Health Care System Process Utility Compared;

Using a Hierarchical Ordered Probit Regression Analysis

Renske Kok

Erasmus University
March 2008

Abstract

Health system performance is mostly measured in terms of health outcomes. While health outcomes can be seen as a measure for outcome utility, they do not measure process utility. Yet a large share of overall health care expenditures is used to respond to patients' wishes, even though this may not affect health outcomes. We argue that the WHO concept of responsiveness can be used to capture process utility. Moreover, when paired with a hierarchical ordered probit, it can be corrected for the subjectivity of responses. We improve the correction technique and use much better data to compare the process utility of 56 health care systems.

Introduction

People make use of health care to improve their health. However, next to health gain people may also feel taken care of, or listened to when receiving care. Not only health, but also the process of being treaded and cared for generates utility. This is illustrated by the large part of the health care budget that is allocated in order to improve process utility. On average 11.15% of the health care budget in OECD countries is spent on long-term inpatient nursing care alone (OECD Health data 2006). Process utility in health care is obviously important in long term care, where the aim is not mainly the improvement of health. Nevertheless, it is essential to understand that process utility is also important in cure, and large part of the cure budget is spent in order to improve the process. The total part of the health care budget spent on process utility will therefore be considerably larger then 11.15%, which indicates the importance and impact of process utility on health system performance.

Process utility in health care is an independent source of utility; it can influence the utility gained or lost by health care independent of the outcome health. The uncertainty of treatments, the information asymmetry between medical specialist and patient, and the large effect of negative process utility in health care emphasizes the importance of the process utility. Judging health care system performance purely on the ability to increase health will therefore not give a representative image of its performance. Unfortunately, measuring process utility in health care is not a straightforward task. In earlier studies patient satisfaction was used as a measure for this purpose. Unfortunately, this is a highly subjective measure and leads to incomparable outcomes. Surveys targeting at process utility or the quality aspect of health care do exist. A well-known example is CAHPS. This survey is however targeted at the quality of a specific health provider or a health plan in contrast to the process utility of an entire health system. In the Netherlands the CQ-QUOTE survey evaluates quality aspects of health care and health insurances. However, no correction for reporting behaviour and weighing of preferences is included in both surveys, which complicates comparability of the outcomes.

The World Health Organisation (WHO) made a large leap by incorporating a comparable measure for process utility into a framework for measuring health system performance in the World Health Report 2000 (WHR 2000). The World Health Organisation (WHO) is the first to make an extensive international comparison of health care systems incorporating the process aspect of health care systems in the conceptual framework of performance. They refer to this as the responsiveness of the health care system. Next to responsiveness this framework includes measures for health and fairness of financing. The WHO defines responsiveness as: *A module to measure the quality of aspects of the interaction between individuals and the health system that have the potential to improve well-being,*

focusing on those aspects which are additional to the improvement in health. Responsiveness is a composite score of eight domains (Valentine, 2003c):

- Autonomy: the level of involvement of individuals in medical decisions;
- Choice: the freedom an individual has to choose a provider;
- Communication: the clearness of the explanation by the medical supplier;
- Confidentiality: the way the health services ensure privacy;
- Dignity: the respectfulness in treatment and communication;
- Quality of basic amenities: the quality of the surroundings;
- Prompt attention: convenient travel and short waiting times;
- Support: the contact with the outside world and maintenance of regular activities.

The latter is solely applicable to inpatient care. The WHO made use of vignettes in the estimation of the separate domain-scores. These vignettes enable corrections for differences in reporting behavior, which makes sure that variation in scores are attributional to differences in attained responsiveness.

In the World Health Report 2000 (WHR 2000) the WHO first published health system performance scores, including responsiveness. This report received much attention in the media and in the scientific journals. Several authors indicated some limitations of the study that partly have been acknowledged by the WHO. First, the quality of the data on which the responsiveness measure was based was regularly questioned, since data from 1791 key informants in only 35 countries was used and extrapolated to the remaining countries. In addition, the WHO only included users of the health care systems in their calculations, and did not adjust for people who are excluded from health care and to which the system clearly is very unresponsive. Consequently, a system that excludes many potential users may appear more responsive than a system that includes all. Further, some authors questioned the determination of the weights for aggregation of the domains of responsiveness, as they do not vary across countries. It is argued that due to cultural and socio-economic factors preferences for various aspects of responsiveness differ across countries, which should be taken into account. Finally, Richardson et al. (2003) recalculated the efficiency scores of health system performance of the WHO for stratified clusters of countries. He concluded that the performance of the model using the OECD data suggests that countries should be stratified and analyzed separately.

The WHO acknowledged much of the criticisms, and formulated partly new methodology in the book: Health system Performance Assessment; Debates, Methods and Empiricism. Second, in 2002 they executed a large household survey in 67 different countries concerning health system performance, which is more representative than the data used for the WHR 2000. This paper will elaborate on the concept of responsiveness developed by the

WHO. It begins with highlighting the theoretical grounds of the concept from an economic perspective, after which the methodology used by the WHO will be applied to the new data of the household survey of 2002. The methodology is relatively new and therefore is open for improvement, which will be explored in this paper.

The focus will be on the methodology of correction for reporting behavior and weighing of domains. The WHO uses an advanced econometric model to incorporate a correction for the differences in reporting behavior across countries, which is done by including vignettes in the questionnaire and in the model. This methodology is used in health more often, where it has been proven to be a significant addition to the model, but has a small influence on the outcomes (Bago d'Uva et al. 2008). The effect and the magnitude of the response heterogeneity will be evaluated by comparing the final outcomes without vignettes with the calculations with the vignettes. In the Performance Assessments book of the WHO a new method of establishing domain-weights is discussed. Here, the weights are country-specific in contrast to the universally determined weights used in the WHR 2000. Next to these weighing methods, clustering of countries in the assessments of weights could be an appropriate method as performed earlier by Richardson on the efficiency scores. If countries are clustered homogenously to cultural and socio-economic factors, preferences will not differ dramatically and therefore the same weights can be applied. These three methods of determining weights will be performed and evaluated in this paper.

Finally, the ultimate goal of the evaluation of these methodological issues is the calculation of responsiveness-scores. These calculations will be based upon the ambitious methodology of the WHO in the WHR 2000, and the newly available data of the household survey in 2002. This data is a large improvement for the analysis, considering that most of the criticisms received upon the WHR 2000 concerned the quality of the data. The WHO-methodology will form the foundation of the calculations, attributed by new parts that are based upon often-mentioned criticisms and methods proposed in the book of the WHO in 2003.

Process utility

The concept of responsiveness developed by the WHO is an example of a measure of process utility. The theory of process utility is relatively new in economics and responsiveness has not been linked to this concept before. This theoretical grounding of the concept highlights the importance and justifies its inclusion into health system performance assessment.

Benz (2005) and Frey (2004, 2005) presented the theory of process utility, interchangeably used with procedural utility, and questioned the harsh concentration on outcomes by economists. They argue that the view that people only base their choices on the expected outcome is constrained, and not very realistic. Choices are also influenced by the

means of reaching the outcomes. In other words, utility gained (or lost) in the procedure needs to be included next to the utility gained (or lost) from the outcome. In this sense the process itself, in this case health care, gains intrinsic value instead of just being instrumental. Consequently, utility gained in the process is considered as an independent source of utility. It should be emphasized that economics is potentially open to the integration of many forms of human needs and desires and thereby does not rule out inclusion of intrinsic value of process utility. The positivistic movement of the 1930s adapted the view that utility cannot be observed directly, but is revealed by the behavior of the people. Although outcomes are observed relatively more simple, that what individuals value could be anything, which leaves the opportunity for the inclusion of process utility.

Although procedural utility may be relatively new in economics, in social sciences this concept is fully integrated. As mentioned by Benz, psychologists have developed an understanding of the psychological needs of the human being. The "self-determination theory" by Deci and Ryan (2000) states that autonomy, competence and relatedness are the three most essential needs of humans. Where, autonomy is the desire to self-organize one's own actions, competence is the human will to control the environment and experience oneself as capable and effective, and finally relatedness refers to the feeling of being connected to others in love and care, and feeling accepted in a social group. Procedures addressing these basic psychological needs generate utility independent of the outcomes of that procedure. This theoretical ground of self-determination makes hypothesis testing of procedural utility possible. More detailed description and related approaches to procedural utility in economics and empirical evidence of the relevance of procedural utility can be found in articles of Benz (2005) and Frey (2004, 2005)

The domains of responsiveness directly link to the basic psychological needs of the Self-determination theory of Ryan and Deci. If the performance of the health system is satisfied on all domains of responsiveness, the basic psychological needs of autonomy, competence and relatedness of the individual are satisfied. The presence or absence of responsiveness to these domains independently influences the utility gained (or lost). Disregarding process utility, and thereby responsiveness, will lead to sub-optimal allocation of resources, and to an image of the performance of the health care system which is not representative.

To conclude, health system performance evaluation cannot be complete without incorporating the utility gains or losses of the process. Not solely the expected outcome but also the process influence the choices of people, and should therefore not be neglected. This is even more important in health care, because of the uncertainty of treatments, the information asymmetry between medical specialist and patient, and the large effect of negative utility. Not

including these gains or losses will lead to an incomplete and disturbed image of health care performance.

Methodology

The research methodology used to calculate responsiveness scores consists of two main parts. First, scores of performance per domain will be estimated using the ordered probit and the hierarchical ordered probit (HOPIT) model. By comparing the outcomes the influence of response heterogeneity will be evaluated. Second, the scores on the domains will be aggregated into an overall responsiveness score. The domains are weighted since it is conceivable that not all domains are equally important. Three different methods of determining weights are considered, which are universal, country-level and clustered weighing.

Estimation of the domain scores

The scores of the domains of responsiveness are based upon self-reported ordinal responses. It is assumed that a certain 'latent scale' represents the true level of responsiveness performance on the domains. This 'latent scale' is measured on a continuous scale. Cut-points represent the position on the latent scale where a respondent shifts from one category to another. These cut-points do not need to be equal for all respondents. Socio-economic and demographic characteristics may influence reporting behavior, and thereby the position of the cut-points on the latent scale of responsiveness. This is referred to as reporting heterogeneity, state-dependent reporting bias, response category cut-point shift or differential item functioning (DIF) (M. Jones 2007, King et al. 2004, Tandon et al. 2002). The concept of reporting heterogeneity is visualized in figure 1 with an example of self-reported respect. For each individual, the thresholds between the categories are located differently on the latent scale of respect. If, for example, all individuals have experienced a level of respect X, all individuals will report a different level of respect (good, moderate, and very bad).

Using the ordered probit model to predict domain scores, which is generally applied to ordinal responses, would result in biased outcomes since the ordered probit model assumes fixed cut-points, and thereby does not adjust for reporting bias. The relatively new hierarchical ordered probit (HOPIT) model overcomes this shortcoming by allowing cut-points on the 'latent scale' to vary across individuals. This variation in cut-points is based on responses to vignettes. A vignette is a description of a concrete level of responsiveness on a given domain that respondents are asked to evaluate with relation to the same main question and on the same categorical response scale as the main self-report question (Salomon et al, 2004). These vignettes have a fixed level of responsiveness; thereby, the variation in ratings of vignettes cannot result from differences in performance, but only from reporting

heterogeneity. In order for this to hold two assumptions are needed. First, response consistency must be satisfied, which means that respondents must rate the vignettes in the same way as they rate the self-report questions. Van Soest et al. tested this assumption in a study comparing a subjective and an objective measure of drinking problems. In this study the assumption of response consistency was satisfied although the vignettes did not give complete information; the number of drinks per occasion was given, but not the number of events per time period. The use of vignettes improved the fit of the model and raised the correlation between the subjective and the objective measure, indicating the essence of the use of vignettes in subjective measures. Second, all respondents must understand the level of responsiveness represented in the vignettes in the same way, irrespectively of their age, sex, income, education, country of residence or other socio-demographic variables (Salomon et al., 2001). If these assumptions are satisfied, vignettes can be used to estimate the individual cut-point shifts on the latent scale of true responsiveness of a given domain.

To conclude, the HOPIT model consists of two parts. The first part estimates cut-points using responses to vignettes, which reflects reporting heterogeneity. The second part utilizes responses on the self-reported main questions, and thereby represents the relationship between the individual's own responsiveness and the observables (Tandon et al., 2002, Bago d'Uva et al., 2008). So, the observations of responsiveness will all be estimated on the same scale, therefore the estimated betas reflect true differences of responsiveness instead of both differences in responsiveness and reporting heterogeneity.

