

# Analysis of the validity of the vignette approach to correct for heterogeneity in reporting health system responsiveness

Nigel Rice · Silvana Robone · Peter Smith

Received: 21 September 2009 / Accepted: 4 March 2010 / Published online: 28 March 2010  
© Springer-Verlag 2010

**Abstract** Despite the growing popularity of the vignette methodology to deal with self-reported, categorical data, the formal evaluation of the validity of this methodology is still a topic of research. Some critical assumptions need to hold in order for this method to be valid. In this paper we analyse the assumption of “vignette equivalence” using data on health system responsiveness contained within the World Health Survey. We perform several tests to check the assumption of vignette equivalence. First, we use a test based on the global ordering of the vignettes. A minimal condition for the assumption of vignette equivalence to hold is that individual responses are consistent with the global ordering of vignettes. Secondly, using the hierarchical ordered probit model (HOPIT) model on the pool of countries, we undertake sensitivity analyses, stratifying countries according to the Inglehart–Welzel scale and the Human Development Index. The results of this analysis are robust, suggesting that the vignette equivalence assumption is not contradicted. Thirdly, we model the reporting behaviour of the respondents through a two-step regression procedure to evaluate whether the vignettes construct is perceived by respondents in different ways. Overall, across the analyses the results do not contradict the assumption of vignette equivalence and accordingly lend support to the use of the vignette methodology when analysing self-reported data and health system responsiveness.

**Keywords** Health system responsiveness · Anchoring vignettes · Vignette equivalence · Hierarchical ordered probit model

**JEL Classification** I10

## Introduction

In recent years the concept of responsiveness has been promoted as a desirable measure to evaluate the performance of health systems. Responsiveness relates to a system’s ability to respond to the legitimate expectations of potential users about non-health enhancing aspects of care [1]. In broad terms, it can be defined as the way in which individuals are treated and the environment in which they are treated, and encompasses the notion of an individual’s experience of contact with the health system [2].

One of the most ambitious attempts to implement a cross-country comparative instrument aimed at measuring health system performance is the World Health Survey (WHS), which includes modules on the responsiveness of a system to user preferences. Respondents are asked to rate their experiences of health systems using a 5-point categorical scale (ranging from “very good” to “very bad”). A common problem with such data is that individuals, when faced with the instrument, are likely to interpret the meaning of the response categories in a way that differs systematically across populations or population sub-groups according to their preferences and norms (for example, see Salomon et al. [3]). Accordingly, the response categories will not be comparable across populations if they do not correspond to the same underlying level of the responsiveness construct. We refer to this phenomenon as “reporting heterogeneity”.

N. Rice · S. Robone (✉) · P. Smith  
Centre for Health Economics, University of York,  
Alcuin Block A, Heslington, York YO10 5DD, UK  
e-mail: sr517@york.ac.uk

P. Smith  
Business School, Imperial College London,  
South Kensington Campus, London SW7 2AZ, UK

Recently, the use of anchoring vignettes has been promoted as a means for controlling for reporting heterogeneity across populations or population sub-groups. Vignettes represent hypothetical descriptions of a fixed level of a latent construct, such as responsiveness. Since these are fixed and predetermined, systematic variation across individuals in the rating of the vignettes can be attributed to differences in reporting behaviour [4]. The idea is to use information from the vignettes to adjust self-reported experiences of health system performance to increase cross population comparability by removing the influence of reporting heterogeneity.

In recent years anchoring vignettes have been utilised to address the issue of heterogeneous reporting behaviour in many studies regarding, for example, health and health-related behaviours (i.e. [3–7]), health system responsiveness [2, 8–10], happiness and job satisfaction [11, 12], national identity [13] and state effectiveness [14]. Despite the growing popularity of the vignette methodology to address the issue of reporting heterogeneity, the formal evaluation of the validity of the approach remains a topic of research [15–18]. Two critical assumptions need to hold in order for the method to be valid. The first, termed response consistency, implies that individuals classify the vignettes in a way that is consistent with the rating of their own experiences of health system responsiveness. This implies that the mapping used from the latent levels of responsiveness given by the vignettes to the response categories is the same as the mapping used to translate latent responsiveness of own experiences of contact with health services to the available response categories. The second assumption, termed vignette equivalence, implies that “the level of the variable represented by any one vignette is perceived by all respondents in the same way and on the same unidimensional scale” [7, p. 194]. This assumption implies that, conditional on the socio-economic characteristics that determine reporting behaviour, for each vignette there is an actual (unobserved) level of responsiveness that all individuals agree to, irrespective of their country of residence, their socio-demographic characteristics or the level of responsiveness they actually face.

In this paper, we focus attention on the assumption of vignette equivalence.<sup>1</sup> A limited number of other studies have tried to assess the validity of this assumption. These were focussed on self-reports of the ratings of work disability [5], mobility [19], visual acuity and political efficacy [7, 21], job satisfaction [11] and life satisfaction for income [21], largely making use of non-parametric methods using tests based on the global ordering of the vignettes. Our study explores the validity of the vignette equivalence

assumption, making reference to the concept of responsiveness and using data from the WHS. Moreover, we adopt several strategies to assess the validity of the vignette equivalence assumption, using both non-parametric and parametric methods. The use of a two-step regression procedure to evaluate whether a vignette construct is perceived in the same way across respondents is novel in this context.

## Data and methods

### Data

To assess the validity of the vignette equivalence assumption we use data from the 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 Ustun et al. [22]). The basic survey mode was an in-person interview, consisting of either a 90-min in-household interview (53 countries), a 30-min face-to-face interview (13 countries) or a computer-assisted telephone interview (4 countries). In total, 70 countries participated in the WHS 2002–2003. All surveys were drawn from nationally representative frames with known probability resulting in sample sizes of between 600 and 10,000 respondents across the countries surveyed. Data collection was on a modular basis covering different aspects of health and health systems, including information on health state valuation, health system responsiveness and health system goals. Samples have undergone extensive quality assurance procedures, including the testing of the psychometric properties of the responsiveness instrument [23], and close attention has been paid to the issue of comparability [22].

The WHS responsiveness module gathers basic information on health care utilisation for both inpatient and outpatient services. In the analysis that follows we make reference only to inpatient services. The measurement of responsiveness was obtained by asking respondents to rate their most recent experience of contact with the health system within a set of eight domains by responding to set questions. The domains consist of “autonomy” (involved in decisions), “choice” (of health care provider), “clarity of communication” (of health care personnel), “confidentiality” (e.g. talk privately), “dignity” (respectful treatment and communication), “prompt attention” (e.g. waiting times), “quality of basic facilities” and “access to family and community support”.<sup>2</sup> The following five response

<sup>1</sup> Other studies focus on the assumption of response consistency when trying to assess the validity of the anchoring vignettes methodology, i.e. [6, 15, 17].

<sup>2</sup> The long-form questionnaire uses two questions items per domain, while the short-form questionnaire uses only one. We use the eight items that are common to the long- and short-form questionnaire.

categories were available to respondents when rating their experience of health systems: “very good”, “good”, “moderate”, “bad”, and “very bad”.

The WHS further contains information on respondent characteristics. We make use of age, gender, level of education and income. These variables have been extensively used in the studies investigating differential reporting behaviour in self-reported measure of health [2, 4, 19] and health-related disabilities [5]. Level of education is a continuous variable measuring the number of years in education. Gender is a dummy variable coded 0 for women and 1 for men. Income is derived from a measure of permanent income based on information on the physical assets owned by households. The approach to its measurement, which relies on a variant of the hierarchical ordered probit model (HOPIT) to improve cross-country comparability, is provided by Ferguson et al. [24]. We construct dummy variables to indicate the tertiles of the within-country distribution of household permanent income to which individuals belong. For the analysis presented here, the first income tertile is considered as the base category.

The WHS contains a number of vignettes describing the experiences of hypothetical individuals within each of the eight domains of responsiveness. The vignettes have been divided into four sets (A–D) with each set containing five vignettes for each item present across two domains. For example, Set A contains five vignettes for each of the two items in the domain of “Dignity” and five vignettes for each of the two items in “Prompt Attention”. Due to 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. The response scale available to respondents answering the vignettes is the same as the scale available when reporting their own experiences of health system responsiveness. Examples of the WHS vignettes are provided in Table 1 for the domains “Confidentiality”, “Choice”, “Clarity of communication” and “Quality of basic amenities”.

We attempt to take into consideration the different levels of socio-economic development of countries to assess whether this influence the perception of the vignettes by making use of the Human Development Index (HDI) to stratify the countries into high, medium and low HDI groups. The HDI is a composite index of human development that combines indicators of life expectancy, educational attainment and income [25]. We also try to take into account the presence of different values and norms in different countries and evaluate if those values and norms affect the way individuals perceive the vignettes. To do this, we stratify our sample on the basis of the Inglehart–Welzel Cultural Map of the World, represented in Fig. 1

(<http://www.worldvaluessurvey.org>) [26].<sup>3</sup> This map reflects the presence of a strong correlation between a large number of basic values common to several countries. If we focus on European countries only, according to the Inglehart–Welzel map it is possible to identify three sets of countries that share similar social norms and values: Catholic countries, Protestant countries and ex-communist countries. At a broader level, if we consider all countries across the world, the basic values can be represented across two major dimensions of cross-cultural variation: Traditional/Secular-rational and Survival/Self-expression values (<http://www.worldvaluessurvey.org>). The first dimension reflects the contrast between societies in which religion is considered as an important element of life and those in which it is not. The second dimension reflects the contrast between industrial and post-industrial societies. In the former societies emphasis is given to economic and physical security while in the latter societies there is an increasing emphasis on subjective well-being, self-expression and quality of life. We follow this stratification in the analysis that follows.<sup>4</sup>

## Methods

### *Consistent and near-consistent ordering of vignettes*

We assess the vignette equivalence assumption by first considering the global ordering of the vignettes. A minimal condition for the assumption of vignette equivalence to hold is that individual responses are consistent with the global ordering of vignettes. The global ordering for a domain can be obtained by pooling all the responses across countries and considering the average categorical response for each vignette [19]. Similar tests of the vignette equivalence assumption based on the global ordering of vignettes, but for health-related disabilities, job satisfaction and self reported measures of health, have been undertaken by [5, 9, 11, 21]. Due to the presence of stochastic measurement errors we cannot expect all individuals to order the vignettes in exactly the same way as each other. Adopting

<sup>3</sup> This map has been utilised to assess the validity of the vignette equivalence assumption also by Kristensen and Johansson [11].

