60.2
	Moderate		7.4	9.7	19.4	16.1	34.9	33.5
	Bad		2.3	1.4	3.8	2.2	1.3	1.4
	Very Bad		0.4	0.0	0.5	0.0	0.1	0.0
<i>Domain: Communication</i>								
Health care providers explained things	Very Good		16.3	16.4	21.3	28.6	7.6	17.7
	Good		71.8	69.5	55.5	41.3	56.9	41.3
	Moderate		8.0	12.3	18.1	18.6	33.5	31.1
	Bad		3.2	1.8	4.6	10.1	1.8	9.2
	Very Bad		0.7	0.0	0.5	1.3	0.2	0.6
Enough time for questions	Very Good		16.1	16.6	19.9	26.2	7.0	16.8
	Good		71.0	67.7	53.2	41.7	52.7	39.2
	Moderate		8.5	13.4	21.8	21.0	37.9	32.2
	Bad		3.7	2.2	4.4	9.8	2.3	10.9
	Very Bad		0.7	0.0	0.6	1.3	0.2	1.0

Table 10: Ex ante frequencies of each reported level of responsiveness versus ex post probabilities (from HOPIT model)