An ordered probit model is estimated to explore the magnitude of response heterogeneity on the self-reported level of responsiveness per domain by comparing the outcomes with the outcomes of the HOPIT model. As mentioned earlier, the ordered probit model assumes the thresholds between the categories to be equal for all individuals. In other words: it assumes that response heterogeneity does not exist or that it doesn't vary systematically among the population. A continuous latent variable Y_{ij}^*, with *i* indicating the individual and *j* the domain, is constructed as followed:

$$Y_{ij}^* = \beta x_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

Here, x_i is a vector of covariates and β the coefficients to be estimated. The latent scale Y_{ij}^* corresponds to the observed ordinal responses y_{ij} in the following way:

$$y_{ij} = 1 \qquad \text{if } Y_{ij}^* < \tau_1$$

$$y_{ij} = 2 \qquad \text{if } \tau_1 \le Y_{ij}^* < \tau_2$$

$$y_{ij} = 3 \qquad \text{if } \tau_2 \le Y_{ij}^* < \tau_3$$

$$y_{ij} = 4 \qquad \text{if } \tau_3 \le Y_{ij}^* < \tau_4$$

$$y_{ij} = 5 \qquad \text{if } Y_{ij}^{*} \geq \tau_4$$

Where, $\tau_1 < \tau_2 < \tau_3 < \tau_4$ represent the thresholds on the latent scale. These thresholds are equal for all observations.

In addition, the HOPIT model is estimated. The model allows for the thresholds to vary across observations based on the information gained from vignettes. In the estimation of the cut-points an unobserved latent variable Y_{ij}^{v*}, with a mean of μ_{ij}^{v} and a variance of 1 is assumed. With *i* indicating the individual, *j* the domain, and the *v* the vignette component. Formally, the latent variable Y_{ij}^{v*} is described by:

$$Y_{ij}^{v*} = \alpha + \varepsilon_{ij}^{v}, \qquad \varepsilon_{ij}^{v} \sim N(0,1)$$

The observed vignette ratings y_{ij}^{v} correspond to Y_{ij}^{v*} by person-specific thresholds τ_i^k:

$$y_{ij}^{v} = k, \text{ if } \tau_i^{k-1} \leq Y_{ij}^{v*} < \tau_i^{k}$$

where, k=1,…., K are the observed responses, $\tau^0 < \tau_i^1 < .... < \tau_i^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. The model assumes that the order of the cut-points does not vary across different vignettes within a particular domain. Furthermore, the cut-points are functions of covariates:

$$\tau_i^k = X_i' \gamma^k$$

The second part of the HOPIT model consists of the estimation of the respondent's self-report question, which corresponds to the vignettes. The ordinal responses are assumed to correspond to a 'true latent variable' of responsiveness of a domain Y_i^{s*}, with mean μ_i^s and variance σ^2, superscript *s* refers to the self-report questions. The latent variable Y_i^{s*} can be described by the following function:

$$Y_i^{s*} = Z_i \beta + \varepsilon_i^s, \qquad \varepsilon \sim N(0, \sigma^2)$$

Here, Z is a vector of covariates, and β is a vector of coefficients that needs to be estimated. The observed categorical responses on the self-report question y_i^s relates to Y_i^{s*} by:

$$y_i^s = k, \text{ if } \tau_i^{k-1} \leq Y_i^{s*} < \tau_i^{k}$$

Here, k=1,…., K are the observed responses, $\tau^0 < \tau_i^1 < .... < \tau_i^{K-1} < \tau^K$, $\tau^0 = -\infty$, $\tau^K = \infty$, and τ_i^k are defined as in the first part of the model.

In the ordered probit and the HOPIT model predicted values of performance per domain are estimated on a latent scale. These values need to be rescaled, since the latent scale has no quantitative interpretation. In the World Health Survey 2000 the vignette with the 'best' performance was set to one and the vignette with the 'worst' performance was set to zero. Predicted values above the highest vignette were considered as 'talent', meaning that higher responsiveness scores were not necessary, and were set equal to one. This approach suggests that the 'best' vignette is the highest needed responsiveness of that particular

domain, and the 'worst' vignette is the worst possible responsiveness of the domain. However, for responsiveness no natural zero point exists. It is implausible to maintain that the 'worst' vignette is the worst responsiveness level attainable, and the best vignette is the maximum responsiveness level needed. Alternatively, the observation with the worst performance is set to zero and the observation with the best performance is set to one. This implies that a score of one is not the best achievable responsiveness-level, but the best score attained in the sample. Similar, a score of zero is not the worst imaginable responsiveness-level, but the worst responsiveness accomplished in the sample.

In sum, the HOPIT model adjusts the ratings of the self-reported questions to correct for reporting bias. As a result, the individual responsiveness scores of the domains are on a comparable scale, and are scaled relative to the best and worst performance attained in the sample.

Weights

The scores of the separate domains aggregate into one overall responsiveness score. It is inappropriate to use equal weights for all domains, since these domains are not considered equally important. In this paper three methods of weighing are considered: universal weighing, country-level weighing, and clustered weighing.

The weights are based upon self-report ordinal ratings. These ratings can be assumed to refer to a continuous 'latent importance scale' with cut-points or thresholds between the categories. The distances between cut-points on this 'latent scale' do not have to be of equal length. For example, it is likely that the distance between not important and slightly important is much larger compared to the distance between extremely important and very important. This property complicates a simple summation of these importance scores.

The method proposed by the WHO (Valentine et al. 2003b) to determine weights consists of two somewhat arbitrary choices. The WHO calculates country-level weights by running an ordered probit model with the importance scores as dependent variables and interaction terms of country dummies multiplied by domain dummies as independent variables. Here, one interaction term must be set to zero to identify the model. The WHO has chosen the first domain of the first country in the alphabet, which has no influence on the absolute differences between the estimated coefficients, but the relative differences will alter. These relative differences between coefficients are used to determine weights. In order to generate weights from the coefficients, the lowest coefficient is set as anchor weight, which also is arbitrary to some extent. This anchor weight is set to a value of 2%. The WHO argued: "*This choice reflects the notion that the mean weight in any particular country is unlikely to be zero even for the domain with the lowest relative importance. However, a low non-zero value has been chosen for the anchor because values for the minimum imply greater variation*

across domains and countries." Both these choices can affect the outcome of the responsiveness score importantly.

If the ordinal ratings are cardinalized, which is a transformation from an ordinal to an interval scale, a simple interval regression becomes possible. Wagstaff and Van Doorslaer (1994) discuss this kind of cardinalisation of ordinal data. The importance score y_{ij} consists of 5 categories, which is related to the latent scale of importance Y_{ij}^{*}:

$$y_{ij} = k, \text{ if } \tau_i^{k-1} \le Y_{ij}^{*} < \tau_i^{k}$$

where, k=1,…., K are the observed responses, $\tau^0 < \tau_i^1 < .... < \tau_i^{K-1} < \tau^K$ and $\tau^0 = -\infty$, $\tau^K = \infty$. If information of some sort of 'golden standard' is available then the thresholds on the 'latent scale' can be obtained relatively easy. However, this kind of information on importance is not available. A strategy to obtain values of the thresholds on the latent scale is assuming a distribution of the observations on the latent scale. Using this distribution, the length of the categories on the latent scale, and thereby the thresholds can be estimated.

As mentioned before, cardinalisation enables interval regression. Herein, interaction terms of domain dummies and country dummies are used as independent variables. This estimation is described by:

$$(y_l^k, y_u^k) = D_i^{'}\beta$$

Where, y_u^k represents the upper limit, y_l^k the lower limit, $y_l^1 = -\infty$, $y_u^5 = \infty$, D represents the interaction terms and β a vector of coefficients. Next, the predicted values on the latent scale are used to construct weights.

In the set-up of the questionnaires 'one' represents 'extremely important' and 'five' represents 'not important'. In order to construct weights it is convenient for the highest score to represent the highest level of importance. Therefore, the predicted values of importance are transformed by taking the value of five minus the predicted value. Thereto, it is assumed that a predicted value of five is equal to an importance of zero. This is justified since the value of five is labeled as 'not important' in the questionnaire. Thereafter, these transformed predicted values of importance scores are used to generate weights. Universal weights are calculated by dividing the sum of the mean importance score of one domain by the sum of the mean importance scores of all domains. The average predicted score per country is used to ensure equal weighing of all countries. Country-level weights are calculated similarly. These weights are formed by dividing the mean importance score of one domain of a country by the sum of the mean importance scores of all domains of this country. Finally, weights of clustered countries are calculated. Essentially, the method of calculation of the weights does not differ from the method for country-level and universal weights, except that here the mean importance scores of the clusters are applied.

Generating the responsiveness score

The responsiveness score is a weighted average of domain scores and is calculated separately for in- and outpatient care, both using the same methodology. In the Health System Performance Assessment; Debates, Methods and Empiricism (Murray et al., 2003) the WHO proposed to attach equal weights to in- and outpatient care scores in the calculation of the final responsiveness score, which seems rather arbitrary. Unfortunately, preferences concerning the importance of in- and outpatient care are not available. As an indication of the importance of in- and outpatient care, the division of costs to both sectors is used to generate weights, since it can be expected that more is paid to the sector that is valued as most important.

Data

The data used in this analysis is from The WHO Multi-country Survey Study on Health and Responsiveness 2000-2001. This survey consisted of both a long and a short questionnaire. Data from 67 countries was available; the long questionnaire was exercised in 53 countries and the short questionnaire in 14 countries.

Responsiveness variables: responsiveness scores and vignettes

The respondents rated the domains of responsiveness from very bad to very good based on their actual experiences. The related questions of the questionnaire are presented in appendix 1. The distributions of the responses are displayed in table 1. In addition, this table shows the distribution of the responses to the vignettes. Respondents were asked to rate hypothetical cases for two domains. Consequently, each vignette was rated by approximately one quarter of the sample since there are eight domains in total. The vignettes were identical for all respondents regardless if they answered questions concerning inpatient or outpatient care.

Responses with missing data on one of the domains or with missing demographical information were omitted from the dataset. Countries with less than 30 observations on one of the domains were omitted from the dataset in order to enable convergence of the HOPIT model. As result Austria, Belgium, Denmark, Ethiopia, Guatemala, Ireland, Luxembourg, Mali, Namibia, Senegal and Swaziland were omitted from the sample. The remaining data consists of 48362 observations for inpatient care and 88395 observations for outpatient care divided over 56 countries.

Importance score

Questions regarding the preferences of respondents for the different aspects of responsiveness were included in the long questionnaire. After a short description of the domain, the respondents had to rate the importance of this domain. Per domain they could choose from

extremely important to not important at all. Table 2 displays the distribution of these ratings. It is noticeable that most respondents found the domains extremely or very important.

Socio-economic and demographic variables

Variables that indicate socio-economic status and demographic characteristics are included in the estimation of the model. These variables are: age, sex, years of education and self reported health. These variables are expected to influence the attained level of responsiveness. The way individuals are treated may differ between individuals of different ages, and differ between men and women. In addition, years of education may be of influence on how patients are treated. Responsiveness towards higher educated patients may be higher, since the higher social status of these patients. Finally, self-reported health is included as it is expected that unhealthy patients are treated differently than relative healthy patients. Furthermore, these socio-economic and demographic variables are used in the estimation of the cut-points, because these variables can be expected to influence reporting behavior. For example, women may rate performance systematically lower than men while the actual experience does not differ.

Finally, these variables are included in the model that estimates the importance ratings. These variables are expected to influence the preferences for the different items of responsiveness. Descriptive statistics of these demographic variables are displayed in table 3a-b. Self-reported health is rated from 1 to 5, i.e. from very good to very bad. Two separate tables for inpatient and outpatient care are given since the analyses are done separate.

Data from additional sources

The data described before is available in the World Health Survey 2002. However, importance ratings are only available in the long-questionnaire, therefore the data for short questionnaire countries is estimated, for which additional data is required. Extra variables in this estimation are: GDP per capita and healthy life expectancy in 2002. GDP per capita is expressed in purchaser power parity US dollar. Both variables were retrieved from the website www.worldmapper.org. The indicated source of the value for GDP per capita is: United Nations Development Programme (UNDP) Human Development Report 2004 Table 13. Source in turn is the World Bank. 2004. World Development Indicators 2004. CD-ROM. Washington, DC.; aggregates calculated for the Human Development Report Office by the World Bank. The indicated source for healthy life expectancy is: United Nations Environment Programme (UNEP) (2005): Series: Healthy Life Expectancy (HALE). Finally, data of the OECD Health Data is used to construct weights for inpatient and outpatient care. The expenditures on inpatient and outpatient care are from 2002, which corresponds with the year of the household survey of the WHO. The data is obtained from the online database OECD Health Data 2006.