<sup>4</sup> “Self Secular” = Austria, Belgium, Denmark, Germany, Spain, Finland, France, Great Britain, Greece, Israel, Italy, Luxemburg, Netherlands, Slovenia, Sweden. “Self-Traditional” = Brazil, Dominican Republic, Ecuador, Guatemala, Ireland, Portugal, Uruguay. “Survival-Traditional” = United Arab Emirates, Burkina Faso, Bangladesh, Chad, Cote d’Ivoire, Congo, Comoros, Ethiopia, Ghana, India, Kenya, Lao, Sri Lanka, Malaysia, Mauritania, Malay, Morocco, Myanmar, Mauritius, Malawi, Namibia, Nepal, Pakistan, Philippines, Senegal, Swaziland, Tunisia, South Africa, Zambia, Zimbabwe. “Survival Secular” = Bosnia, China, Croatia, Czech Republic, Georgia, Hungary, Kazakhstan, Latvia, Russia, Slovakia, Ukraine, Vietnam.

**Table 1** Examples of vignettes for the domain of confidentiality, choice, communication and quality of basic facilitiesDomain: confidentiality, choice<sup>a</sup>

1. [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?

2. 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?

Domain: clarity of communication and quality of basic facilities

[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?

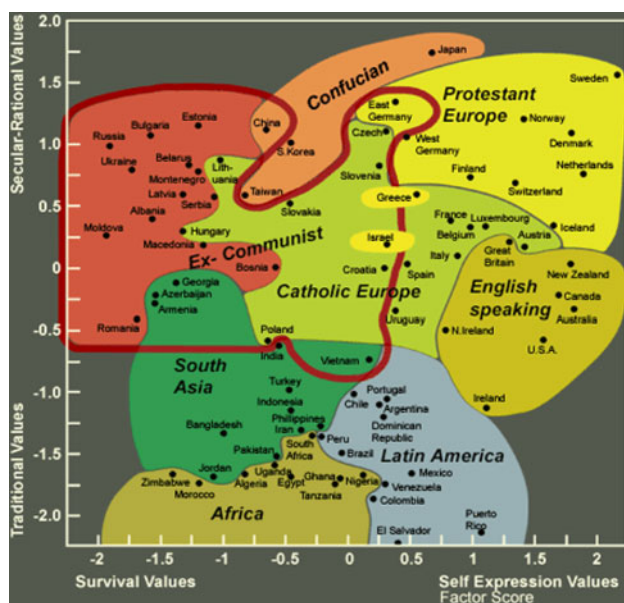
Q2: How would you rate the amount of space [Wing] had?

[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 or treatment?

<sup>a</sup> The vignettes above provide examples only and are 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"



**Fig. 1** Inglehart–Welzel cultural map of the world. Source: <http://www.worldvaluessurvey.org/>

the approach of Murray et al. [19], we define a consistent ordering as “a set of categorical vignette ratings that could be consistent with the global ordering in the latent variable space, if ambiguities were resolved in favour of the global ordering” [p. 373].<sup>5</sup> Accordingly, for each domain and for each country we compute the percentage of respondents

<sup>5</sup> For an example of consistent vignette ordering, consider Murray et al. [19], Fig. 30.3.

that gave an ordering of vignettes consistent with the global ordering, or had an ordering where only one vignette moved one or two ranks or two vignettes moved one rank each. Further, we compute the average percentage of respondents in each country that gave an ordering of vignettes consistent or near-consistent with the global ordering, where countries have been stratified by HDI groups and by the Inglehart–Welzel map groups.<sup>6</sup>

#### *Spearman rank order correlation coefficient*

Individuals' ordering of the vignettes might differ due either to measurement errors (caused, for example, by incorrect phrasing, translation or implementation of the vignette questions) or to problems of multidimensionality and variation in the cultural construct of a domain [19].<sup>7</sup> An analysis of the more common alternative patterns of vignette ordering can provide information about the relative importance of the problem of measurement error versus the problems of multidimensionality and variation in the cultural construct of a domain. Measurement error is generally associated with a large number of alternative orderings (due to chance). The prevalence of multidimensionality or

<sup>6</sup> The average is computed assigning the same weight to each country within a group.

<sup>7</sup> As an example “running a marathon” could be viewed as a multidimensional construct. Some individuals may view running a marathon as evidence of a high level of mobility and some as a result of exceptional talent. Others might consider it as an attribute related to health, whilst others might as an attribute related to sport [19].



cultural variation in a construct should, however, lead us to observe a limited number of alternative orderings, “reflecting some other weighting of the components of a multidimensional construct or alternative cultural constructs” [19, p. 376]. Multidimensionality of the responsiveness construct provides evidence of a violation of the vignette equivalence assumption. The Spearman rank order correlation coefficient (SROCC), which quantifies the extent to which an ordering is consistent with the global ordering of vignettes, has been suggested as a means to investigate the relative importance of the two sources of difference in ratings of vignettes [19].<sup>8</sup> For each domain we compute the SROCC between the vignettes rankings of each respondent and the global ranking.

We calculate the frequency distribution, together with several descriptive statistics, of the SROCCs across all individuals in the WHS dataset for the eight domains considered.<sup>9</sup> First, for each domain, we compute the percentage of individuals who report an ordering of vignettes that is positive and the percentage of individuals for which the correlation coefficient between the individual and the global ordering of vignettes is larger than 0.5. Secondly, following Murray et al. [19], we report the number of different rank order correlation coefficients observed in each domain and the number that occur with a frequency greater than 1%. The greater the number of different rank order correlation coefficients reported in each domain together with a smaller number occurring with a large frequency, the higher the probability that alternative orderings are due to measurement errors rather than to multidimensionality or cultural variation. We also show the median SROCC for each domain and the average SROCC across domains for each country.<sup>10</sup>

### The HOPIT model

An alternative way to check the vignette equivalence assumption implies estimating a model for responsiveness that takes into account possible biases due to reporting

<sup>8</sup> Perfect agreement of the rankings leads to a coefficient of 1, perfect disagreement  $-1$ , and independence 0.

<sup>9</sup> We do not include in the analysis individuals who gave the same evaluation of all the vignettes (i.e. they judge all the vignettes as excellent responsiveness). Indeed, for these individuals it is not possible to compute the Spearman rank order correlation coefficient between their ranking and the global ordering ranking. However, we perform a robustness check including in the sample the observations about respondents who gave the same evaluation of all the vignettes. Referring to the domain “Confidentiality”, we perform the robustness check by just moving one vignette of one rank, in a consistent way with the global ordering. The results obtained including these observations are extremely similar to those not including them.

<sup>10</sup> The average SROCCs have been computed assuming equal weight for each individual.

heterogeneity. This approach, adopted by Kristensen and Johansson [11] when considering self-reported job satisfaction, consists of firstly estimating a model on a pool of countries. Secondly, the sample is split into groups of countries according to the values, social norms, economic development, etc. that characterise these countries. Models are then estimated on the sub-samples and the coefficients are compared to those obtained from the pooled sample. If the model is robust and the vignette equivalence assumption is not violated, then we would expect the coefficient to be similar in the two samples. However, if the differences in culture and values across the country groups lead individuals to interpret the meaning of vignettes differently (and thus to violate the vignette equivalence assumption), we should observe very different estimated coefficients across the country groups [11].

Since the data on responsiveness in the WHS are self-reported and categorical, we use the HOPIT model developed by Tandon et al. [27] (also see Terza [28]), to adjust for reporting behaviour. The model can be specified in two parts. The first part draws on the use of the anchoring vignettes to provide a source of information that enables the thresholds to be modelled as functions of relevant covariates (*reporting behaviour equation*). The second part maps the relevant covariates to underlying self-reported health system responsiveness while controlling for differences in reporting behaviour obtained through the first step (*responsiveness equation*). A more formal description of the two parts of the model is reported in “Appendix” (also see Rice et al. [9]). The use of vignettes to identify reporting heterogeneity relies on the assumptions of response consistency and vignette equivalence described in the Introduction.

As a preliminary analysis, we apply the HOPIT model across the pool of 27 European countries present in the WHS, using the domain “Dignity”. For the purposes of our model, we use the dummies for country of residence together with individual specific characteristics (age, gender, level of education and income) as relevant covariates in both the reporting behaviour and the responsiveness equation. Austria is taken as the baseline country. We then stratify the European countries in three groups according to the Inglehart–Welzer map to reflect similar cultures, social norms and values. We finally re-estimate the HOPIT model for each of the three groups of countries.

We further extend the analysis by considering all the countries present in the WHS.<sup>11</sup> Mexico, which has the largest sample size, is taken as the baseline country. Countries are stratified into four groups according to the Inglehart–Welzer map (“Self-Traditional”, “Self-Secular”, “Survival-Traditional”, “Survival-Secular”) and the

<sup>11</sup> We exclude only Australia, Norway and Turkey since data on “Dignity” are not available for these countries.

HOPIT model is estimated separately for each of these groups of countries.

We also consider the possibility that differences in the level of socioeconomic development of countries might induce individuals to interpret the meaning of vignettes differently. Accordingly, we stratify the countries in the WHS according to their level of HDI and again apply the HOPIT model for each of these groups of countries.

