Cultural Values: Can They Explain Differences in Health Utilities between Countries?

Bram Roudijk , **A. Rogier T. Donders, and Peep F. M. Stalmeier;**
Cultural Values Group

Introduction. Health utilities are widely used in health care. The distributions of utilities differ between countries; some countries more often report worse than dead health states, while mild states are valued more or less the same. We hypothesize that cultural values explain these country-related utility differences. **Research Question.** What is the effect of sociodemographic background, methodological factors, and cultural values on differences in health utilities? **Methods and Analyses.** Time tradeoff data from 28 EQ-5D valuation studies were analyzed, together with their sociodemographic variables. The dependent variable was Δu, the utility difference between mild and severe states. Country-specific cultural variables were taken from the World Values Survey. Multilevel models were used to analyze the effect of sociodemographic background, methodology (3L v. 5L), and cultural values on Δu. Intraclass correlation (ICC) for country variation was used to assess the impact of the predicting variables on the variation between countries. **Results.** Substantial variation in Δu was found between countries. Adding cultural values did not reduce ICCs for country variation. Sociodemographic background variables were only weakly associated with Δu and did not affect the ICC. Δu was 0.118 smaller for EQ-5D-5L studies. **Discussion.** Δu varies between countries. These differences were not explained by national cultural values. In conclusion, despite correction for various variables, utility differences between countries remain substantial and unexplained. This justifies the use of country-specific value sets for instruments such as the EQ-5D.

Health utilities are commonly used in cost-utility analysis of drugs and interventions in health care.[1] They provide the quality weight in the quality-adjusted life year (QALY) model and are usually measuring preferences for hypothetical health states derived from instruments such as the SF-6D, EQ-5D, or the Health Utility Index (HUI).[2–4] Differences between countries have been observed in valuations of health utilities by the general public.[5,6] In particular in the EQ-5D valuation system, for severe health states, the differences between countries can be as large as 0.4.[7] Similarly, differing amounts of health states considered worse than dead have been observed, which are health states that are assigned negative utilities. For example, the 2016 value set for England has 5% of its health states valued worse than dead, while the 2017 Indonesian value set has 35% of its health states valued worse than dead.[8,9]

Differences between countries persist despite efforts to harmonize valuation studies. Several sources may contribute to these differences such as the sociodemographic

Corresponding Author
Bram Roudijk, Radboud University Medical Center, Radboud Institute for Health Sciences, Geert Grooteplein 21, PO Box 9101, Nijmegen, Gelderland 6500 HB, the Netherlands. (Bram.Roudijk@Radboudumc.nl).

backgrounds of the respondents, different valuation methods, and differing cultural values between countries. First, sociodemographic factors such as age, sex, education, and marital status have been shown to be related to utilities for health states, albeit weakly.[10,11] Also, respondents' self-reported health and self-description are related to the valuation of health states.[12–14] Second, different methods of valuation might also affect the outcomes of valuation studies. There are a variety of methods to value health states, such as the standard gamble (SG), visual analog scale (VAS), discrete choice experiments (DCEs), and, most commonly used, the time tradeoff (TTO).[15,16] The results of valuation studies differ systematically by valuation method.[1] Also, a variety of other methodological factors affect utilities, such as layouts, indifference procedures, scale anchors, and transformations of values for worse than dead health states.[17,18] If these factors explain differences in health state valuations within studies, they might also explain differences in health state valuations between countries.

Cultural values have also been hypothesized to explain differences in utilities between studies.[5,6] Cultural values can be defined as what should be judged as good or evil by a group.[19] Cultural values have been operationalized by pioneers such as Hofstede and Inglehart.[20–22] There is some evidence that cultural values are related to health; for instance, the cultural values of Inglehart were shown to be related to self-reported health.[23] Furthermore, it was shown that utility differences between countries were related to Hofstede's cultural values.[24] The aim of this study is to test whether the variation in utilities is caused by differences in sociodemographic background, methodological factors, or cultural values. Our research question is as follows: What is the effect of country, sociodemographic profile, methodological factors, and cultural values on differences in health utilities?

Methods

General Approach

We are interested in the determinants of variation in health utilities between countries, which we aim to explain by sociodemographic background, methodological factors, and cultural values. We focus on differences between utilities for mild and severe health states for reasons explained in the analyses section.