Results

Calculation of domain scores

In order to assess the presence and magnitude of response heterogeneity both an ordered probit model and a hierarchical ordered probit (HOPIT) model are estimated. These models obtain the predicted values on the latent scale of responsiveness per domain. In both models the independent variables were age, gender, health, years of education and country-dummies. The same independent variables were used to estimate the cut-points on the latent scale of the domains in the HOPIT model. Table 4 presents the output of the estimation of the cutpoint between good and very good of the domain dignity of outpatient care. Solely the output of this cutpoint is shown due to length constraints.

Reporting heterogeneity in rating responsiveness is found to be present, since most of the coefficients are significant, which proves that respondents from different countries, education levels, and health status use different values for the cut-point between good and very good on the latent scale of dignity. The influence of reporting behaviour on the final responsiveness score will be evaluated in a later stage.

Calculation of weights

The second part in calculating the overall responsiveness score consists of weighing and aggregating the predicted domain scores. Three methods to obtain these domain-weights are explored in this paper: universal weights, clustered weights and country-level weights. Weights are calculated using importance-ratings of the household survey. First, these ordinal self-reported ratings were cardinalized. Consequently, an assumption concerning the distribution of the self-reported ratings was necessary since no 'golden standard' of importance was available. In figures 2a-b the distribution of the importance-scores is displayed. As shown in these figures, the distribution is obviously skewed to the left. Therefore, log-normal distribution of the data was assumed. The values of the cut-points between categories on the latent importance scale are calculated using this distribution. The importance ratings were not included in the short questionnaire, so these ratings had to be estimated using data from the long questionnaire. Cardinalisation enabled an interval regression on the data, in which GDP in Purchaser Power Parity, years of education, healthy life expectance, age, age squared, sex and domain dummies were independent variables. This combination of variables was found to be the best predictor of the importance ratings. Table 5 illustrates the coefficients and the z-values.

Next, the results of this interval regression were used in order to form clusters of countries. The variable with the largest explanatory power was determined as basis for this clustering. This was executed using fully standardized coefficients. Fully standardized coefficients are completely unit free and therefore appropriate to explore the explanatory

power of variables. Healthy life expectancy was found to explain most of the variation in importance ratings, and was therefore used to form three clusters of countries. The calculated country-level, universal and clustered weights are displayed in table's 6a-d. The range in weights is the smallest for cluster three countries, which is the cluster with the highest life expectancy, with the lowest weight of 0.1413 and the highest weight of 0.1455. The largest range in weights is in country-level weights of Zimbabwe, where the domain choice gets 0.0413 and dignity gets 0.1696. Both clustered and universal weights do not vary much. Most variation in weights is found in country-level weights of developing countries, in contrast to developed countries that have more equal distributed weights.

Responsiveness-score

The composite responsiveness scores on inpatient- and outpatient care were finally aggregated into a final responsiveness score using the average division of health care costs in OECD countries. Unfortunately, no data of non-OECD countries was available. The final responsiveness scores of the countries are displayed in table's 7a-c. Scores based on clustered weights are presented in separated rankings, since countries in cluster 1 have different weights than countries in cluster 2 and 3, implicating that comparison between clusters is not straightforward anymore. A ranking of country-level weights is formed since weights vary per country and reflect the preferences of countries separate. The rankings using the different weighing methods are very similar, indicating a small influence of weighing method on the results. Some countries, for example France and Paraguay, switch ranking, but no dramatic changes are caused. In all weighing methods Finland keeps the best performance on overall responsiveness.

The influence of response heterogeneity on the outcomes

Both the predicted values of the domain scores of the HOPIT and the ordered probit model were used to calculate the final responsiveness scores in order to assess the influence of the reporting bias on the outcomes. The results using universal weighting are presented in table 8. The differences between the results of the two methods are large. When using the ordered probit model Greece has the highest overall responsiveness score, however when using the HOPIT model Greece falls to rank 11. This indicates that Greeks are more likely to report higher ratings relative to other countries for the same performance on responsiveness. In contrary, the Dutch score estimated with the HOPIT model (0.658) is higher than the score estimated with the ordered probit model (0.575), indicating that the Dutch are more critical and tend to rate performance lower than people in other countries.

Discussion

In the World Health Report 2000 the WHO included responsiveness into the measurement of health system performance, which in our view represents process utility. The inclusion into the framework is essential since utility is not only determined by the outcome health, but also by the process of health care. The large part of the health care budget spent on the process of health care stresses its importance. Therefore, neglecting process utility will lead to an incomplete and disturbed image of total performance. The methodology used by the WHO in the World Health Report was criticized on some aspects in the scientific literature. The WHO acknowledged some of the critiques and proposed improved methods in Health System Performance Assessment in 2003. In this paper the responsiveness-scores are calculated using new data of the World Health Survey 2002 and partly the suggested improved methods. Three main aspects were the focus in this paper: assessment of the response heterogeneity in the self-reported responsiveness, exploring methods for the determination of weights and the relative influence of these weights, and calculating final responsiveness-scores.

The presence of reporting heterogeneity in self-reported responsiveness was assessed by exploring the significance of the coefficients in the estimation of the cut-points in the HOPIT model. Most coefficients were found to be significant indicating that respondents from different countries, ages, sex, health status and education level rate responsiveness on a different scale. Put in other words: response heterogeneity is present. This result is similar to the findings in reporting heterogeneity in health where reporting heterogeneity was also found to be present. However, in health the results did not differ much when correcting for this reporting bias. This is different for responsiveness. The rankings of the countries based on the final responsiveness scores using respectively the ordered probit and the HOPIT model presented in table 8 show large differences. These differences in results indicate the large influence of reporting heterogeneity especially on the ranking of countries. Countries reporting relatively high ratings as compared to the attained responsiveness, so countries with a higher responsiveness score using the ordered probit model compared to the HOPIT, are mainly developing countries. This indicates the higher standards of developed countries and thereby the relatively higher thresholds on the latent scale of responsiveness of these countries. In sum, reporting heterogeneity is found to be present in the reporting of health system responsiveness, and not correcting for this heterogeneity will lead to biased results.

Next, the weights for aggregation of the scores were calculated. The results of the three methods, especially of universal weights and clustered weights, did not vary dramatically. For universal weights the lowest weight is 0.1105 for choice and the highest is 0.1339 for communication for inpatient care. First, this relative small variation is partly due to the method of questioning. The answers to the importance of every domain were mostly in the

categories extremely and very important. A conjoint analysis study could make a better evaluation of the importance of the domains. Respondents are obliged to make trade-offs in these studies, which reflects a more realistic situation since policy makers need to make trade-offs in policy determination due to limited health care budgets. Second, the universal weights and the clustered weights are basically averages of the overall preferences. If in country A autonomy is found to be more important, and in country B communication is found to be more important, then the aggregate weights will balance each other out.

Most variation is found within the country-level weights. The lowest weight is 0.04 to the domain choice of outpatient care for Zimbabwe, and the highest is 0.1824 to the domain quality of basic amenities for Congo. Countries with most variation are low-income countries. High-income countries, like Spain, have less variation in weights. A possible explanation is that these countries are wealthier and therefore demand a high standard on all aspects of responsiveness. In contrast, developing countries are less demanding, and probably regard the elementary domains to be most important. Furthermore, the observed variation in weights is much smaller for short questionnaire countries, for which there are two possible explanations. First, the responsiveness ratings were not included in the short questionnaire. Consequently, these ratings were estimated using the importance ratings of the long questionnaire. Obviously, not all variation could be explained by the independent variables, resulting that the predicted values of importance and consequently the weights for short questionnaire countries varied less compared to the values of long questionnaire countries. Second, short questionnaire countries are mainly high-income countries. High-income countries with the long questionnaire also showed little variation in weights, which indicates that high-income countries tend to value all domains more equal as compared to low-income countries.

The final rankings in table's 7a-c resulted from the weighted aggregation of the domain scores. It should be noted that the predicted outcomes are rescaled to the best and worst observation, and thus is based upon the observed variation. The influence of this rescaling on the scores depends on the lowest and the highest observation and on the distance from the natural zero point – the worst responsiveness possible. Here we are restricted to the use of the observed variation, since the natural zero point of responsiveness is unknown.

The three different methods of weighing give some insight into country specific preferences and their attainment. For example, Norway scores higher when country-level weights are used compared to universal weights. Therefore, the health care system in Norway is more responsive to the preferences of its population than for the 'universal preferences'. Generally, the values of the scores do not vary much per method. Some countries switch places if different weighing methods are applied, but no dramatic changes in ranking occur. The relatively small influence of weighing methods on the final rankings may not primarily be contributed to the weights. It needs to be realized that the effective weights are of influence

on the final responsiveness score. The effective weight depends upon the nominal weight but also upon the variation in the scores to which they are applied (Richardson et al., 2003). For example, if there is little variation in the scores than the weights do not have an influence on the rankings of the countries. Nonetheless, it should be realized that little differences could have a larger influence than first assumed if the responsiveness scores are used to compute complete health system performance, depending on the weight attached to responsiveness.

The determination of a favorable method of weighing is complicated. De Silva (2000) states: "If a normative judgment existed with regard to the relative importance of the elements, then country level weights would be contrary to the objective of achieving a universally determined optimal standard of health system responsiveness." If such a normative judgment indeed exists then universal weights are justified and countries can be compared and ranked. However, no such normative judgment is available for responsiveness. Restricting on averages of values of the elements, in the form of universal domain weights, leads to weights that do not apply to any country, and therefore will not reflect preferences of any country. The second method applied is country-level weighing. Country-level weighing is a direct reflection of preferences of the population, however it reduces the comparability of the final scores. The final scores are interactions of the weights and the domain scores, and therefore depend on both. This could entail that in an extreme case a country with a higher score on each domain could rank lower than another country with lower scores on all domains. The alternative is the use clustered weighing. Here, countries with comparable preferences are compared and weighted identical. This has the advantages that weights have a meaning, since countries are clustered in homogenous groups and the final scores within a cluster are comparable since identical weights are used. The aim of country rankings is to learn about the state of the health care system and to generate opportunities to improve it. If countries are hard to compare, like countries in different stages of development, than the primary goal of ranking is not achieved. By clustering countries according to Healthy Life Expectancy, the comparability and usefulness of the rankings are enlarged, since countries in the clusters are more similar in terms of economic and health system development.

Some countries are ranked unexpectedly high with all weighing methods. The most unexpected results appear to be the high rankings of the Dominican Republic and Paraguay. The Dominican Republic and Paraguay are both low middle-income countries and the level of the health care system is far lower compared to Western Countries, such as the Netherlands. Unexpectedly, the Dominican Republic attains a fairly identical responsiveness-score and Paraguay attains even a higher score than the Netherlands. These unexpected results give the impression of data problems. The mean scores of the vignettes and the self-reports on the domain of autonomy of inpatient care of the Dominican Republic, Paraguay and all other countries are plotted in figure 3. The mean self-reported scores in the Dominican Republic

and Paraguay are higher than the "best vignette", while in the other countries the value of the self-reported score falls between the second and the third vignette. This confirms the presence of data-problems with the vignettes in both countries, since it does not seem realistic that the level of responsiveness in these countries is higher than the level of the best vignette.

In the calculation of the responsiveness score non-users were excluded. These non-users are the part of the population that wants to make use of health care, but are excluded. Therefore, the responsiveness score measures the responsiveness of care that is actually delivered. Consequently, countries with high rates of non-users, such as Paraguay, have a high ranking in spite of the low coverage of health care. It seems unreasonable that such a country scores higher on responsiveness than a country serving all people in need of health care. The inclusion of non-users into the analysis is not straightforward and therefore it is not performed in this study. To give an impression of the consequences of inclusion of non-users into the analysis, results of calculations with non-users are displayed in appendix 2. Many differences in rankings using both methods are found. Some countries score significantly lower when non-users are included, not only indicating the good responsiveness to users, but also the lack of ability to serve the entire population.