#### *Assessment of multidimensionality of the constructs represented by vignettes*

An analysis of the characteristics of individuals described in the vignettes offers a further tool to check the vignette equivalence assumption. If the person described in a vignette is characterized by specific socio-demographic characteristics, it is possible that respondents are influenced by these characteristics, which may induce them to perceive the vignettes differently to other respondent. This would represent a violation of the vignette equivalence assumption. As an example, consider a vignette about “Autonomy” representing an elderly person. Some respondents may feel that elderly people are incapable of making appropriate decisions about treatments and may have lower expectations about the level of autonomy afforded to elderly individuals. Other respondents, however, could consider elderly people equally able to be involved in decisions about treatments as young people and hence would have the same expectations about the level of autonomy for elderly and young people. Specifying the age of the person described in the vignette may therefore induce some respondents to perceive the construct as representing “autonomy for elderly people” and for others to perceive it as “autonomy” in general.

Information on the characteristics of the individual described in the vignette have been used to assess vignette equivalence in a study by Kapteyn et al. [5]. The authors use responses obtained from two internet surveys on work disability conducted in the Netherlands and in the US. Vignettes were presented to respondents by randomly using either a female or a male name (i.e. faced with the same vignette, some respondents rated the health conditions for a woman while others rated the same condition for a man). Variability across the ratings allowed the authors to model the reporting behaviour of respondents as a function of the gender of the individual described in the vignette by explicitly including this variable as a regressor in the HOPIT model.<sup>12</sup> They reported that “for a given vignette description, a male vignette person is seen as more work disabled than a female vignette person, by both male and

female respondents” [5, p. 469]. In a similar vein, we evaluate whether individuals judge vignettes differently according to the gender of the person presented in a vignette and whether the person suffers from physical pain. We choose these individual characteristics for two reasons. First, on practical grounds, vignettes tend to represent “neutral” individuals, with little information on personal characteristics. Gender and pain are two of a very limited set of characteristics we can identify in the 20 vignettes considered. Secondly, while Kapteyn et al. [5] suggest that respondents tend to judge the vignettes differently according to whether the person in the vignette is female or male, Bago d’Uva [4] suggests that the elderly and the young interpret the construct of a vignette differently where the vignette describes a situation of physical pain.

For our analysis, we consider the pool of countries present in the WHS and, for illustration, make reference to the set of vignettes contained in the domains of “Dignity” and “Prompt attention”.<sup>13</sup> This set comprises 20 vignette questions answered by 858,570 individuals across all countries. Unfortunately, in the WHS there is no variability within a vignette in the gender of the individual described. The gender of the individual represented in each vignette is fixed and, accordingly, we are unable to adopt the methodology of Kapteyn et al. [5]. However, since within each domain of responsiveness in the WHS respondents are asked to evaluate a set of vignettes, we can exploit the variability in gender that is present across the vignettes within a given domain. To exploit this variability we perform a two-stage analysis using an estimated dependent variable regression model (EDV), as described by Lewis and Linzer [29]. In the first stage we model the reporting behaviour of respondents using a standard ordered probit model. We regress respondent ratings of the vignettes on the socio-demographic characteristics of the respondents and on a set of vignette-specific dummy variables [30, p. 61].<sup>14</sup> We then “store” the coefficients of the vignette-specific dummy variables.<sup>15</sup> In the second stage we regress the coefficients of the vignette-specific dummies on a dummy variable indicating if the person in the vignette is female, and on a dummy indicating if the person is in pain. Given the small sample size of the data we use in the second step regression, we correct for the potential presence of heteroskedasticity using the Efron

<sup>12</sup> See pp. 463–464 of Kapteyn et al. [5] for a formal description of the model estimated by the authors.

<sup>13</sup> This set of vignettes is coded as Set A in the WHS. We are unable to perform our analysis on a pool of all the vignettes contained in the responsiveness module, since each set is evaluated by a different group of respondents.

<sup>14</sup> The first vignette of the set (q7501) is assumed to be the base category.

<sup>15</sup> The strategy adopted by STATA (the software we utilize for the empirical estimates) for identification in the ordered probit model is to set the constant term to zero. Therefore, we assume the coefficient of the base reference vignette-dummy to be equal to zero.

robust standard error estimator [31], as suggested by Lewis and Linzer [29].

## Results

### Consistent and near-consistent ordering of vignettes

Using the data on health system responsiveness contained in the WHS, Table 2 reports the percentage of respondents for each domain in each country that gave an ordering of vignettes consistent with the global ordering, or had an ordering where only one vignette moved one or two ranks or two vignettes moved one rank each.<sup>16</sup> For each domain, there was no substantial variation across countries. For all countries (with few exceptions) more than 90% of respondents report consistent or near-consistent vignette orderings. For each domain, this percentage is equal to or greater than 95% in at least 52 countries. These preliminary results provide support for the assumption of vignette equivalence.

Table 3 presents the average percentage of respondents in each country that gave an ordering of vignettes consistent or near consistent with the global ordering, where countries are stratified by HDI groups and by the Inglehart–Welzel map groups. Average percentages are reported for each domain. In general, the average percentages are slightly higher for High HDI countries compared to Medium and Low HDI countries, and for countries characterised by “Secular-Rational” values compared to “Traditional” ones. However, the variation across HDI groups and across the Inglehart–Welzel grouping of countries is very small. These results provide further evidence that individuals across different countries tend to interpret the vignettes in a consistent way.

### Spearman rank order correlation coefficient

Table 4 provides frequency distributions for the SROCCs for an illustrative domain, “Clarity of Communication”, and Table 5 provides descriptive statistics across all domains. For each domain, the majority of the individuals reports an ordering of vignettes that is positive and highly correlated with the global ordering (the percentage of individuals whose SROCC is positive is between 87 and 95%, and the percentage of individuals with a SROCC larger than 0.5 is between 64 and 90%). The number of different rank order correlation coefficients reported in each domain appears to be high, and varies quite substantially (between 59 and 145) across domains.

<sup>16</sup> Australia, Turkey and Guatemala are excluded from the analysis since data on vignettes are not reported for all the domains considered.

Accordingly, in some domains there is a large number of alternative orderings (i.e. “Prompt Attention” and “Quality of Facilities”), while for others the number of ordering is small (i.e. “Clarity of communication”, “Autonomy” and “Social Support”). The number of SROCCs that occur with a frequency greater than 1% does not appear to be particularly large (on average 19) and it varies across domains much less than the number of alternative orderings.<sup>17</sup> Overall, the results suggest that vignettes ordering inconsistencies are more likely to occur because of measurement errors than because of the multidimensionality or cultural variation in the constructs of a domain. However, the possibility of some problem of multidimensionality appears to be higher in some domains (domains presenting a smaller number of alternative orderings, i.e. “Autonomy”) than in others.

Figure 2 shows the median SROCC across the data for each domain.<sup>18</sup> For most domains the vignettes appear to work well, with the median correlation assuming values between 0.85 and 0.95. Only the domains “Confidentiality” and “Choice” appear to have a slightly worse performance, presenting a median correlation that varies between 0.75 and 0.80. Figure 3 shows the median value of the SROCC across domains in each country. This value ranges from very high levels observed for Bangladesh and Comoros Islands (1.00 each) to more moderate values for Cote d’Ivoire and Namibia (0.84 and 0.74, respectively). However, the coefficient is greater than 0.90 in the majority of countries. The high values presented by the average SROCCs imply that cultural differences in the interpretation of vignettes across countries may not be of great concern.<sup>19</sup>

Table 6 provides the average SROCCs across all countries for individuals belonging to different socioeconomic groups. We perform this analysis following the suggestion of King et al. [7, p. 200], that “the key in detecting multidimensionality [of the vignette construct] is searching for inconsistencies that are systematically related to any measured variable”. In particular, Table 6 provides the SROCC between the ordering of vignettes defined at global level and the median ordering given by individuals within different education groups. The same information is provided for individuals stratified according to their level of income and gender. The vignettes appear to be ordered in a

<sup>17</sup> The coefficient of variation of the number of alternative orderings is 14.35, while for the number of SROCCs that occur with a frequency greater than 1% it is 0.91.

<sup>18</sup> For each domain, we have computed the median SROCC on the basis of tables analogous to Table 4.

<sup>19</sup> We are not aware of any study that explicitly defines a threshold of acceptability for the rank order correlation coefficient above which we can assume that vignette equivalence holds. However, according to Murray et al. [19], a rank order correlation coefficient greater than 0.9 strongly corroborates the assumption of vignette equivalence.