In order to give a quick assessment of the validity of the study, the correlation between the results of this study and the results of the World Health Report 2000 is calculated. The correlation for all weighing methods is around 0.69. This number indicates the differences between both results, which is due to several reasons. First, the analyses are based on different datasets. In this research, a household survey is used as compared to the key-informants in the WHR 2000. Moreover, much of the data was estimated in the WHR, while most data in this paper is directly observed in the household survey. Second, the methodology used here is not the same as in 2000 with the largest difference in the method of determining weights and the rescaling of the predicted observations.

A theoretical economic framework that justifies the inclusion of responsiveness into health system performance framework is included in the beginning of this paper. In evaluating the methodology of the measure of responsiveness, reporting heterogeneity was found to be present and its influence on the outcomes of responsiveness large. The results would be biased without correcting for this. Further, three methods of attaining weights were compared. The methodology of clustered weighing was argued to be best applicable, since forming homogenous groups of countries ensures both comparability and reflection of preferences in weights. Besides, the best lessons are learned from countries with similar conditions, and difficulties. Finally, rankings of countries based on the three methods were presented with Finland as the number one in all rankings. Improvements in the data and methodology of vignettes, importance-scores (weights) and non-users are still needed in order

to optimize the estimation of process utility in health care; however these results give a good indication of the attainment of responsiveness of health care systems over the world.

References

Ammar et al., Report of the Scientific Peer Review Group on Health Systems Performance Assessment, Responsiveness - Level and Distribution, members of scientific peer review group with institutional affiliations, 2001, www.who.int/health-systems-performance/sprg/report_of_sprg_on_hspa.htm,

T.B. Bago d'Uva, E. van Doorslaer, M. Lindeboom, O. O'Donnell, S. Chatterji. Does Reporting Heterogeneity bias the Measurement of Health Disparities?. Health Economics 2008; 17(3); 351-75.

M. Benz. The Relevance of Procedural Utility for economics, Institute for empirical research in Economics. Working Paper Series 2005, University of Zurich,

R.J. Blendon, M. Kim, J.M. Benson. The public versus the WHO on health system performance. Health affairs 2001; 20-3; 10-20

A. Domingues Uga, C.M. de Almeida, C. Landmann Szwarcwald, C. Travassos, F. Viacava, J. Mendes Ribeiro, N. do Rosario Costa, P.M. Buss, S. Porto. Considerations on methodology used in the World health organization 2000 report. Cad. Saude Publica; 17(3); 705:712

Deci, E.L. & R.M. Ryan. 'The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior.' Psychological Inquiry 2000*;* 11(4); 227-268.

B.S. Frey, M. Benz, A. Stutzer. Introducing procedural utility: not only what, but also How Matters. Journal of institutional and Theoretical Economics 2004; Vol. 160; Nr. 3; 377-401

B.S. Frey, A. Stutzer. Beyond Outcomes: Measuring Procedural Utility, Berkeley Program in Law & Economics. Working Paper Series, University of California, Oxford Economic Papers 2005.

A.M. Jones, N. Rice, T. Bago d'Uva, S. Balia. Applied Health Economics. Routledge Advance Texts in Economics and Finance 2007

McKee. The World Health Report 2000: advancing the debate, European regional consultation on the WHR 2000, Copenhagen Denmark, 2001

G. King, C.J.L. Murray, J.A. Salomon, A. Tandon. Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. American Political Science Review 2004; 98; 197-207

KGN Kramers, PW Achterberg, EA van der Wilk. Leren van het buitenland; Gezondheid en zorg internationaal vergeleken: een samenvatting van drie studies, 2001, RIVM rapport 271558 004

C.J.L. Murray, D.B. Evans. Health System Performance Assessment: Goals, Framework and Overview, in Health system Performance Assessment; Debates, Methods and Empiricism, edited by C.J.L. Murray, D.B. Evans, World Health Organization, 2003

C.J.L. Murray, A. Tandon, J.A. Salomon, C.D. Mathers, R. Sadana. Cross-population comparability of evidence for health policy. Global Programme on Evidence for Health Policy Discussion Paper No. 46, WHO 2002

P. Musgrove. Judging health systems: reflections on WHO's methods. The Lancet 2003; 361; 1817-20

V. Navarro. Assessment of the World Health Report 2000. the Lancet 2000, Vol. 356, issue 9241, pg 1598-1601

V. Navarro. The new conventional wisdom: an evaluation of he WHO report "Health systems: improving performance. International Journal of health Services 2001; Vol. 31, Nr. 1, pg 23- 33

E. Nord. Measures of goal attainment and performance in the World Health Report 2000; a brief, critical consumer guide. Health Policy 2002; 59; 183-191

Oswaldo Cruz Foundation, Brazilian Ministry of Health: Report of the Workshop "Health Systems Performance: The World Health Report 2000". 2000, Rio de Janeiro.

K.M. Pedersen. The World health report 2000: dialogue or the deaf?. Health Economics 2002; 11; 93-101

J. Richardson, J. Wildman. I.K. Robertson. A critique of the World Health Organisation's evaluation of health system performance. Health Economics 2003;12; 355-366

J.A. Salomon, A. Tandon, C.J.L. Murray. Using vignettes to improve cross-population comparability of health surveys: concepts, design, and evaluation techniques. Global Programme on Evidence for Health Policy Discussion Paper No. 41, WHO 2001

J.A. Salomon, A. Tandon, C.J.L. Murray. World Health Survey Pilot Study Collaborating Group, Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes, 2004, published online on www.BMJ.com

Sen, K. Amertya. Rationality and Social Choice. American Economic Review 1995; 85(1);1-24

Sen, K. amertya. Maximization and the Act of Choice. Econometrica 1997; 65(4); 745-779

A. de Silva. A Framework for Measuring Responsiveness, GPE Discussion Paper Series: No. 32, 2000, WHO

A. van Soest, L. Delaney, C. Harmon, A. Kapteyn, J.P. Smith. Validating the Use of Vignettes for Subjective Threshold Scales. IZA Discussion Papers nr. 2860, Institute for the study of Labor, 2007

A. Tandon, C.J.L. Murray, J.A. Salomon, G. King, Statistical Models for Enhancing Cross-Population Comparability. Global Programme on Evidence for Health Policy. Discussion Paper 2002; No. 42, WHO, Geneva, Switzerland

N.B. Valentine, R. Lavallee, B. Liu, G.J. Bonsel, C.J.L. Murray. Chapter 44: Classical Psychometric Assessment of the Responsiveness Instrument in the WHO Multi-country survey study on health and responsiveness 2000-2001, in Health System Performance Assessment; Debates, Methods and Empiricism, edited by C.J.L. Murray, D.B. Evans, World Health Organization, 2003a

N.B. Valentine, J. A. Salomon. Chapter 45: Weights for Responsiveness Domains: Analysis of Country variation in 65 national sample surveys, in Health System Performance assessment; Debates, Methods and Empiricism, edited by C.J.L. Murray, D.B. Evans, World Health Organization, 2003b

N.B. Valentine, A. de Silva, K. Kawabata, C. Darby, C. J.L. Murray, D.B. Evans. Chapter 43: Health system responsiveness: Concepts, domains and operationalization, in Health system Performance Assessment; Debates, Methods and Empiricism, edited by C.J.L. Murray, D.B. Evans, World Health Organization, 2003c

A. Wagstaff, E. van Doorslaer. Measuring inequalities in health in the presence of multiple-category morbidity indicators, Health Economics 1994; 3; 281-291

A. Williams, Science or marketing at WHO? A commentary on 'World Health 2000'. Health economics 2001a; 10; 93-100

A. Williams. Science or marketing at WHO? Rejoinder from Alan Williams. Health economics 2001b; 10; 283-285

World Health Organization. The World health report 2000; Health Systems: Improving Performance, http://www.who.int/whr/2000/en/whr00_en.pdf

Tables and figures

Figure 1: Response heterogeneity

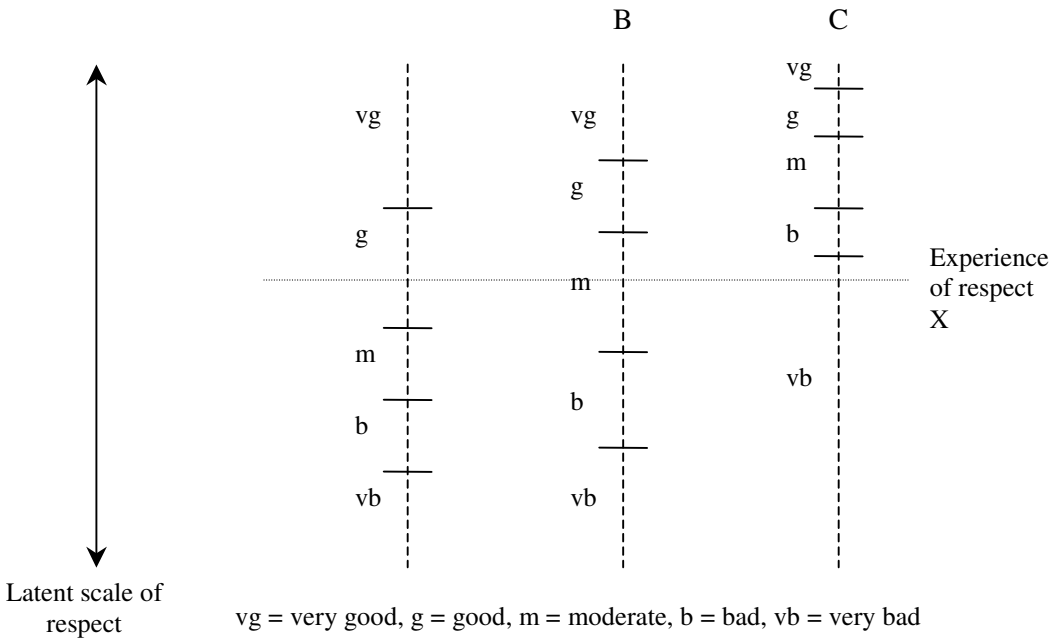

Table 1: Frequencies of responsiveness score and vignettes by domain

	Inpatient care						Outpatient care					
	Score	*Vig1*	*Vig2*	*Vig3*	*Vig4*	*Vig5*	*Score*	*Vig1*	*Vig2*	*Vig3*	*Vig4*	*Vig5*
Autonomy												
Very good	14.71	1.83	2.44	13.32	41.51	26.02	13.87	2	2.73	12.76	39.28	25.43
Good	50.37	7.43	7.22	35.12	39.64	45.83	54.61	7.42	7.62	34.41	40.56	46.67
Moderate	24.09	20.64	19.77	28.21	12.44	19.56	22.69	20.32	20.49	29.44	13.38	19.21
Bad	8.63	51.69	48.86	19.26	4.9	6.47	7.37	53.13	49.09	19.02	5.23	6.67
Very bad	2.2	18.41	21.71	4.09	1.51	2.12	1.46	17.13	20.07	4.37	1.55	2.02
N	48362	11807	11798	11801	11786	11690	88395	21464	21456	21444	21408	21282
Choice												
Very good	13.90	4.20	27.35	2.93	7.89	47.30	15	3.9	26.07	3.07	7.82	44.58
Good	44.40	17.31	43.48	10.13	28.15	34.85	51.34	17.03	44.16	10.74	28.5	37.59
Moderate	24.68	37.98	17.47	25.69	33.55	11.16	21.56	39.6	17.42	25.67	34.12	11.14
Bad	12.07	33.23	9.62	42.08	24.06	4.63	9.17	32.28	9.95	42.9	23.76	4.43
Very bad	4.95	7.27	2.08	19.16	6.35	2.05	2.93	7.19	2.4	17.63	5.8	2.26
N	48362	11869	11851	11831	11833	11815	88395	21789	21765	21756	21744	21703
Communication												
Very good	21.02	38.90	49.07	6.87	1.73	1.61	20.07	37.5	47.19	6.74	1.62	1.49
Good	55.46	47.26	39.24	24.45	4.39	6.84	59.67	49.05	40.85	24.77	4.39	7.29
Moderate	17.70	9.77	7.37	38.26	8.93	25.97	16.24	9.45	7.68	38.61	8.88	27.35
Bad	4.55	3.29	3.02	24.47	40.72	46.69	3.36	3.31	3.08	23.97	43.63	46.3
Very bad	1.27	0.78	1.30	5.95	44.23	18.89	0.66	0.69	1.21	5.9	41.49	17.57
N	48362	11994	11984	11975	11978	11959	88395	21681	21657	21650	21653	21640
Confidentiality												
Very good	17.41	1.94	5.39	2.30	2.68	43.65	17.58	2.05	5.13	2.68	2.78	40.84
Good	54.73	7.13	22.33	8.21	8.54	36.19	59.44	7.1	22.25	8.56	9.26	37.72
Moderate	21.03	14.18	23.43	11.31	19.66	11.91	18.53	13.87	23.93	11.88	19.98	12.6
Bad	5.61	50.80	33.47	43.61	51.78	6.45	3.8	52.28	34.26	45.43	51.49	7.04
Very bad	1.22	25.95	15.38	34.58	17.34	1.81	0.66	24.69	14.43	31.45	16.49	1.8
N	48362	11886	11861	11846	11836	11836	88395	21812	21773	21787	21763	21755

	Inpatient care						Outpatient care					
	Score	*Vig1*	*Vig2*	*Vig3*	*Vig4*	*Vig5*	*Score*	*Vig1*	*Vig2*	*Vig3*	*Vig4*	*Vig5*
Dignity												
Very good	22.77	3.43	32.46	50.31	11.41	1.42	21.47	3.43	30.51	47.23	11.42	1.37
Good	56.36	13.79	45.62	37.93	35.95	3.51	61.17	14.8	46.74	39.97	37.78	3.6
Moderate	16.31	23.85	13.07	6.34	31.83	6.47	14.46	24.84	13.95	6.83	30.82	6.84
Bad	3.45	43.66	6.58	3.26	17.29	29.84	2.31	43.39	6.35	3.61	16.3	30.16
Very bad	1.11	15.27	2.28	2.16	3.52	58.76	0.6	13.55	2.44	2.37	3.68	58.03
N	48362	12011	11993	12003	11987	11988	88395	21949	21926	21948	21937	21916
Prompt attention												
Very good	18.88	31.95	1.63	2.59	47.79	1.81	13.37	29.35	1.61	2.93	46.88	1.86
Good	51.31	44.77	4.03	4.24	38.36	11.13	51.56	47.13	4.42	4.84	38.7	11.86
Moderate	19.81	16.56	9.41	6.56	9.12	37.12	23.36	17.05	9.07	6.43	9.44	37.34
Bad	7.47	5.78	39.16	25.31	3.27	38.81	9.22	5.27	39.29	25.99	3.56	37.89
Very bad	2.54	0.94	45.78	61.30	1.46	11.13	2.49	1.2	45.6	59.82	1.43	11.05
N	48362	12020	11997	11992	11993	11982	88395	21952	21928	21927	21939	21916
Quality of basic amenities												
Very good	20.37	35.33	63.59	1.85	1.32	1.30	19.5	33.13	61.62	1.66	1.27	1.16
Good	51.81	47.15	28.14	3.97	4.38	4.88	56.82	48.47	29.59	3.86	4.39	5.07
Moderate	19.83	13.14	5.66	6.31	12.18	17.28	19	13.92	5.64	6.05	12.17	17.14
Bad	5.79	3.68	2.06	27.21	46.88	47.14	3.35	3.7	2.52	29.86	48.37	48.83
Very bad	2.21	0.71	0.55	60.67	35.24	29.41	1.3	0.79	0.61	58.58	33.79	27.8
N	48362	11990	11974	11978	11966	11958	88395	21674	21670	21654	21652	21639
Support												
Very good	19.88	2.51	1.63	4.98	47.39	33.02						
Good	55.35	11.27	4.11	24.49	38.46	45.19						
Moderate	19.51	32.82	8.99	43.78	9.06	16.18						
Bad	4.22	44.76	45.34	22.34	3.51	4.57						
Very bad	1.04	8.65	39.93	4.41	1.58	1.04						
N	48362	11814	11807	11811	11806	11788						

Table 2: Frequencies of importance ratings by domain

	Extremely important	*Very important*	*Moderately important*	*Slightly important*	*Not important*	*N*
Autonomy	28.87	52.04	13.93	3.7	1.47	87739
Choice	28.98	50.81	13.89	3.98	2.34	87764
Communication	38.86	51.77	7.49	1.46	0.42	87619
Confidentiality	37.77	51.82	7.83	1.69	0.9	87777
Dignity	39.33	53.3	6.26	0.8	0.31	87688
Prompt attention	35.84	52.52	9.72	1.58	0.34	87776
Quality of basic amenities	38.91	52.22	7.48	1.19	0.2	87770
Support	30.95	51.77	13.21	3.11	0.95	87781

Table 3a: Descriptive statistics for inpatient care

	Number of obs	*Mean age*	*St dev age*	*% women*	*Mean # of yrs of education*	*St dev # of yrs of education*	*Mean self reported health*	*St dev self reported health*
All	48362	40.26	15.72	65	8.11	5.06	2.41	0.94
United Arab Emirates	251	35.49	11.70	61	11.04	5.23	1.78	0.92
Burkina Faso	592	33.98	12.31	67	3.35	5.29	2.22	0.87
Bangladesh	743	36.78	13.31	56	5.81	4.44	2.75	0.97
Bosnia Herzegovina	326	47.36	16.49	58	9.08	4.39	2.60	1.15
Brazil	1510	40.08	15.87	69	6.68	4.89	2.61	0.92
China	568	45.86	15.94	61	8.12	4.81	2.62	0.97
Cote d'Ivoire	330	35.68	13.08	51	6.26	6.00	2.40	0.89
Congo	249	35.55	13.17	61	5.91	5.51	2.47	1.07
Comoros	459	42.19	16.31	57	3.06	4.79	2.58	1.00
Czech Republic	300	47.01	18.11	62	12.56	2.83	2.57	0.95
Germany	319	51.59	16.70	58	10.89	2.95	2.54	0.92
Dominican Republic	1789	37.53	15.10	68	7.51	4.98	2.47	0.86
Ecuador	797	37.87	14.29	66	8.89	4.38	2.51	0.89
Spain	1838	51.44	17.00	62	8.97	5.21	2.53	0.97
Estonia	325	47.72	17.31	69	12.06	3.26	2.86	0.85
Finland	354	50.64	16.79	56	11.64	3.77	2.47	0.86
France	210	46.58	16.16	62	13.44	4.51	2.28	0.87
United Kingdom	304	46.93	18.20	70	12.00	2.93	2.36	1.00
Georgia	403	46.04	18.22	60	11.94	3.64	2.92	0.96
Ghana	756	39.80	14.17	60	6.50	5.29	2.14	0.98
Greece	285	52.69	18.93	52	9.13	4.28	2.40	1.05
Croatia	330	51.07	16.11	64	10.47	3.99	2.66	1.11
Hungary	501	48.62	17.13	61	11.49	3.74	2.71	0.97
India	1911	37.84	14.79	54	5.30	5.04	2.60	1.04
Israël	455	44.17	16.96	61	13.34	3.98	1.98	0.97
Italy	235	50.31	16.92	59	11.02	4.90	2.51	0.94
Kazakhstan	885	42.68	14.04	70	12.99	2.75	2.76	0.73
Kenya	871	34.88	12.70	70	8.44	4.29	2.31	0.94
Lao	933	38.80	12.55	53	5.09	4.26	2.11	0.90
Sri Lanka	1855	40.04	14.34	59	8.01	3.95	2.08	0.85
Latvia	324	51.75	18.84	71	11.47	3.44	3.11	0.91
Morocco	266	35.91	11.18	64	8.71	4.74	2.69	1.03
Mexico	8309	37.22	15.08	74	7.78	4.75	2.34	0.85
Myanmar	417	40.88	14.54	69	6.86	4.53	2.20	0.73

Mauritania	510	37.34	14.19	68	3.48	4.73	2.29	0.82
Mauritius	1328	41.28	14.42	60	8.13	3.96	2.46	1.07
Malawi	1268	32.09	12.21	71	6.60	4.16	1.71	0.91
Malaysia	1685	38.41	12.57	68	8.95	4.22	2.14	0.76
Netherlands	232	46.83	17.49	75	12.42	3.51	2.41	0.83
Norway	225	45.47	15.84	38	12.27	4.11	2.10	1.00
Nepal	1377	37.17	14.37	61	4.13	4.73	2.49	0.87
Pakistan	955	38.35	14.51	48	4.45	5.29	2.24	0.91
Philippines	2125	37.88	13.40	63	8.99	3.65	2.46	0.71
Portugal	208	49.84	17.44	65	6.95	4.14	2.97	0.91
Paraguay	1364	36.18	14.08	73	7.79	4.74	2.05	0.86
Russia	861	50.34	17.63	69	11.77	3.59	3.04	0.79
Slovakia	473	39.87	14.82	72	13.24	2.91	2.42	0.90
Slovenia	196	47.03	17.04	56	11.69	3.11	2.53	0.97
Sweden	240	51.30	17.74	68	11.64	3.61	2.56	1.07
Chad	324	35.94	14.00	47	2.79	4.48	2.48	0.87
Tunisia	1226	40.11	14.29	65	6.87	5.60	2.36	0.95
Ukraine	895	46.64	17.27	70	12.32	3.27	3.14	0.84
Uruguay	680	45.25	16.92	56	9.94	4.44	2.16	0.75
Vietnam	791	40.54	13.35	58	7.79	3.75	2.66	0.79
South Africa	620	38.70	14.01	60	8.99	5.77	2.20	1.06
Zimbabwe	779	34.55	13.15	76	8.30	3.40	2.52	0.97

Table 3b: Descriptive statistics for outpatient care

	Number of obs	*Mean age*	*St dev age*	*% women*	*Mean #r of yrs education*	*St dev yrs of education*	*Mean self reported health*	*St dev self reported health*
All	88395	40.5619	15.63	58	7.76	5.2	2.28	0.88
United Arab Emirates	431	37.34	11.87	47	11.93	5.30	1.75	0.73
Burkina Faso	1010	34.36	12.71	56	2.90	4.93	2.19	0.83
Bangladesh	2226	36.76	13.81	50	5.48	4.36	2.63	0.91
Bosnia Herzegovina	403	47.43	17.28	57	9.50	4.53	2.36	0.96
Brazil	2230	41.32	15.35	57	7.16	5.01	2.42	0.88
China	1395	44.68	14.14	56	7.18	4.40	2.32	0.91
Cote d'Ivoire	536	34.38	13.01	44	6.06	6.00	2.34	0.89
Congo	171	33.08	11.62	48	8.48	5.48	2.41	0.93
Comoros	530	40.97	15.93	58	3.89	5.35	2.51	1.01
Czech Republic	349	45.29	16.31	53	12.47	2.52	2.25	0.87
Germany	333	48.73	16.50	59	10.95	3.00	2.11	0.80
Dominican Republic	1249	43.14	16.14	55	6.85	4.92	2.44	0.87
Ecuador	1280	39.69	14.54	57	9.00	4.63	2.45	0.82
Spain	2615	51.06	17.22	61	9.11	5.33	2.25	0.84
Estonia	339	47.86	15.19	69	12.40	3.19	2.67	0.77
Finland	399	51.01	15.68	56	12.12	4.08	2.24	0.78
France	255	42.64	15.04	62	14.23	4.26	1.98	0.73
United Kingdom	311	50.83	17.77	68	11.86	2.75	2.09	0.91
Georgia	729	48.15	17.23	59	12.17	3.44	2.91	0.94
Ghana	1321	39.36	14.69	60	5.89	5.22	2.01	0.91
Greece	378	48.79	17.00	53	10.22	4.39	1.99	0.95
Croatia	439	50.88	15.97	63	10.34	3.95	2.39	1.03