**Table 2** Percent of consistent and near-consistent ordering by domain and country

	Prompt attention	Dignity	Clarity of communication	Autonomy	Confidentiality	Choice	Quality of facilities	Social support
ARE	0.97	0.97	0.98	0.94	0.98	0.99	0.99	0.96
AUT	0.97	0.95	0.98	0.92	0.98	0.98	0.99	0.96
BEL	0.98	0.98	0.99	0.99	0.97	0.95	0.99	0.98
BFA	0.97	0.96	0.99	0.97	0.97	0.97	0.98	0.98
BGD	1.00	0.97	0.99	0.98	0.97	0.98	0.99	0.98
BIH	0.98	0.95	0.97	0.99	0.95	0.99	0.99	0.96
BRA	0.99	0.98	0.98	0.98	0.97	0.97	0.99	0.99
CHN	0.99	0.99	1.00	1.00	1.00	0.99	1.00	0.99
CIV	0.96	0.97	0.97	0.96	0.98	0.96	0.97	0.97
COG	0.99	0.99	0.99	0.97	0.98	0.98	0.99	0.99
COM	0.95	0.89	0.94	0.95	0.96	0.94	0.97	0.94
CZE	0.99	0.99	0.99	0.98	0.99	0.99	1.00	0.98
DEU	0.99	0.99	0.99	0.97	1.00	1.00	1.00	0.97
DNK	0.98	0.95	0.96	0.99	0.94	0.96	0.96	0.99
DOM	0.99	0.96	0.98	0.99	0.98	0.98	0.99	0.99
ECU	0.97	0.94	0.97	0.97	0.98	0.98	0.98	0.98
ESP	0.98	0.97	0.98	0.98	0.96	0.96	0.99	0.98
EST	0.99	0.98	0.99	0.99	1.00	0.99	0.99	0.99
ETH	0.97	0.97	0.99	0.96	0.99	0.99	0.99	0.99
FIN	0.99	0.98	0.99	1.00	0.98	0.98	1.00	0.99
FRA	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
GBR	0.98	0.98	0.99	0.98	0.98	0.98	1.00	0.98
GEO	0.98	0.97	1.00	0.99	0.99	0.99	1.00	0.98
GHA	0.98	0.95	0.98	0.98	0.98	0.97	0.98	0.99
GRC	0.99	0.97	0.99	0.99	0.99	0.98	0.99	0.96
HRV	0.99	0.99	0.99	0.99	0.99	0.99	1.00	0.99
HUN	0.99	1.00	0.99	0.99	0.99	0.98	0.98	0.99
IND	0.97	0.93	0.97	0.96	0.96	0.96	0.98	0.97
IRL	0.82	0.79	0.90	0.72	0.73	0.74	0.93	0.73
ISR	1.00	0.99	0.98	0.97	1.00	1.00	0.99	0.99
ITA	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.99
KAZ	0.95	0.92	0.96	0.68	0.76	0.81	0.98	0.77
KEN	0.96	0.94	0.98	0.98	0.98	0.96	0.99	0.98
LAO	1.00	0.98	1.00	0.99	0.99	0.98	1.00	0.99
LKA	0.94	0.92	0.95	0.95	0.93	0.94	0.95	0.97
LUX	0.98	0.97	0.99	0.99	0.97	0.99	0.99	0.99
LVA	0.99	0.98	0.98	0.99	0.99	0.95	0.99	0.99
MAR	0.98	0.98	0.99	0.97	0.98	0.97	0.99	0.98
MEX	0.95	0.92	0.97	0.96	0.95	0.96	0.97	0.97
MLI	1.00	0.99	0.99	1.00	0.99	0.97	0.99	0.98
MMR	1.00	0.98	0.99	0.96	0.99	0.98	0.99	0.97
MRT	0.96	0.90	0.90	0.92	0.91	0.89	0.93	0.96
MUS	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.98
MWI	0.98	0.95	0.97	0.98	0.98	0.97	0.98	0.97
MYS	1.00	0.99	1.00	0.99	0.99	0.99	1.00	0.99
NAM	0.95	0.94	0.89	0.97	0.94	0.92	0.92	0.97
NLD	0.99	0.98	1.00	1.00	1.00	0.99	0.99	1.00
NOR	0.97	0.97	0.99	0.97	0.97	0.96	0.97	0.97



**Table 2** continued

	Prompt attention	Dignity	Clarity of communication	Autonomy	Confidentiality	Choice	Quality of facilities	Social support
NPL	0.98	0.96	0.99	0.99	0.98	0.98	0.99	0.99
PAK	0.98	0.98	0.99	0.99	0.99	0.99	1.00	0.99
PHL	1.00	0.99	0.99	0.99	1.00	1.00	0.99	0.99
PRT	0.99	0.99	0.97	0.99	0.99	0.99	0.98	0.98
PRY	0.98	0.96	0.98	0.97	0.98	0.98	0.99	0.98
RUS	0.98	0.96	0.99	0.98	0.98	0.97	0.99	0.99
SEN	0.94	0.92	0.93	0.95	0.92	0.93	0.96	0.96
SVK	0.98	0.97	0.97	0.99	0.98	0.98	0.98	0.99
SVN	0.99	0.95	0.94	0.98	0.97	0.91	0.98	0.98
SWE	0.98	0.98	0.99	0.98	0.99	0.99	0.99	0.99
SWZ	0.92	0.88	0.88	0.92	0.98	0.97	0.91	0.96
TCD	0.97	0.95	0.98	0.99	0.90	0.86	0.98	0.98
TUN	0.94	0.88	0.99	0.99	0.99	0.99	1.00	0.99
UKR	0.95	0.92	0.98	0.97	0.93	0.96	0.98	0.96
URY	0.98	0.95	0.97	0.98	0.99	0.98	0.97	0.98
VNM	0.98	0.96	0.98	0.98	0.97	0.97	0.99	0.98
ZAF	0.92	0.92	0.96	0.94	0.92	0.92	0.97	0.97
ZMB	0.97	0.95	0.98	0.97	0.96	0.95	0.99	0.98
ZWE	0.99	0.97	1.00	0.99	0.98	0.98	1.00	0.99

**Table 3** Average percent consistent and near-consistent ordering, by Human Development Index (HDI) groups and by Inglehart–Welzel map groups

	Prompt attention	Dignity	Clarity of communication	Autonomy	Confidentiality	Choice	Quality of facilities	Social support
Average across all countries	0.97	0.96	0.98	0.97	0.97	0.96	0.98	0.97
Countries by HDI group								
High	0.98	0.97	0.98	0.97	0.97	0.97	0.99	0.97
Low	0.97	0.95	0.97	0.97	0.96	0.95	0.98	0.98
Medium	0.97	0.95	0.97	0.96	0.96	0.96	0.98	0.97
Countries by Inglehart value map								
Self-secular	0.98	0.97	0.98	0.98	0.98	0.97	0.99	0.98
Survival-secular	0.98	0.97	0.98	0.96	0.96	0.97	0.99	0.97
Self-traditional	0.96	0.94	0.97	0.94	0.95	0.95	0.98	0.95
Survival-traditional	0.97	0.95	0.97	0.97	0.97	0.96	0.98	0.98

similar way across the different socio-economic groups. The exception is individuals with a high level of education for the domain “Confidentiality”. For these individuals the ordering of the vignettes is less close to the global ordering, since the SROCC assumes values inferior to 0.8.

#### The HOPIT model

Table 7 presents the results from the responsiveness and reporting behaviour equation of the HOPIT model estimated on the pool of the 27 European countries present in

the WHS. For brevity, only the results related to the first cut point (the cut point separating the response category “very bad” from “bad”) are presented in the table. Results relating to other cut points are available on request. Belonging to the top income tertile, compared to the bottom, appears to be significantly related to experiencing a high level of responsiveness, while being a woman is negatively related to responsiveness (although this effect does not attain statistical significance). Elderly people and more educated people appear to face higher levels of responsiveness, but only for the former is the association

**Table 4** Spearman's rank order correlation coefficient between individual ordering of vignettes and the global ordering, for the domain "Clarity of Communication"

Spearman rank order correlation coefficient	<i>N</i>	%	Cum (%)
-1.000	4	0.03	0.03
-0.973	11	0.08	0.10
-0.949	8	0.06	0.16
-0.917	39	0.27	0.43
-0.913	47	0.33	0.76
-0.892	21	0.15	0.90
-0.884	16	0.11	1.01
-0.811	5	0.03	1.05
-0.791	2	0.01	1.06
-0.750	14	0.10	1.16
-0.730	10	0.07	1.23
-0.707	35	0.24	1.47
-0.667	3	0.02	1.49
-0.649	16	0.11	1.60
-0.632	1	0.01	1.61
-0.583	39	0.27	1.88
-0.559	91	0.63	2.51
-0.530	13	0.09	2.60
-0.487	8	0.06	2.66
-0.456	59	0.41	3.07
-0.406	24	0.17	3.24
-0.354	49	0.34	3.58
-0.324	3	0.02	3.60
-0.316	3	0.02	3.62
-0.250	71	0.49	4.11
-0.177	36	0.25	4.36
-0.162	34	0.24	4.60
-0.083	14	0.10	4.69
-0.081	9	0.06	4.76
0.000	301	2.09	6.85
0.081	15	0.10	6.95
0.083	19	0.13	7.08
0.158	1	0.01	7.09
0.162	43	0.30	7.39
0.177	44	0.31	7.69
0.250	95	0.66	8.35
0.316	2	0.01	8.37
0.324	28	0.19	8.56
0.354	99	0.69	9.25
0.406	37	0.26	9.50
0.456	132	0.92	10.42
0.474	5	0.03	10.46
0.487	34	0.24	10.69
0.530	149	1.03	11.73
0.559	335	2.33	14.05
0.583	176	1.22	15.27

**Table 4** continued

Spearman rank order correlation coefficient	<i>N</i>	%	Cum (%)
0.632	17	0.12	15.39
0.649	296	2.06	17.45
0.667	140	0.97	18.42
0.707	246	1.71	20.13
0.730	343	2.38	22.51
0.750	597	4.14	26.65
0.791	122	0.85	27.50
0.811	282	1.96	29.46
0.884	1,040	7.22	36.68
0.892	1,287	8.94	45.62
0.913	1,520	10.55	56.17
0.917	2,043	14.18	70.35
0.949	400	2.78	73.13
0.973	1,952	13.55	86.68
1.000	1,918	13.32	100.00
Total	14,403	100	

statistically significant. On average, individuals in Eastern European countries appear to face lower levels of responsiveness than in Austria, while we can not draw general conclusions for individuals in western European countries.