Hungary	383	45.31	17.19	58	12.08	3.48	2.41	0.82
India	4083	38.32	14.45	52	5.16	5.07	2.37	1.03
Israel	401	42.48	16.02	56	13.89	3.57	1.75	0.85
Italy	405	46.16	15.68	61	11.59	4.75	2.20	0.76
Kazakhstan	2322	41.43	12.63	67	13.06	2.77	2.55	0.66
Kenya	2048	36.72	14.54	58	7.91	4.60	2.17	0.89
Lao	657	37.36	13.82	53	5.17	4.35	2.01	0.86
Sri Lanka	1917	41.39	14.80	52	8.68	4.16	2.03	0.83
Latvia	252	46.98	17.80	67	11.89	3.27	2.71	0.83
Morocco	593	35.80	12.63	47	9.39	5.13	2.66	0.98
Mexico	24657	40.42	15.79	62	7.22	4.92	2.27	0.84
Myanmar	1258	40.36	14.82	59	6.53	4.46	2.18	0.74
Mauritania	416	37.04	13.62	65	3.87	4.87	2.21	0.74
Mauritius	1614	41.10	13.94	46	8.65	4.11	2.23	0.95
Malawi	2180	34.35	14.91	53	6.53	4.21	1.66	0.88
Malaysia	1828	41.14	14.39	49	8.51	4.62	2.07	0.70
Netherlands	562	41.87	18.70	68	13.28	3.51	2.02	0.66
Norway	274	43.66	15.69	47	12.55	4.08	1.90	0.80
Nepal	3117	36.40	14.18	58	3.49	4.41	2.35	0.89
Pakistan	3119	37.10	14.22	42	4.29	5.27	2.05	0.83
Philippines	3434	37.80	13.30	60	8.88	3.54	2.44	0.69
Portugal	390	47.66	17.52	62	7.36	4.31	2.72	0.96
Paraguay	2246	39.75	16.07	53	7.61	4.74	2.01	0.85
Russia	1570	50.05	16.75	66	11.92	3.56	2.81	0.78
Slovakia	979	36.64	13.99	68	13.41	2.87	2.12	0.82
Slovenia	350	45.88	17.37	54	11.78	3.44	2.45	0.90
Sweden	248	47.33	16.63	58	12.72	3.37	2.31	1.00
Chad	306	33.23	12.19	48	3.47	5.27	2.24	0.85
Tunisia	2298	40.14	15.27	54	7.65	5.87	2.22	0.93
Ukraine	937	45.60	16.57	71	12.46	3.31	2.93	0.83
Uruguay	950	45.69	16.42	57	10.46	4.59	2.06	0.72
Vietnam	1575	40.54	13.65	58	7.61	3.82	2.47	0.76
South Africa	436	39.11	14.75	58	8.50	5.63	2.17	0.99
Zimbabwe	1661	36.69	15.02	66	7.62	3.97	2.40	0.93

Figure 2a: Distribution of importance scores of inpatient care

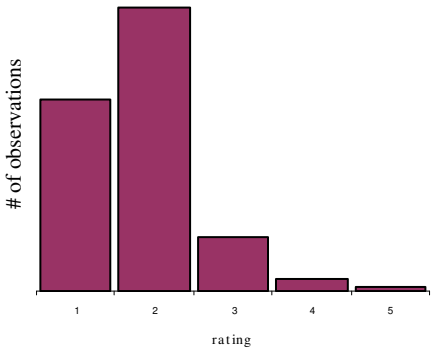

Figure 2b: Distribution of importance scores of outpatient care

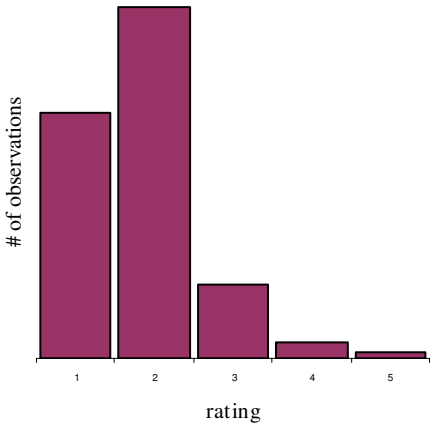

Table 4: Coefficients of cutpoint between very good and good for the domain of dignity of outpatient care using the HOPIT model

Variable	*Coefficient*	*P*>|*z*|	*Variable*	*Coefficient*	*P*>|*z*|
age	*0.0006	0.057	KEN	0.2951	0.000
gender	*-0.0045	0.597	LAO	-0.2179	0.000
health	-0.0348	0.000	LKA	*0.0431	0.202
yrsedu	0.0120	0.000	LVA	*-0.1231	0.122
ARE	0.1294	0.038	MAR	0.5475	0.000
BFA	-0.0980	0.029	MEX	-0.2195	0.000
BGD	0.0658	0.039	MMR	-0.6482	0.000
BIH	0.2773	0.000	MRT	-0.1592	0.027
BRA	0.0778	0.016	MUS	-0.2119	0.000
CHN	-0.3821	0.000	MWI	0.5293	0.000
CIV	-0.1346	0.018	MYS	-0.2050	0.000
COG	*-0.2110	0.056	NLD	-0.3517	0.000
COM	-0.3120	0.000	NOR	-0.2211	0.006
CZE	*0.1311	0.074	NPL	-0.3121	0.000
DEU	*-0.0526	0.477	PAK	-0.2391	0.000
DOM	-0.3834	0.000	PHL	-0.7318	0.000
ECU	*0.0204	0.608	PRT	-0.2070	0.002
ESP	-0.0730	0.026	PRY	0.0806	0.013
EST	*-0.1021	0.144	RUS	-0.0749	0.040
FIN	-0.2450	0.001	SVK	0.2872	0.000
FRA	-0.3478	0.000	SVN	*0.0985	0.127
GBR	*0.0561	0.445	SWE	*-0.0607	0.469
GEO	-0.1611	0.002	TCD	*-0.1156	0.148
GHA	0.0853	0.026	TUN	0.0651	0.041
GRC	0.2696	0.000	UKR	-0.1852	0.000
HRV	*0.0568	0.390	URY	-0.1890	0.000
HUN	0.3499	0.000	VNM	-0.1424	0.000
IND	*-0.0292	0.279	ZAF	0.1379	0.024
ISR	0.2888	0.000	ZWE	0.0790	0.025
ITA	*0.0432	0.499	_cons	-0.6291	0.000
KAZ	-0.1899	0.000			

* not found significant at 5% confidence level

Table 5: Coefficients of interval regression to estimate short-questionnaire

	Inpatient care		Outpatient care	
Variable	*Coefficient*	*Z-value*	*Coefficient*	*Z-value*
GDPaut	-0.00008	-51.71	-0.00008	-62.4
GDPchoi	-0.00010	-57.05	-0.00009	-64.99
GDPcom	-0.00006	-38.69	-0.00006	-45.04
GDPconf	-0.00006	-40.43	-0.00005	-37.58
GDPdign	-0.00005	-31.62	-0.00005	-36.18
GDPprom	-0.00005	-31.29	-0.00005	-38.03
GDPqual	-0.00005	-31.41	-0.00004	-32.68
GDPsup	-0.00006	-40.41	-	-
YRSEDUaut	-0.01247	-7.72	-0.00292	-2.46
YRSEDUchoi	*-0.00196	-1.21	0.00627	5.24
YRSEDUcom	-0.01991	-12.32	-0.01718	-13.97
YRSEDUconf	-0.01247	-7.74	-0.01314	-11.05
YRSEDUdign	-0.01518	-9.36	-0.01830	-15.27
YRSEDUprom	-0.01576	-9.61	-0.01162	-9.89
YRSEDUqual	-0.01576	-9.86	-0.01474	-12.45
YRSEDUsup	*-0.00121	-0.75	-	-
HLEaut	0.02823	26.41	0.03390	38.88
HLEchoi	0.02302	20.8	0.02259	24.9
HLEcom	0.01238	11.92	0.01563	18.37
HLEconf	0.02059	21.91	0.01987	27.31
HLEdign	0.01302	14.2	0.00963	13.79
HLEprom	0.01317	13.11	0.01383	18.18
HLEqual	0.01439	14.4	0.02018	25.58
HLEsup	0.01105	12.15	-	-
AGEaut	0.01354	6.37	0.01411	8.72
AGE2aut	-0.00006	-2.58	-0.00007	-3.85
AGEchoi	0.06420	30.13	0.06034	37.6
AGE2choi	-0.00060	-25.67	-0.00058	-32.01
AGEcom	*-0.00317	-1.51	-0.00590	-3.71
AGE2com	0.00006	2.47	0.00009	4.78
AGEconf	-0.01260	-5.93	-0.01178	-7.4
AGE2conf	0.00019	8.34	0.00017	9.63
AGEdign	-0.02667	-12.44	-0.02645	-16.56
AGE2dign	0.00026	11.19	0.00026	14.79
AGEprom	-0.02641	-12.37	-0.02475	-15.47
AGE2prom	0.00027	11.58	0.00027	14.98
AGEqual	-0.01958	-9.12	-0.01149	-6.92
AGE2qual	0.00022	9.41	0.00015	7.93
AGEsup	0.02622	11.99	-	-
AGE2sup	-0.00021	-8.95	-	-
SEXaut	*0.01646	1.03	0.04005	3.47
SEXchoi	0.09978	6.24	0.08026	6.96
SEXcom	-0.06850	-4.34	*-0.00890	-0.77
SEXconf	-0.09273	-5.85	*-0.01784	-1.55
SEXdign	-0.08606	-5.43	-0.03688	-3.22
SEXprom	-0.10643	-6.56	0.00287	0.25
SEXqual	-0.05462	-3.45	*0.02439	2.12
SEXsup	-0.06930	-4.34	-	-
AUT	-2.07458	-37.82	-3.09837	-58.3
CHOI	-2.68961	-43.29	-3.07136	-55.31
COM	-0.55033	-8.81	-1.01909	-15.51

CONF	-1.03778	-17.96	-1.54928	-27.18
DIGN	0.15164	2.06	0.50097	8.66
PROM	0.42535	6.1	*0.00318	0.08
QUAL	-0.34198	-5.37	-1.29325	-20.42
SUP	-1.34437	-21	-	-
constant	1.66782	57.59	1.71512	75.83

* not found significant at 5% confidence level

Table 6a: Universal weights

	Inpatient care	*Outpatient care*
Autonomy	0.1120	0.1259
Choice	0.1105	0.1232
Communication	0.1339	0.1528
Confidentiality	0.1289	0.1484
Dignity	0.1355	0.1532
Prompt attention	0.1281	0.1452
Quality of basic amenities	0.1333	0.1512
Support	0.1180	-

Table 6b: Clustered weights

	Cluster 1		**Cluster 2**		**Cluster 3**	
	inpatient	*outpatient*	*inpatient*	*outpatient*	*inpatient*	*outpatient*
Autonomy	0.1208	0.1369	0.1207	0.1370	0.1234	0.1415
Choice	0.1189	0.1339	0.1210	0.1380	0.1237	0.1413
Communication	0.1263	0.1443	0.1281	0.1463	0.1273	0.1455
Confidentiality	0.1269	0.1449	0.1254	0.1436	0.1257	0.1444
Dignity	0.1292	0.1473	0.1283	0.1464	0.1256	0.1430
Prompt attention	0.1275	0.1456	0.1259	0.1433	0.1241	0.1417
Quality of basic amenities	0.1284	0.1470	0.1279	0.1455	0.1252	0.1426
Support	0.1219	-	0.1227	-	0.1249	-

Table 6c: Country-level weights for outpatient care

Country	*Aut*	*Choi*	*Com*	*Conf*	*Dign*	*Prom*	*Qual*
Bangladesh	0.1327	0.1319	0.1385	0.1512	0.1546	0.1435	0.1476
Bosnia Herzegovina	0.1360	0.1366	0.1449	0.1471	0.1491	0.1463	0.1401
Brazil	0.1359	0.1314	0.1527	0.1358	0.1499	0.1419	0.1524
Burkina Faso	0.1147	0.0927	0.1582	0.1497	0.1628	0.1615	0.1604
Chad	0.1342	0.1061	0.1432	0.1416	0.1624	0.1524	0.1602
China	0.1006	0.1352	0.1478	0.1462	0.1692	0.1468	0.1543
Comoros	0.0863	0.1057	0.1566	0.1712	0.1773	0.1412	0.1618
Congo	0.1164	0.0985	0.1420	0.1685	0.1410	0.1513	0.1824
Cote d'Ivoire	0.0939	0.0597	0.1656	0.1603	0.1783	0.1616	0.1807
Croatia	0.1298	0.1396	0.1652	0.1312	0.1459	0.1415	0.1470
Czech Republic	0.1171	0.1216	0.1674	0.1568	0.1560	0.1373	0.1439
Dominican Republic	0.1391	0.1415	0.1464	0.1393	0.1446	0.1409	0.1481
Ecuador	0.1278	0.1316	0.1518	0.1491	0.1512	0.1405	0.1479
Estonia	0.1083	0.1315	0.1679	0.1500	0.1587	0.1474	0.1362
Finland	0.1524	0.1365	0.1528	0.1502	0.1338	0.1351	0.1391
France	0.1519	0.1590	0.1459	0.1429	0.1309	0.1316	0.1379
Georgia	0.1027	0.1459	0.1569	0.1251	0.1628	0.1476	0.1589
Germany	0.1402	0.1511	0.1470	0.1471	0.1391	0.1350	0.1405
Ghana	0.0922	0.0893	0.1393	0.1673	0.1653	0.1680	0.1786
Greece	0.1328	0.1409	0.1477	0.1482	0.1463	0.1400	0.1442
Hungary	0.1144	0.1152	0.1647	0.1472	0.1501	0.1489	0.1595
India	0.1310	0.1363	0.1360	0.1403	0.1565	0.1442	0.1557
Israel	0.1376	0.1358	0.1495	0.1505	0.1440	0.1389	0.1437
Italy	0.1353	0.1458	0.1495	0.1435	0.1474	0.1400	0.1383
Kazakhstan	0.1258	0.1356	0.1429	0.1539	0.1578	0.1436	0.1403
Kenya	0.0971	0.0665	0.1505	0.1639	0.1766	0.1710	0.1744
Lao	0.1283	0.1067	0.1458	0.1381	0.1557	0.1632	0.1621
Latvia	0.1259	0.1268	0.1733	0.1404	0.1565	0.1486	0.1286
Malawi	0.0935	0.0621	0.1565	0.1657	0.1756	0.1732	0.1734
Malaysia	0.1314	0.0924	0.1650	0.1437	0.1525	0.1465	0.1684
Mauritania	0.1367	0.1399	0.1389	0.1489	0.1495	0.1423	0.1438
Mauritius	0.1215	0.0755	0.1800	0.1612	0.1688	0.1219	0.1711

Mexico	0.1399	0.1330	0.1490	0.1472	0.1457	0.1404	0.1448
Morocco	0.1059	0.1114	0.1670	0.1428	0.1651	0.1441	0.1638
Myanmar	0.1233	0.1283	0.1495	0.1449	0.1567	0.1413	0.1561
Nepal	0.1326	0.1345	0.1577	0.1062	0.1530	0.1547	0.1613
Netherlands	0.1523	0.1419	0.1492	0.1444	0.1369	0.1359	0.1394
Norway	0.1401	0.1381	0.1519	0.1518	0.1372	0.1392	0.1417
Pakistan	0.1047	0.1226	0.1278	0.1531	0.1762	0.1660	0.1494
Paraguay	0.1359	0.1327	0.1509	0.1386	0.1460	0.1444	0.1514
Philippines	0.1240	0.1143	0.1675	0.1358	0.1549	0.1454	0.1580
Portugal	0.1330	0.1335	0.1451	0.1466	0.1536	0.1452	0.1431
Russia	0.1031	0.1028	0.1635	0.1547	0.1688	0.1492	0.1579
Slovakia	0.1359	0.1329	0.1615	0.1564	0.1429	0.1345	0.1358
Slovenia	0.1192	0.1108	0.1641	0.1544	0.1552	0.1515	0.1448
South Africa	0.1337	0.1201	0.1537	0.1537	0.1485	0.1405	0.1498
Spain	0.1194	0.1194	0.1563	0.1503	0.1524	0.1449	0.1496
Sri Lanka	0.1206	0.1372	0.1464	0.1466	0.1512	0.1449	0.1531
Sweden	0.1495	0.1313	0.1492	0.1537	0.1386	0.1358	0.1419
Tunisia	0.1126	0.1175	0.1527	0.1600	0.1613	0.1428	0.1533
United Arab Emirates	0.1158	0.1234	0.1518	0.1594	0.1608	0.1403	0.1485
United Kingdom	0.1511	0.1434	0.1464	0.1478	0.1381	0.1346	0.1386
Ukraine	0.1237	0.1328	0.1468	0.1474	0.1522	0.1430	0.1542
Uruguay	0.1169	0.1305	0.1632	0.1510	0.1517	0.1384	0.1484
Vietnam	0.0956	0.1144	0.1503	0.1346	0.1725	0.1630	0.1695
Zimbabwe	0.1252	0.0413	0.1666	0.1692	0.1696	0.1648	0.1633

Table 6d: Country-level weights for inpatient care

Country	*Aut*	*Choi*	*Com*	*Conf*	*Dign*	*Prom*	*Qual*	*Sup*
Bangladesh	0.1210	0.1201	0.1259	0.1350	0.1358	0.1255	0.1343	0.1024
Bosnia Herzegovina	0.1211	0.1199	0.1266	0.1274	0.1291	0.1281	0.1242	0.1235
Brazil	0.1161	0.1159	0.1349	0.1196	0.1323	0.1259	0.1359	0.1194
Burkina Faso	0.1048	0.0909	0.1408	0.1306	0.1415	0.1365	0.1415	0.1134
Chad	0.1095	0.0930	0.1198	0.1391	0.1601	0.1437	0.1470	0.0877
China	0.0943	0.1267	0.1353	0.1296	0.1500	0.1343	0.1385	0.0912
Comoros	0.0958	0.1016	0.1394	0.1510	0.1544	0.1169	0.1428	0.0981
Congo	0.1017	0.1070	0.1414	0.1300	0.1179	0.1332	0.1370	0.1318
Cote d'Ivoire	0.0957	0.0890	0.1377	0.1400	0.1481	0.1355	0.1491	0.1050
Croatia	0.1143	0.1197	0.1416	0.1175	0.1284	0.1304	0.1289	0.1191
Czech Republic	0.1049	0.1059	0.1464	0.1347	0.1427	0.1102	0.1256	0.1296
Dominican Republic	0.1225	0.1232	0.1278	0.1230	0.1264	0.1242	0.1287	0.1242
Ecuador	0.1139	0.1158	0.1346	0.1293	0.1310	0.1242	0.1303	0.1209
Estonia	0.1033	0.1160	0.1477	0.1153	0.1415	0.1310	0.1272	0.1180
Finland	0.1271	0.1234	0.1382	0.1300	0.1297	0.1267	0.1304	0.0944
France	0.1246	0.1375	0.1322	0.1323	0.1281	0.1259	0.1278	0.0915
Georgia	0.0996	0.1279	0.1312	0.1143	0.1385	0.1296	0.1390	0.1199
Germany	0.1319	0.1238	0.1348	0.1310	0.1304	0.1258	0.1289	0.0935
Ghana	0.0942	0.0868	0.1218	0.1422	0.1450	0.1505	0.1550	0.1045
Greece	0.1156	0.1305	0.1314	0.1345	0.1314	0.1230	0.1289	0.1048
Hungary	0.1035	0.1045	0.1379	0.1299	0.1325	0.1283	0.1352	0.1281
India	0.1147	0.1249	0.1172	0.1265	0.1411	0.1317	0.1376	0.1062
Israel	0.1252	0.1191	0.1362	0.1297	0.1313	0.1266	0.1303	0.1017
Italy	0.1190	0.1344	0.1344	0.1268	0.1333	0.1268	0.1295	0.0958
Kazakhstan	0.1132	0.1217	0.1298	0.1346	0.1406	0.1290	0.1228	0.1084
Kenya	0.0879	0.0600	0.1342	0.1454	0.1574	0.1503	0.1528	0.1119
Lao	0.1153	0.1017	0.1289	0.1201	0.1375	0.1346	0.1373	0.1247
Latvia	0.1143	0.1184	0.1475	0.1073	0.1301	0.1326	0.1272	0.1226
Malawi	0.0834	0.0608	0.1328	0.1437	0.1529	0.1515	0.1505	0.1244
Malaysia	0.1187	0.0798	0.1463	0.1289	0.1341	0.1294	0.1477	0.1150
Mauritania	0.1210	0.1218	0.1271	0.1299	0.1327	0.1229	0.1289	0.1157
Mauritius	0.1076	0.0667	0.1633	0.1461	0.1545	0.1121	0.1490	0.1008

Mexico	0.1230	0.1161	0.1308	0.1281	0.1272	0.1231	0.1271	0.1246
Morocco	0.0866	0.0904	0.1454	0.1287	0.1487	0.1345	0.1427	0.1230
Myanmar	0.1147	0.1158	0.1282	0.1296	0.1383	0.1306	0.1346	0.1082
Nepal	0.1190	0.1208	0.1330	0.1039	0.1343	0.1338	0.1382	0.1171
Netherlands	0.1286	0.1348	0.1343	0.1321	0.1289	0.1244	0.1270	0.0901
Norway	0.1344	0.1213	0.1393	0.1346	0.1294	0.1263	0.1298	0.0849
Pakistan	0.0940	0.1115	0.1127	0.1358	0.1640	0.1524	0.1350	0.0945
Paraguay	0.1206	0.1162	0.1309	0.1225	0.1297	0.1263	0.1314	0.1224
Philippines	0.1109	0.1065	0.1427	0.1166	0.1315	0.1286	0.1362	0.1271
Portugal	0.1256	0.1261	0.1310	0.1302	0.1300	0.1250	0.1286	0.1034
Russia	0.1011	0.0981	0.1437	0.1232	0.1477	0.1339	0.1377	0.1147
Slovakia	0.1148	0.1218	0.1413	0.1302	0.1193	0.1209	0.1270	0.1247
Slovenia	0.1120	0.1046	0.1439	0.1319	0.1297	0.1289	0.1277	0.1213
South Africa	0.1220	0.1132	0.1303	0.1328	0.1315	0.1202	0.1293	0.1206
Spain	0.1063	0.1133	0.1368	0.1296	0.1351	0.1260	0.1315	0.1214
Sri Lanka	0.1093	0.1232	0.1271	0.1302	0.1357	0.1305	0.1385	0.1054
Sweden	0.1309	0.1230	0.1353	0.1314	0.1292	0.1262	0.1294	0.0946
Tunisia	0.0972	0.1048	0.1381	0.1420	0.1437	0.1273	0.1361	0.1109
United Arab Emirates	0.1103	0.1037	0.1308	0.1422	0.1402	0.1205	0.1364	0.1159
United Kingdom	0.1326	0.1326	0.1334	0.1309	0.1281	0.1237	0.1261	0.0927
Ukraine	0.1114	0.1170	0.1309	0.1257	0.1354	0.1235	0.1357	0.1205
Uruguay	0.1077	0.1159	0.1456	0.1331	0.1340	0.1202	0.1298	0.1137
Vietnam	0.0850	0.0987	0.1407	0.1160	0.1603	0.1530	0.1606	0.0857
Zimbabwe	0.1156	0.0525	0.1403	0.1403	0.1436	0.1409	0.1383	0.1286

Table 7a, 7b: Final responsiveness scores

	Universal weights			Country-level weights	
	Country	*Score*		*Country*	*Score*
1	Finland	0.7396	1	Finland	0.7343
2	Paraguay	0.7114	2	France	0.7139
3	France	0.7112	3	Paraguay	0.7131
4	Sweden	0.6960	4	Sweden	0.6950
5	United Kingdom	0.6788	5	United Kingdom	0.6796
6	Norway	0.6738	6	Norway	0.6634
7	Netherlands	0.6576	7	Dominican Republic	0.6605
8	Dominican Republic	0.6567	8	Netherlands	0.6575
9	Germany	0.6551	9	Germany	0.6541
10	Georgia	0.6376	10	Georgia	0.6428
11	Greece	0.6374	11	Greece	0.6397
12	Myanmar	0.6370	12	Myanmar	0.6384
13	Uruguay	0.6330	13	Uruguay	0.6352
14	United Arab Emirates	0.6219	14	United Arab Emirates	0.6223
15	Israel	0.6204	15	Israel	0.6217
16	Czech Republic	0.6151	16	Czech Republic	0.6148
17	Spain	0.6105	17	Spain	0.6102
18	Mexico	0.6040	18	Mexico	0.6055
19	Slovenia	0.5783	19	Slovenia	0.5774
20	Brazil	0.5690	20	Malaysia	0.5722
21	Malaysia	0.5688	21	Brazil	0.5702
22	Hungary	0.5550	22	Hungary	0.5574
23	Bosnia Herzegovina	0.5536	23	Croatia	0.5562
24	Estonia	0.5524	24	China	0.5556
25	Croatia	0.5522	25	Bosnia Herzegovina	0.5544
26	Ecuador	0.5522	26	Estonia	0.5529
27	China	0.5505	27	Ecuador	0.5523
28	Latvia	0.5490	28	Latvia	0.5487
29	Philippines	0.5425	29	Philippines	0.5435
30	Portugal	0.5399	30	Portugal	0.5429
31	Italy	0.5256	31	Italy	0.5299
32	Vietnam	0.4976	32	Vietnam	0.5069
33	India	0.4867	33	India	0.4925
34	Kazakhstan	0.4834	34	Kazakhstan	0.4836
35	Lao	0.4793	35	Lao	0.4819
36	South Africa	0.4715	36	Bangladesh	0.4736
37	Bangladesh	0.4713	37	South Africa	0.4717
38	Comoros	0.4474	38	Mauritius	0.4456
39	Tunisia	0.4387	39	Comoros	0.4442
40	Nepal	0.4300	40	Tunisia	0.4400
41	Burkina Faso	0.4279	41	Nepal	0.4350
42	Slovakia	0.4250	42	Burkina Faso	0.4296
43	Mauritius	0.4210	43	Slovakia	0.4287
44	Congo	0.4138	44	Pakistan	0.4160
45	Pakistan	0.4135	45	Congo	0.4127
46	Zimbabwe	0.4069	46	Zimbabwe	0.4075
47	Ukraine	0.3987	47	Ghana	0.4050
48	Ghana	0.3956	48	Ukraine	0.3994
49	Russia	0.3943	49	Russia	0.3943
50	Kenya	0.3779	50	Malawi	0.3810
51	Sri Lanka	0.3764	51	Kenya	0.3793
52	Cote d'Ivoire	0.3743	52	Sri Lanka	0.3793
53	Malawi	0.3739	53	Cote d'Ivoire	0.3779
54	Mauritania	0.3731	54	Mauritania	0.3746
55	Morocco	0.2882	55	Morocco	0.2873
56	Chad	0.2266	56	Chad	0.2246