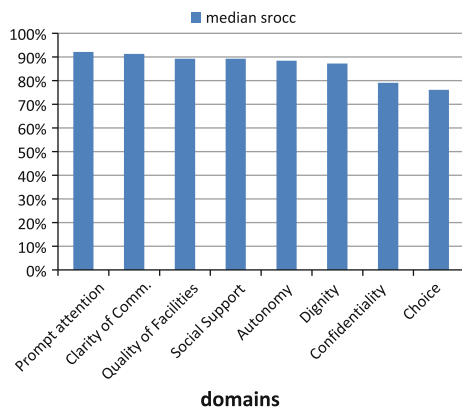
We stratify the European countries into three groups, according to the Inglehart–Welzer map, to reflect similar cultures, social norms and values. When we estimate the HOPIT model for each of the three groups of European countries separately (Catholic, Protestant and ex-communist), the coefficients for the country dummy variables are very robust both in the responsiveness equation and in the reporting behaviour equation. The coefficients retain the same sign when compared to the coefficients for the model where all the European countries are pooled together. Further, few of them change substantially. These results lend further support to the assumption of vignette equivalence.

Table 8 presents the results of the HOPIT model estimated across the full pool of countries and on "Self-Traditional", "Self-Secular", "Survival-Traditional", "Survival-Secular" countries separately. Again, the coefficients for the country dummy variables, both in the responsiveness and in the reporting behaviour equation, appear robust. Similar results, presented in Table 9, are obtained when the HOPIT model is estimated separately for countries stratified according to their level of HDI.<sup>20</sup> For both the responsiveness equation and the reporting behaviour equation, the coefficients for the country dummy

<sup>20</sup> Only the result related to the first cut point in the reporting bias equations is reported in Tables 8 and 9. Results related to the other cut points are available on request.

**Table 5** Descriptive statistics about the Spearman rank order correlation coefficient, by domain

	No. of different rank order correlation coefficients	Individuals whose correlation coefficient is positive (%)	Individuals whose correlation coefficient is >0.5 (%)	No. of rank order correlation coefficients that occur with frequency >1%
Prompt attention	145	93	85	13
Dignity	98	94	87	17
Clarity of communication	61	93	88	16
Autonomy	59	90	81	21
Confidentiality	80	88	64	22
Choice	125	87	70	26
Quality of facilities	143	95	90	16
Social support	59	93	85	19
Average	96	92	81	19

**Fig. 2** Median Spearman rank order correlation coefficient (SROCC) across domains

variables again appear robust. These results provide further evidence in favour of the assumption of vignette equivalence.

Test for multidimensionality of the constructs represented by vignettes

When we perform the two-stage analysis described in the section “[Assessment of multidimensionality of the constructs represented by vignettes](#)”, neither the regressors nor the constant term in the second step regression are statistically significant at the 95% percent level.<sup>21</sup> This result suggests that the gender of the person represented in the vignettes and his/her condition of pain do not influence the way respondents judge the vignettes.<sup>22</sup> Again, these results provide support to the vignette equivalence assumption.

<sup>21</sup> The results of the first and second step regression are available on request.

<sup>22</sup> The results are not affected by the distribution of the gender of individuals across vignettes, since both women and men are represented in vignettes describing high and low levels of responsiveness.

## Conclusion and discussion

Despite the growing popularity of the vignette methodology to address the issue of systematic reporting heterogeneity in self-reported data, the formal evaluation of the validity of this methodology has remained a topic of research. Two critical assumptions need to hold in order for the method to be valid. This paper presents analyses to assess the validity of the assumption of vignette equivalence using data on health system responsiveness contained within the WHS.

We first performed non-parametric analyses based on the global ordering of the vignettes. Secondly, after estimating a HOPIT model for responsiveness on the pool of countries, we performed sensitivity analyses stratifying the countries in our sample on the basis of the Inglehart–Welzel map and HDI groupings. Thirdly, we adopted a two-step regression procedure to evaluate the possibility that an individuals’ perceptions of the construct described by a vignette differ according to the characteristics of the person described in the vignette. The results derived from our analysis do not contradict the assumption of vignette equivalence. Accordingly, they lend support to the use of the vignette methodology to correct for the presence of reporting heterogeneity.

A potential limitation of our analysis is that, for brevity, only a limited set of domains of responsiveness were used. For the analysis in the section on “[The HOPIT model](#)” we considered only “Dignity”, while in “[Test for multidimensionality of the constructs represented by vignettes](#)”, we refer to “Dignity” and “Prompt Attention”. Some caution is, therefore, required in generalising our results to other domains of the responsiveness construct.

The results refer only to the assumption of vignette equivalence and do not consider response consistency. Recent literature has tried to assess the validity of the latter assumption [6, 17]. The majority of these studies

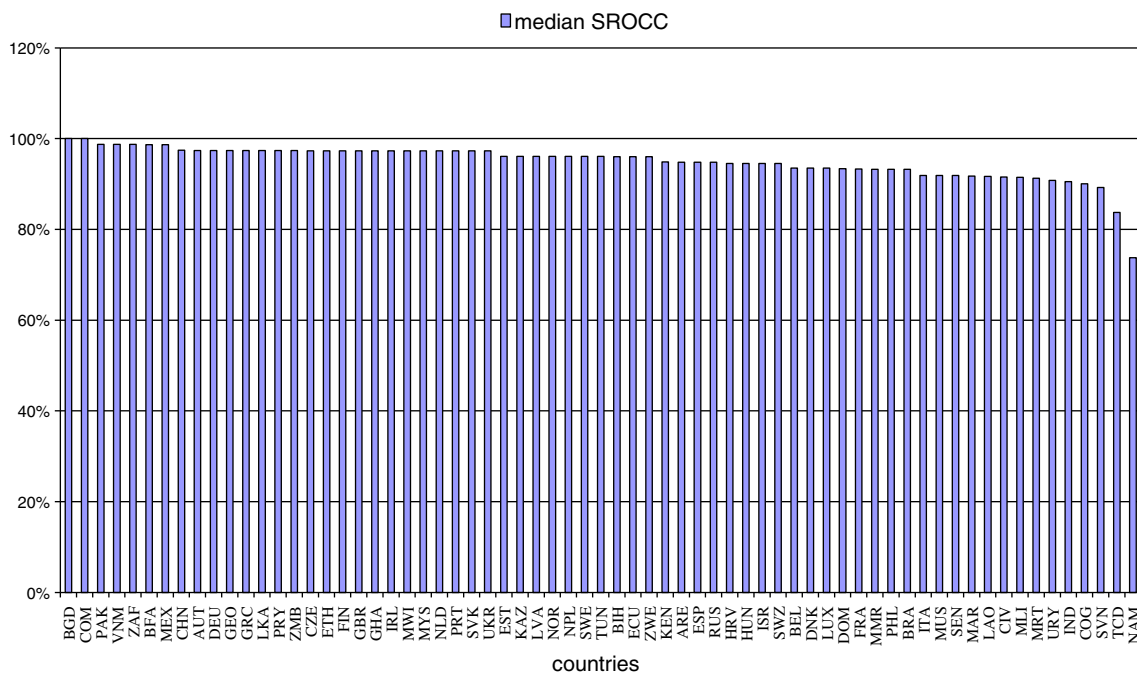


Fig. 3 Median SROCC across countries

Table 6 Average Spearman rank order correlation coefficient (SROCC) across all surveys

	Prompt attention	Dignity	Clarity of communication	Autonomy	Confidentiality	Choice	Quality of facilities	Social support
<b>Education groups</b>								
No formal schooling	0.97	0.95	1.00	1.00	1.00	1.00	0.92	1.00
Less than primary school	0.97	0.97	1.00	0.91	1.00	1.00	1.00	1.00
Primary school completed	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97
Secondary school completed	1.00	0.89	1.00	1.00	1.00	1.00	1.00	1.00
High school completed	0.97	1.00	0.95	1.00	0.79	0.97	1.00	0.97
College completed	0.95	1.00	0.95	1.00	0.79	0.97	1.00	0.95
Post graduate degree completed	0.95	1.00	0.95	0.97	0.75	0.89	1.00	0.95
Average	0.97	0.97	0.97	0.98	0.90	0.98	0.99	0.98
<b>Income quintiles</b>								
1st	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
2nd	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
3rd	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4th	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00
5th	0.95	1.00	0.97	0.91	1.00	0.97	1.00	1.00
Average	0.99	0.99	0.99	0.98	1.00	0.99	1.00	1.00
<b>Gender</b>								
Female	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00
Male	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00

**Table 7** European countries: coefficients and standard errors for the responsiveness equation and the reporting behaviour equation (first cut point) of the hierarchical ordered probit model (HOPIT) model, for the domain “Dignity”, for the pool of countries and for countries stratified by the Inglehart–Welzer value map

	Europe overall		Catholic countries		Communist countries		Protestant countries	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Responsiveness equation, xb								
dumI2	−0.004	0.026	0.041	0.051	0.029	0.033	−0.035	0.067
dumI3	0.063	0.028	0.094	0.057	0.073	0.035	0.124	0.072
Female	−0.010	0.021	0.080	0.040	−0.031	0.028	−0.011	0.055
age_yrs	0.006	0.001	0.009	0.001	0.005	0.001	0.007	0.002
edu_yrs	0.005	0.003	0.007	0.005	0.002	0.004	−0.003	0.009
BEL	0.275	0.108	0.284	0.110				
ESP	−0.099	0.078	−0.094	0.079				
FRA	0.224	0.118	0.236	0.119				
GRC	−0.079	0.098	−0.064	0.099				
ITA	−0.598	0.148	−0.577	0.148				
LUX	0.486	0.106	0.503	0.107				
PRT	−0.138	0.103	−0.120	0.105				
BIH	−0.247	0.095			−0.239	0.093		
CZE	0.053	0.098			0.053	0.096		
EST	−0.045	0.095			−0.035	0.093		
GEO	0.015	0.088			0.029	0.086		
HRV	−0.135	0.097			−0.127	0.095		
HUN	−0.229	0.088			−0.218	0.086		
KAZ	−0.450	0.081			−0.421	0.080		
LVA	−0.221	0.097			−0.203	0.096		
RUS	−0.547	0.080			−0.514	0.078		
SVK	−0.561	0.089			−0.545	0.087		
SVN	−0.192	0.107			−0.176	0.104		
UKR	−0.547	0.083			−0.514	0.082		
DEU	−0.159	0.093					−0.213	0.103
DNK	0.481	0.108					0.524	0.120
FIN	0.508	0.096					0.518	0.106
GBR	0.190	0.095					0.199	0.106
IRL	−0.107	0.105					−0.135	0.117
NLD	0.170	0.104					0.155	0.117
SWE	0.493	0.101					0.521	0.112
_cons	1.577	0.092	1.499	0.133	1.546	0.102	1.828	0.173
Reporting behaviour equation, m1								
dumI2	0.011	0.023	0.001	0.049	0.024	0.030	−0.013	0.056
dumI3	0.052	0.025	0.151	0.054	0.059	0.032	−0.037	0.058
Female	0.098	0.019	0.112	0.038	0.071	0.025	0.135	0.045
age_yrs	−0.002	0.001	−0.004	0.001	−0.002	0.001	−0.003	0.001
edu_yrs	0.007	0.003	−0.002	0.005	0.007	0.004	0.021	0.007
BEL	0.687	0.108	0.810	0.114				
ESP	0.111	0.082	0.098	0.086				
FRA	0.455	0.112	0.552	0.118				
GRC	0.229	0.097	0.243	0.102				
ITA	0.263	0.137	0.282	0.144				
LUX	0.315	0.107	0.371	0.112				