Table 7c: Final responsiveness scores based on clustered weights

	Cluster 1	
	Country	*Score*
1	Myanmar	0.6391
2	Philippines	0.5432
3	India	0.4913
4	Kazakhstan	0.4841
5	Lao	0.4835
6	Bangladesh	0.4742
7	South Africa	0.4730
8	Comoros	0.4480
9	Nepal	0.4316
10	Burkina Faso	0.4275
11	Pakistan	0.4154
12	Congo	0.4149
13	Zimbabwe	0.4054
14	Ghana	0.3941
15	Kenya	0.3773
16	Mauritania	0.3747
17	Cote d'Ivoire	0.3736
18	Malawi	0.3723
19	Chad	0.2310

	Cluster 2	
	Country	*Score*
1	Paraguay	0.7128
2	Dominican Republic	0.6592
3	Georgia	0.6382
4	United Arab Emirates	0.6210
5	Malaysia	0.5680
6	Brazil	0.5672
7	Hungary	0.5567
8	Bosnia Herzegovina	0.5551
9	China	0.5534
10	Ecuador	0.5531
11	Estonia	0.5511
12	Latvia	0.5479
13	Vietnam	0.4983
14	Tunisia	0.4387
15	Mauritius	0.4141
16	Ukraine	0.3983
17	Russia	0.3943
18	Sri Lanka	0.3812
19	Morocco	0.2900

	Cluster 3	
	Country	Score
1	Finland	0.7349
2	France	0.7127
3	Sweden	0.6931
4	United Kingdom	0.6805
5	Norway	0.6652
6	Netherlands	0.6567
7	Germany	0.6560
8	Greece	0.6379
9	Uruguay	0.6304
10	Israel	0.6203
11	Czech Republic	0.6182
12	Spain	0.6077
13	Mexico	0.6062
14	Slovenia	0.5792
15	Croatia	0.5537
16	Portugal	0.5422
17	Italy	0.5272
18	Slovakia	0.4295

Table 8: Responsiveness scores based on universal weights using ordered probit model and HOPIT model

	Ordered Probit		*HOPIT*	
Country	*Rank*	*Score*	*Rank*	*Score*
Greece	1	0,728	11	0,637
Paraguay	2	0,718	2	0,711
Israel	3	0,700	15	0,620
Czech Republic	4	0,686	16	0,615
United Arab Emirates	5	0,685	14	0,622
France	6	0,677	3	0,711
United Kingdom	7	0,675	5	0,679
Sweden	8	0,668	4	0,696
Norway	9	0,653	6	0,674
Germany	10	0,615	9	0,655
Hungary	11	0,614	22	0,555
Finland	12	0,613	1	0,740
Bosnia Herzegovina	13	0,589	23	0,554
Uruguay	14	0,585	13	0,633
Netherlands	15	0,575	7	0,658
Slovenia	16	0,554	19	0,578
Brazil	17	0,553	20	0,569
Spain	18	0,548	17	0,611
Malaysia	19	0,546	21	0,569
Estonia	20	0,529	24	0,552
Portugal	21	0,522	30	0,540
Mexico	22	0,522	18	0,604
Ecuador	23	0,520	26	0,552
Latvia	24	0,517	28	0,549
Georgia	25	0,512	10	0,638
Dominican Republic	26	0,506	8	0,657
South Africa	27	0,499	36	0,471
Slovakia	28	0,489	42	0,425
Croatia	29	0,482	25	0,552
India	30	0,477	33	0,487
Malawi	31	0,476	53	0,374
Myanmar	32	0,471	12	0,637
Italy	33	0,468	31	0,526
Ghana	34	0,467	48	0,396
Zimbabwe	35	0,455	46	0,407
Lao	36	0,446	35	0,479
Kazakhstan	37	0,425	34	0,483
Kenya	38	0,419	50	0,378
China	39	0,418	27	0,550
Tunisia	40	0,413	39	0,439
Bangladesh	41	0,411	37	0,471
Burkina Faso	42	0,383	41	0,428
Sri Lanka	43	0,368	51	0,376
Mauritius	44	0,361	43	0,421
Philippines	45	0,358	29	0,543
Nepal	46	0,357	40	0,430
Vietnam	47	0,355	32	0,498
Cote d'Ivoire	48	0,350	52	0,374

Morocco	49	0,337	55	0,288
Congo	50	0,331	44	0,414
Russia	51	0,323	49	0,394
Mauritania	52	0,311	54	0,373
Comoros	53	0,289	38	0,447
Ukraine	54	0,289	47	0,399
Chad	55	0,278	56	0,227
Pakistan	56	0,265	45	0,413

Figure 3: Mean ratings of vignettes and self-report of autonomy for inpatient care

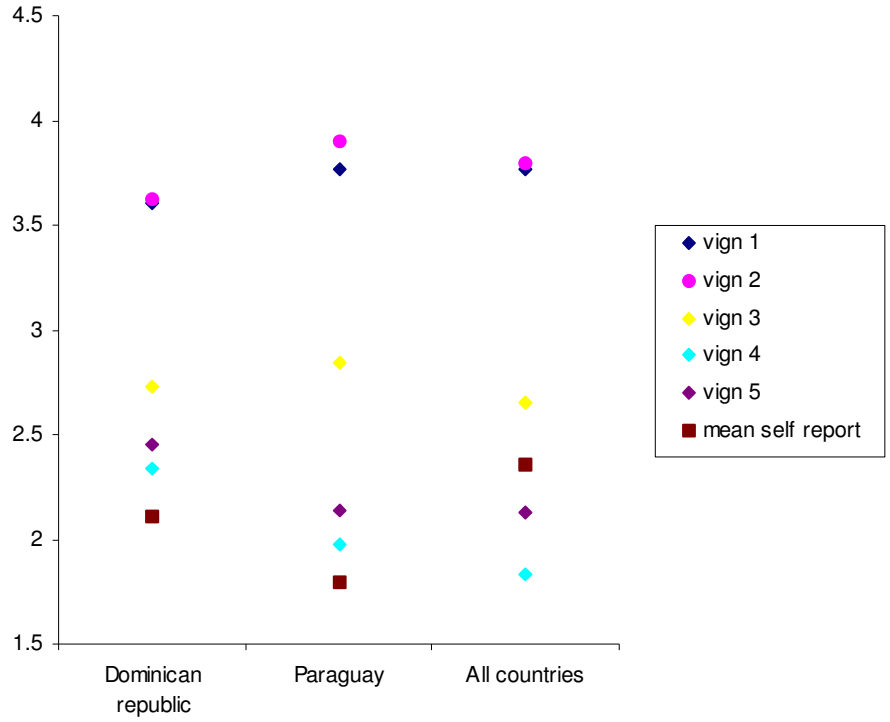

Table 9: Responsiveness scores based on universal weights including and excluding non-users

	including non-users		*excluding non-users*	
Country	*Rank*	*Score*	*Rank*	*Score*
Finland	1	0.7283	1	0.7396
France	2	0.7039	3	0.7112
United Kingdom	3	0.6682	5	0.6788
Netherlands	4	0.6581	7	0.6576
Sweden	5	0.6541	4	0.6960
Dominican Republic	6	0.6525	8	0.6567
Germany	7	0.6511	9	0.6551
Norway	8	0.6484	6	0.6738
Uruguay	9	0.6275	13	0.6330
Greece	10	0.6272	11	0.6374
Paraguay	11	0.6255	2	0.7114
Czech Republic	12	0.6134	16	0.6151
United Arab Emirates	13	0.6120	14	0.6219
Myanmar	14	0.6088	12	0.6370
Spain	15	0.6083	17	0.6105
Israel	16	0.6069	15	0.6204
Georgia	17	0.6006	10	0.6376
Slovenia	18	0.5763	19	0.5783
Mexico	19	0.5727	18	0.6040
Malaysia	20	0.5550	21	0.5688
Brazil	21	0.5524	20	0.5690
Bosnia Herzegovina	22	0.5511	23	0.5536
Croatia	23	0.5490	25	0.5522
Hungary	24	0.5437	22	0.5550
China	25	0.5418	27	0.5505
Latvia	26	0.5339	28	0.5490
Ecuador	27	0.5322	26	0.5522
Portugal	28	0.5301	30	0.5399
Estonia	29	0.5300	24	0.5524
Italy	30	0.5230	31	0.5256
Philippines	31	0.5082	29	0.5425
Vietnam	32	0.4844	32	0.4976
Kazakhstan	33	0.4784	34	0.4834
India	34	0.4750	33	0.4867
Lao	35	0.4576	35	0.4793
Bangladesh	36	0.4571	37	0.4713
South Africa	37	0.4541	36	0.4715
Tunisia	38	0.4299	39	0.4387
Slovakia	39	0.4244	42	0.4250
Mauritius	40	0.4190	43	0.4210
Comoros	41	0.4169	38	0.4474
Pakistan	42	0.4012	45	0.4135
Zimbabwe	43	0.3918	46	0.4069
Congo	44	0.3839	44	0.4138
Burkina Faso	45	0.3803	41	0.4279
Ghana	46	0.3758	48	0.3956

Nepal	47	0.3755	40	0.4300
Russia	48	0.3733	49	0.3943
Ukraine	49	0.3728	47	0.3987
Sri Lanka	50	0.3710	51	0.3764
Kenya	51	0.3577	50	0.3779
Cote d'Ivoire	52	0.3505	52	0.3743
Malawi	53	0.3435	53	0.3739
Mauritania	54	0.3347	54	0.3731
Morocco	55	0.2380	55	0.2882
Chad	56	0.1929	56	0.2266

Appendix 1: Responsiveness questions

The following questions addressed the eight domains of responsiveness:

- 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 the experience of how clearly health care providers explained things to you?
- Confidentiality: How would you rate the way the health services ensured you could talk privately to health care providers?
- Dignity: How would you rate your experience of being greeted and talked to respectfully?
- 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 community support: How would you rate the ease of having family and friends visit you? (only applies to inpatient care).

The complete household questionnaire and the appendix with the vignettes can be downloaded from following website:
http://www.who.int/healthinfo/survey/instruments/en/index.html

Appendix 2: Inclusion of non-users

In order to assess the consequences of the exclusion of non-users to the final scores an analysis including the non-users is performed. In the WHS the following question was included: The last time you [your child] needed health care, did you get health care? This question is used to identify the population excluded from health care. The WHO advised to attach a value of zero to the performance of all domains for the non-users (Valentine, 2003c). Unfortunately, it is not clear whether the population is excluded from outpatient- or inpatient care. Assuming that patients mostly consult an outpatient specialist before being admitted to inpatient care, the non-users are included in the outpatient care sample. After including non-users, the analysis is performed identical to the methods described in the paper. The results of the analysis using universal weights are displayed in table 9.

Table 9 around here

The inclusion of the non-users into the analysis has a large influence on the ranking. This can be explained by the extreme value attached to the responsiveness of non-users. The value of zero is very low and pulls the responsiveness score down.

The analysis is far from perfect since important assumptions are made. First, it is assumed all non-users were demanding outpatient care. Second, the value of zero is attached to all domains of responsiveness, which is arbitrary and pulls down the responsiveness score. For example, it does not seem reasonable to attach a value of zero to the domain of autonomy, because of exclusion of health care. Because of these large assumptions non-users are not included in the main analysis and these results can only be used as an indication of the effect.