**Table 7** continued

	Europe overall		Catholic countries		Communist countries		Protestant countries	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
PRT	-0.172	0.104	-0.229	0.109				
BIH	0.133	0.093			0.124	0.091		
CZE	0.054	0.099			0.056	0.097		
EST	-0.058	0.098			-0.050	0.096		
GEO	-0.389	0.086			-0.384	0.084		
HRV	0.147	0.098			0.142	0.096		
HUN	-0.005	0.091			-0.007	0.089		
KAZ	-0.145	0.084			-0.147	0.083		
LVA	0.424	0.099			0.414	0.097		
RUS	0.084	0.083			0.077	0.082		
SVK	-0.085	0.090			-0.076	0.088		
SVN	0.367	0.102			0.343	0.100		
UKR	0.022	0.087			0.017	0.085		
DEU	0.171	0.096					0.165	0.099
DNK	0.682	0.101					0.696	0.105
FIN	0.176	0.097					0.178	0.100
GBR	0.508	0.094					0.506	0.098
IRL	0.715	0.102					0.737	0.105
NLD	-0.285	0.102					-0.320	0.106
SWE	0.639	0.098					0.651	0.101
_cons	-0.828	0.093	-0.674	0.132	-0.852	0.101	-0.952	0.144

**Table 8** All countries: coefficients and standard errors for the responsiveness equation and the reporting behaviour equation (first cut point) of the HOPIT model, for the domain “Dignity”, for the pool of countries and for countries stratified by the Inglehart–Welzer value map

	All countries		Self SEC		Self TRAD		Sur TRAD		Sur SEC	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Responsiveness equation, xb										
dumI2	0.023	0.011	0.003	0.021	0.029	0.020	0.032	0.013	0.012	0.019
dumI3	0.085	0.012	0.021	0.022	0.067	0.023	0.096	0.015	0.025	0.020
Female	0.027	0.009	0.045	0.018	0.028	0.018	0.039	0.011	0.006	0.016
age_yrs	0.004	0.000	0.005	0.001	0.005	0.001	0.003	0.000	0.003	0.000
edu_yrs	0.005	0.001	0.004	0.002	0.003	0.002	0.005	0.001	0.002	0.002
AUT	0.184	0.071	0.158	0.069						
BEL	0.424	0.078	0.388	0.077						
DEU	0.036	0.056	0.020	0.056						
DNK	0.621	0.077	0.581	0.075						
ESP	0.093	0.028	0.079	0.028						
FIN	0.651	0.060	0.610	0.060						
FRA	0.383	0.089	0.346	0.087						
GBR	0.343	0.060	0.313	0.059						
GRC	0.116	0.063	0.098	0.062						
ISR	0.112	0.065	0.091	0.064						
ITA	-0.355	0.122	-0.360	0.119						
LUX	0.628	0.075	0.584	0.073						

Table 8 continued

	All countries		Self SEC		Self TRAD		Sur TRAD		Sur SEC	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
NLD	0.333	0.071	0.305	0.069						
SVN	0.009	0.075	-0.009	0.073						
SWE	0.623	0.068	0.583	0.067						
BRA	0.201	0.030			0.202	0.030				
DOM	0.150	0.030			0.149	0.030				
ECU	0.016	0.038			0.023	0.038				
GTM	0.106	0.035			0.104	0.036				
IRL	0.097	0.074			0.095	0.075				
PRT	0.051	0.070			0.042	0.070				
PRY	0.323	0.032			0.332	0.032				
URY	0.320	0.042			0.316	0.043				
ARE	0.108	0.063					0.101	0.062		
BFA	-0.092	0.044					-0.094	0.043		
BGD	-0.186	0.035					-0.181	0.035		
CIV	-0.114	0.054					-0.111	0.053		
COG	-0.048	0.072					-0.049	0.071		
COM	-0.016	0.052					-0.009	0.051		
ETH	-0.534	0.070					-0.520	0.068		
GHA	-0.130	0.039					-0.123	0.038		
IND	-0.105	0.025					-0.097	0.025		
KEN	-0.350	0.034					-0.344	0.034		
LAO	-0.140	0.042					-0.135	0.041		
LKA	-0.466	0.026					-0.446	0.025		
MAR	-0.555	0.049					-0.543	0.048		
MLI	-0.103	0.152					-0.082	0.148		
MMR	0.187	0.050					0.183	0.049		
MRT	-0.366	0.051					-0.353	0.050		
MUS	0.091	0.031					0.088	0.031		
MWI	-0.183	0.031					-0.186	0.030		
MYS	-0.053	0.030					-0.056	0.029		
NAM	-0.034	0.034					-0.028	0.033		
NPL	-0.293	0.030					-0.283	0.030		
PAK	-0.199	0.032					-0.192	0.032		
PHL	0.131	0.024					0.129	0.023		
SEN	-0.415	0.071					-0.395	0.069		
SWZ	-0.235	0.059					-0.225	0.057		
TCD	-0.225	0.054					-0.214	0.053		
TUN	-0.346	0.029					-0.329	0.028		
ZAF	-0.079	0.045					-0.071	0.044		
ZMB	-0.187	0.033					-0.184	0.032		
ZWE	-0.136	0.036					-0.139	0.035		
BIH	-0.039	0.058							-0.034	0.056
CHN	0.065	0.043							0.067	0.041
CZE	0.223	0.063							0.222	0.061
EST	0.132	0.059							0.142	0.058
GEO	0.187	0.048							0.198	0.046

Table 8 continued

	All countries		Self SEC		Self TRAD		Sur TRAD		Sur SEC	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
HRV	0.050	0.061							0.059	0.059
HUN	-0.024	0.048							-0.027	0.047
KAZ	-0.229	0.035							-0.208	0.035
LVA	-0.023	0.063							-0.014	0.061
RUS	-0.315	0.032							-0.283	0.032
SVK	-0.342	0.049							-0.327	0.047
UKR	-0.320	0.039							-0.295	0.039
VNM	-0.318	0.035							-0.308	0.033
_cons	1.105	0.022	1.189	0.037	1.200	0.037	1.012	0.026	1.274	0.034
Reporting behaviour equation, m1										
dumI2	0.039	0.010	0.065	0.018	0.068	0.017	0.043	0.012	0.054	0.017
dumI3	0.068	0.010	0.080	0.019	0.079	0.019	0.062	0.013	0.066	0.018
Female	0.028	0.008	0.052	0.014	0.035	0.014	0.007	0.009	0.036	0.013
age_yrs	-0.001	0.000	-0.001	0.000	-0.001	0.000	0.000	0.000	-0.001	0.000
edu_yrs	0.009	0.001	0.005	0.002	0.013	0.002	0.006	0.001	0.008	0.002
AUT	0.059	0.075	0.087	0.076						
BEL	0.675	0.071	0.732	0.072						
DEU	0.219	0.054	0.245	0.054						
DNK	0.672	0.061	0.704	0.062						
ESP	0.163	0.025	0.173	0.025						
FIN	0.210	0.054	0.247	0.055						
FRA	0.464	0.076	0.532	0.077						
GBR	0.528	0.051	0.558	0.052						
GRC	0.273	0.055	0.300	0.056						
ISR	0.320	0.054	0.372	0.055						
ITA	0.307	0.107	0.340	0.107						
LUX	0.336	0.069	0.380	0.070						
NLD	-0.207	0.063	-0.157	0.063						
SVN	0.381	0.063	0.431	0.063						
SWE	0.649	0.056	0.682	0.057						
BRA	0.479	0.025			0.488	0.025				
DOM	0.002	0.030			0.001	0.030				
ECU	0.006	0.032			0.003	0.032				
GTM	-0.449	0.032			-0.437	0.033				
IRL	0.673	0.064			0.657	0.065				
PRT	-0.081	0.065			-0.079	0.065				
PRY	0.140	0.027			0.144	0.027				
URY	0.165	0.035			0.152	0.035				
ARE	-0.010	0.050					-0.005	0.050		
BFA	0.210	0.039					0.198	0.039		
BGD	0.083	0.028					0.080	0.027		
CIV	0.159	0.043					0.149	0.043		
COG	0.212	0.073					0.205	0.072		
COM	0.223	0.049					0.200	0.048		

Table 8 continued

	All countries		Self SEC		Self TRAD		Sur TRAD		Sur SEC	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
ETH	-0.016	0.038					-0.013	0.037		
GHA	0.231	0.034					0.222	0.033		
IND	-0.116	0.024					-0.123	0.023		
KEN	0.276	0.028					0.271	0.028		
LAO	-0.641	0.054					-0.622	0.053		
LKA	-0.122	0.024					-0.125	0.024		
MAR	0.671	0.038					0.661	0.037		
MLI	-0.276	0.151					-0.281	0.149		
MMR	-0.582	0.044					-0.564	0.043		
MRT	0.077	0.054					0.061	0.054		
MUS	0.387	0.027					0.382	0.027		
MWI	0.225	0.027					0.220	0.026		
MYS	-0.188	0.029					-0.171	0.029		
NAM	0.092	0.031					0.079	0.030		
NPL	-0.117	0.027					-0.119	0.027		
PAK	0.001	0.026					0.002	0.026		
PHL	-0.240	0.020					-0.230	0.019		
SEN	-0.076	0.051					-0.090	0.050		
SWZ	0.174	0.035					0.161	0.035		
TCD	-0.003	0.048					-0.027	0.047		
TUN	0.123	0.025					0.110	0.025		
ZAF	0.320	0.042					0.305	0.041		
ZMB	0.182	0.026					0.173	0.026		
ZWE	0.127	0.029					0.133	0.028		
BIH	0.171	0.049							0.176	0.049
CHN	-0.449	0.042							-0.420	0.042
CZE	0.122	0.057							0.133	0.058
EST	0.010	0.057							0.022	0.058
GEO	-0.309	0.034							-0.302	0.035
HRV	0.201	0.057							0.216	0.057
HUN	0.056	0.045							0.087	0.045
KAZ	-0.099	0.030							-0.078	0.031
LVA	0.448	0.057							0.469	0.058
RUS	0.120	0.029							0.136	0.030
SVK	-0.005	0.042							0.024	0.043
UKR	0.064	0.037							0.094	0.038
VNM	-0.247	0.033							-0.240	0.033
_cons	-1.013	0.018	-0.901	0.031	-0.961	0.029	-1.053	0.022	-0.928	0.029

test this assumption by comparing self-reported data to objective data (for example, comparing self-reported data on health to objectively measured levels of health). Unfortunately, the WHS does not contain objective measures of the level of responsiveness faced by respondents. Hence, we are currently unable to test this assumption in the WHS.

Our study provides an original contribution to the literature on anchoring vignettes by exploring the validity of the vignette equivalence assumption with reference to the concept of responsiveness. We adopt several strategies to assess the validity of the vignette equivalence assumption, employing both non-parametric and parametric methods. Overall, our results do not provide strong evidence to

**Table 9** All countries: coefficients and standard errors for the responsiveness equation and the reporting behaviour equation (first cut point) of the HOPIT model, for the domain “Dignity”, for the pool of countries and for countries stratified by HDI group

	All countries		High HDI		Medium HDI		Low HDI	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Responsiveness equation, xb								
dumI2	0.023	0.011	0.010	0.017	0.043	0.013	-0.004	0.020
dumI3	0.085	0.012	0.042	0.019	0.112	0.015	0.012	0.022
Female	0.027	0.009	0.020	0.015	0.029	0.011	0.043	0.017
age_yrs	0.004	0.000	0.005	0.000	0.004	0.000	0.003	0.001
edu_yrs	0.005	0.001	0.002	0.002	0.006	0.001	0.001	0.002
ARE	0.108	0.063	0.102	0.062				
AUT	0.184	0.071	0.168	0.070				
BEL	0.424	0.078	0.418	0.077				
BIH	-0.039	0.058	-0.049	0.058				
CZE	0.223	0.063	0.223	0.063				
DEU	0.036	0.056	0.027	0.056				
DNK	0.621	0.077	0.608	0.076				
ESP	0.093	0.028	0.082	0.028				
EST	0.132	0.059	0.133	0.059				
FIN	0.651	0.060	0.635	0.060				
FRA	0.383	0.089	0.372	0.088				
GBR	0.343	0.060	0.336	0.060				
GRC	0.116	0.063	0.100	0.063				
HRV	0.050	0.061	0.045	0.061				
HUN	-0.024	0.048	-0.039	0.048				
IRL	0.097	0.074	0.085	0.073				
ISR	0.112	0.065	0.106	0.065				
ITA	-0.355	0.122	-0.370	0.121				
LUX	0.628	0.075	0.612	0.074				
LVA	-0.023	0.063	-0.034	0.063				
MUS	0.091	0.031	0.080	0.031				
MYS	-0.053	0.030	-0.052	0.029				
NLD	0.333	0.071	0.333	0.070				
PRT	0.051	0.070	0.043	0.069				
SVK	-0.342	0.049	-0.334	0.049				
SVN	0.009	0.075	-0.002	0.074				
SWE	0.623	0.068	0.614	0.067				
URY	0.320	0.042	0.317	0.042				
BGD	-0.186	0.035			-0.184	0.035		
BRA	0.201	0.030			0.196	0.029		
CHN	0.065	0.043			0.064	0.042		
COG	-0.048	0.072			-0.049	0.071		
COM	-0.016	0.052			-0.009	0.051		
DOM	0.150	0.030			0.144	0.030		
ECU	0.016	0.038			0.013	0.037		
GEO	0.187	0.048			0.180	0.047		
GHA	-0.130	0.039			-0.127	0.038		
GTM	0.106	0.035			0.097	0.034		
IND	-0.105	0.025			-0.100	0.025		
KAZ	-0.229	0.035			-0.226	0.035		



Table 9 continued

	All countries		High HDI		Medium HDI		Low HDI	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
LAO	-0.140	0.042			-0.136	0.041		
LKA	-0.466	0.026			-0.451	0.025		
MAR	-0.555	0.049			-0.547	0.048		
MMR	0.187	0.050			0.180	0.049		
NAM	-0.034	0.034			-0.033	0.033		
NPL	-0.293	0.030			-0.284	0.030		
PAK	-0.199	0.032			-0.192	0.032		
PHL	0.131	0.024			0.124	0.023		
PRY	0.323	0.032			0.308	0.031		
RUS	-0.315	0.032			-0.302	0.032		
SWZ	-0.235	0.059			-0.225	0.057		
TUN	-0.346	0.029			-0.336	0.029		
UKR	-0.320	0.039			-0.312	0.039		
VNM	-0.318	0.035			-0.308	0.034		
ZAF	-0.079	0.045			-0.080	0.044		
BFA	-0.092	0.044					-0.099	0.044
CIV	-0.114	0.054					-0.112	0.053
ETH	-0.534	0.070					-0.506	0.068
KEN	-0.350	0.034					-0.339	0.034
MLI	-0.103	0.152					-0.102	0.148
MRT	-0.366	0.051					-0.357	0.051
MWI	-0.183	0.031					-0.184	0.031
SEN	-0.415	0.071					-0.399	0.069
TCD	-0.225	0.054					-0.221	0.054
ZMB	-0.187	0.033					-0.185	0.032
ZWE	-0.136	0.036					-0.137	0.035
_cons	1.105	0.022	1.256	0.032	1.058	0.025	1.144	0.038
Reporting behaviour equation, ml								
dumI2	0.039	0.010	0.059	0.015	0.042	0.012	0.059	0.016
dumI3	0.068	0.010	0.066	0.017	0.069	0.013	0.070	0.018
Female	0.028	0.008	0.039	0.012	0.026	0.009	0.020	0.013
age_yrs	-0.001	0.000	-0.002	0.000	-0.001	0.000	0.000	0.000
edu_yrs	0.009	0.001	0.007	0.002	0.009	0.001	0.009	0.002
ARE	-0.010	0.050	0.025	0.051				
AUT	0.059	0.075	0.082	0.077				
BEL	0.675	0.071	0.729	0.072				
BIH	0.171	0.049	0.186	0.049				
CZE	0.122	0.057	0.135	0.059				
DEU	0.219	0.054	0.239	0.055				
DNK	0.672	0.061	0.711	0.062				
ESP	0.163	0.025	0.173	0.025				
EST	0.010	0.057	0.027	0.058				
FIN	0.210	0.054	0.241	0.055				
FRA	0.464	0.076	0.518	0.077				
GBR	0.528	0.051	0.559	0.052				
GRC	0.273	0.055	0.295	0.056				
HRV	0.201	0.057	0.224	0.058				

Table 9 continued

	All countries		High HDI		Medium HDI		Low HDI	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
HUN	0.056	0.045	0.088	0.046				
IRL	0.673	0.064	0.716	0.065				
ISR	0.320	0.054	0.357	0.055				
ITA	0.307	0.107	0.332	0.109				
LUX	0.336	0.069	0.371	0.070				
LVA	0.448	0.057	0.485	0.059				
MUS	0.387	0.027	0.415	0.028				
MYS	-0.188	0.029	-0.184	0.030				
NLD	-0.207	0.063	-0.188	0.064				
PRT	-0.081	0.065	-0.083	0.066				
SVK	-0.005	0.042	0.015	0.043				
SVN	0.381	0.063	0.425	0.064				
SWE	0.649	0.056	0.686	0.058				
URY	0.165	0.035	0.172	0.036				
BGD	0.083	0.028			0.088	0.028		
BRA	0.479	0.025			0.478	0.025		
CHN	-0.449	0.042			-0.432	0.042		
COG	0.212	0.073			0.215	0.072		
COM	0.223	0.049			0.222	0.048		
DOM	0.002	0.030			0.007	0.030		
ECU	0.006	0.032			0.012	0.032		
GEO	-0.309	0.034			-0.299	0.034		
GHA	0.231	0.034			0.235	0.033		
GTM	-0.449	0.032			-0.431	0.032		
IND	-0.116	0.024			-0.112	0.023		
KAZ	-0.099	0.030			-0.094	0.030		
LAO	-0.641	0.054			-0.622	0.054		
LKA	-0.122	0.024			-0.120	0.024		
MAR	0.671	0.038			0.671	0.038		
MMR	-0.582	0.044			-0.563	0.043		
NAM	0.092	0.031			0.089	0.031		
NPL	-0.117	0.027			-0.109	0.027		
PAK	0.001	0.026			0.008	0.026		
PHL	-0.240	0.020			-0.229	0.019		
PRY	0.140	0.027			0.146	0.026		
RUS	0.120	0.029			0.122	0.029		
SWZ	0.174	0.035			0.167	0.035		
TUN	0.123	0.025			0.126	0.025		
UKR	0.064	0.037			0.070	0.037		
VNM	-0.247	0.033			-0.238	0.033		
ZAF	0.320	0.042			0.317	0.041		
BFA	0.210	0.039					0.221	0.039
CIV	0.159	0.043					0.168	0.042
ETH	-0.016	0.038					0.004	0.037
KEN	0.276	0.028					0.284	0.028
MLI	-0.276	0.151					-0.255	0.149
MRT	0.077	0.054					0.077	0.054

**Table 9** continued

	All countries		High HDI		Medium HDI		Low HDI	
	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
MWI	0.225	0.027					0.232	0.026
SEN	-0.076	0.051					-0.068	0.050
TCD	-0.003	0.048					0.002	0.048
ZMB	0.182	0.026					0.195	0.026
ZWE	0.127	0.029					0.142	0.028
_cons	-1.013	0.018	-0.883	0.027	-1.027	0.021	-1.036	0.029

suggest that the assumption does not hold and, accordingly, support the use of the anchoring vignette approach to adjust self-reported data for systematic differences in reporting behaviour.

**Acknowledgments** This research was funded by the Economic and Social Research Council under the Public Services Programme, grant number RES-166-25-0038, and under the ESRC Large Grant Scheme, grant number RES-060-25-0045. We would like to thank the World Health Organization for providing access to the World Health Survey and, in particular, Somnath Chatterji, Amit Prasad, Nicole Valentine and Emese Verdes. We are also grateful to the Health, Econometrics and Data Group Seminar Series at the University of York and to the Health Economics Seminars, Erasmus University for helpful comments on an earlier draft.

**Appendix: The HOPIT model**

Reporting behaviour equation

To identify the thresholds 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:

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

where  $K_{ik}$  is the vector of vignettes,  $\eta_k$  is a conformably dimensioned vector of parameters and  $\varepsilon_{ik}^v$  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:

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

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

Should the thresholds represent fixed constants,  $\mu^j$ , common to all individuals, then the above mapping is

common to the ordered probit model. For the HOPIT model the thresholds are assumed to be functions of covariates,  $X$  such that:

$$\mu_i^j = X_i \gamma^j \tag{3}$$

where  $\mu_i^j, j = 1, \dots, 5$  are parameters to be estimated along with  $\eta_k$ . Further, we assume an ordering of the thresholds such that  $\mu_i^1 < \mu_i^2 < \dots < \mu_i^5$ . If we impose the restriction that the covariates affect all thresholds by the same magnitude, then we have parallel cut-point shift. However, if the degree of reporting heterogeneity varies across thresholds such that it is greater at some levels of responsiveness than others, we refer to this as non-parallel shift [30].

Responsiveness equation

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

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

where  $Z_i$  represents a set of regressors predictive of responsiveness. 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:

$$r_i^s = j \quad \text{if } \mu_i^{j-1} \leq R_i^{s*} < \mu_i^j \quad \text{for } \mu_i^0 = -\infty, \mu_i^5 = \infty, \forall i; \tag{5}$$

$j = 1, \dots, 5$

where  $\mu_i^j$  are defined by (3) with  $\gamma^j$  fixed and it is assumed that  $R_{ik}^{v*}$  and  $R_i^{s*}$  are independent for all  $i = 1, \dots, N$  and  $k = 1, \dots, V$ . Note that  $\hat{\sigma}^2$  in Eq. 4 is identified due to the thresholds being fixed through the reporting behaviour equation. It follows that the probabilities associated with each of the five categories are given by:

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

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



## References

1. Murray, C., Frenk, J.: A framework for assessing the performance of health systems. *Bull. World Health Org.* **78**, 717–731 (2000)
2. Valentine, N., De Silva, A., Kawabata, K., Darby, C., Murray, C.J.L., Evans, D.: Health system responsiveness: concepts, domains and operationalization. In: Murray, C.J.L., Evans, D.B. (eds.) *Health systems performance assessment: debates, methods and empiricism*, pp. 573–596. World Health Organization, Geneva (2003)
3. Salomon, J., Tandon, A., Murray, C.J.L., World Health Survey Pilot Study Collaborating Group: Comparability of self-rated health: cross sectional multi-country survey using anchoring vignettes. *Brit. Med. J.* **328**(258), 258–261 (2004)
4. Bago d’Uva, T., van Doorslaer, E., Lindeboom, M., O’Donnell, O.: Does reporting heterogeneity bias the measurement of health disparities? *Health Econ.* **17**(3), 351–375 (2008)
5. Kapteyn, A., Salomon, J., van Soest, A.: Vignettes and self-reports of work disability in the US and the Netherlands. *AER* **97**(1), 461–473 (2007)
6. Van Soest, A., Delaney, A., Harmon, C., Kapteyn, A., Smith, J.P.: Validating the use of vignettes for subjective threshold scales. Discussion Paper, Tilburg University (2007)
7. King, G., Murray, C.J.L., Salomon, J., Tandon, A.: Enhancing the validity and cross-cultural comparability of measurement in survey research. *Am. Polit. Sci. Rev.* **98**(1), 184–191 (2004)
8. Rice, N., Robone, S., Smith, P.C.: Analysis of the validity of the vignette approach to correct for heterogeneity in reporting health system responsiveness. HEDG Working Paper, University of York (2009)
9. Rice, N., Robone, S., Smith, P.C.: International comparison of public sector performance: the use of anchoring vignettes to adjust self-reported data. *Evaluation* **16**(1), 81–101 (2010)
10. Sirven, N., Santos-Eggimann, B., Spagnoli, J.: Comparability of health care responsiveness in Europe using anchoring vignettes from SHARE. IRDES working paper DT15 (2008)
11. Kristensen, N., Johansson, E.: New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Econ.* **15**, 96–117 (2008)
12. Hsee, C.K., Tang, J.N.: Sun and water: on a modulus-based measurement of happiness. *Emotion* **7**, 213–218 (2007)
13. Javaras, K.N., Ripley, B.D.: An “unfolding” latent variable model for likert attitude data: drawing inferences adjusted for response style. *JASA* **102**(478), 454–463 (2007)
14. Grzymala-Busse, A.: *Rebuilding Levithan: party competition and state exploitation in post-communist democracies*. Cambridge University Press, New York (2007)
15. Bago d’Uva, T., Lindeboom, M., O’Donnell, O., van Doorslaer E.: Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. HEDG Working Paper, University of York (2009)
16. Hopkins, D., King, G.: Improving anchoring vignettes: designing surveys to correct interpersonal incomparability. *Public Opin. Q.* (2010) (in press)
17. Gupta N., Kristensen, N., Pozzoli, D.: External validation of the use of vignettes in cross-country health studies. IZA Discussion Paper (2008)
18. Wand, J.: Credible comparisons using interpersonally incomparable data: ranking self-evaluations relative to anchoring vignettes or other common survey questions. Mimeo (2007)
19. Murray, C.J.L., Ozaltin, E., Tandon, A.J., Salomon, J.: Empirical evaluation of the anchoring vignettes approach in health surveys. In: Murray, C.J.L., Evans, D.B. (eds.) *Health systems performance assessment: debates, methods and empiricism*, pp. 369–399. World Health Organization, Geneva (2003)
20. King, G., Wand, J.: Comparing incomparable survey responses: new tools for anchoring vignettes. *Polit. Anal.* **15**(1, Winter), 46–66 (2007)
21. Kapteyn, A., Salomon J., van Soest, A.: Are Americans really less happy with their incomes? Rand Working Paper (2008)
22. Üstün, T.B., Chatterji, S., Mechbal, A., Murray, C.: The world health surveys. In: Murray, C.J.L., Evans, D.B. (eds.) *Health systems performance assessment: debates, methods and empiricism*, pp. 762–796. World Health Organization, Geneva (2003)
23. Valentine, N., Prasat, A., Rice, N., Robone, S., Chatterji, S.: Health systems responsiveness—a measure of the acceptability of health care processes and systems. In: Mossialos, E., Smith, P., Leatherman, S. (eds.) *Performance measurement for health system improvement: experiences, challenges and prospects*, pp. 256–305. WHO European Regional Office, London (2009)
24. Ferguson, B.D., Tandon, A., Gakidou, E., Murray, C.J.L.: Estimating permanent income using indicator variables. In: Murray, C.J.L., Evans, D.B. (eds.) *Health systems performance assessment: debates, methods and empiricism*, pp. 748–760. World Health Organization, Geneva (2003)
25. United Nation Development Programme: *Human Development Report*. New York (2006)
26. Inglehart R.: Inglehart–Welzel cultural map of the World. <http://www.worldvaluessurvey.org>
27. Tandon, A., Murray, C.J.L., Salomon, J.A., King, G.: Statistical models for enhancing cross-population comparability. In: Murray, C.J.L., Evans, D.B. (eds.) *Health systems performance assessment: debates, methods and empiricism*, pp. 727–746. World Health Organization, Geneva (2003)
28. Terza, J.V.: Ordinal probit: a generalization. *Commun. Stat.* **14**(1), 1–11 (1985)
29. Lewis, J.B., Linzer, D.A.: Estimating regression models in which the dependent variable is based on estimates. *Polit. Anal.* **13**, 345–364 (2005)
30. Jones, A.M., Rice, N., D’Uva, T.B., Balia, S.: *Applied health economics*. Routledge, New York (2007)
31. Efron, B.: *The Jackknife, the Bootstrap and other resampling plans*. Society for Industrial and Applied Mathematics, Philadelphia (1982)

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.