Valuation Instrument Used in Various Countries

The preference-based valuation instrument that will be used in this study is the EQ-5D, developed by the EuroQol Group. This tool assesses utilities for health states.[25] The EQ-5D-3L is a health state classification system with 3 levels and 5 dimensions that span the domains of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, in that order. On each of these dimensions, one can have 1) no problems, 2) some problems, or 3) extreme problems. A score of 11321 on the EQ-5D-3L indicates that a hypothetical person has no problems with mobility and self-care, has extreme problems with performing usual activities, some problems with pain or discomfort, and no problems with anxiety or depression.

The EQ-5D-5L was developed to improve the sensitivity of the EQ-5D and to reduce ceiling effects present in the EQ-5D-3L.[26] In addition to the usual 3 levels of severity, 2 intermediate levels are introduced. One can now have 1) no problems, 2) slight problems, 3) moderate problems, 4) severe problems, and 5) extreme problems.

Utility weights are assigned to EQ-5D health states through valuation studies. The EQ-5D instrument assigns utility to health states by employing the TTO or composite TTO (cTTO) methods.[27] For states better than dead, the TTO and cTTO methods allow respondents to choose between 10 years in good health v. 10 years in the health state to be valued. If 10 years in good health is preferred, the respondent is faced with the choice between 10 years in the health state or, for example, 9 years in good health and so on. The TTO, for states worse than dead, gives respondents a choice

between dying immediately or spending x years in the health state, followed by $10 - x$ years in good health.[28] The cTTO, for states worse than dead, gives respondents a choice between (10 years in good health followed by 10 years in the state worse than dead) and (10 years in good health). Subsequently, another 10 years of living in good health can then be traded.[27] Utility can then be assigned to health states using the QALY model, based on the amount of years traded by the respondents for the given health states.[27]

Data Collection

Valuation data were collected from existing EQ-5D valuation studies. Principal investigators (PIs) of EQ-5D valuation studies were contacted through email, to ask for their data. Reminders were sent if the PIs did not respond after a few weeks, and more reminders were sent if necessary. PIs were also contacted at 3 EuroQol Meetings in 2016 and 2017.

Measures

Sociodemographic variables. We selected variables that were assumed to be related to Δu. Age, sex, education, and EQ-5D self-description were collected from the valuation data sets. These variables were shown to be related to utilities in other studies.[10,11] If only age classes were available, the mean of each class was assigned to the age of the respondent. Educational status was coded as low, medium, and high. Low education corresponds to at most finished primary school, middle indicates secondary education, and high indicates at least some tertiary education. Many valuation studies have also collected the respondent's own EQ-5D self-description, indicating whether the respondent had problems on 1 or more domains of the EQ-5D.

Methodological variables. Methodological variables were extracted from the research papers of the included valuation studies. We initially considered known methodological factors that affect the outcomes of utility assessments as a basis for the inclusion of methodological variables.[18] However, EQ-5D-3L studies were fairly homogeneous, as the methodology of most 3L studies was derived from the original Measurement and Valuation of Health (MVH) study conducted in Britain.[3] With the introduction of the EQ-5D-5L, the methodology of the valuation studies was standardized, reducing methodological differences within the 5L studies.[29] A major difference between 3L and 5L studies is that in 5L studies, the cTTO was introduced. Hence, a variable representing whether a study is 3L or 5L was used in our analyses, which captures, among other differences, whether TTO or cTTO is used or whether DCE was done complementary to the cTTO.

Cultural variables. There are several theories on cultural values, for example, the approach of Hofstede and the approach of Inglehart.[20,22] The approach by Inglehart is based on the World Values Survey in which large representative target samples are obtained. Therefore, we use their theory to derive national levels of cultural values on 2 dimensions: traditional v. rational/secular values and survival v. self-expression values.[20,21,30] Traditional values are indicated by a negative score on the traditional/rational-secular dimension and are related to religion (importance of God), authority, national pride, lower levels of tolerance toward homosexuality and abortion, and stronger family ties, while rational-secular values imply the opposite and are indicated by a positive score on the traditional/rational-secular dimension. Survival values are indicated by a negative score on the survival/self-expression dimension and are indicated by low levels of trust, low levels of political activism, and low levels of tolerance for abortion and homosexuality, while self-expression values imply the opposite.

Analyses

General Approach

The dependent variable was Δu, which represented the observed TTO utility difference between mild and severe states. We used Δu instead of utilities themselves, as mild states are valued similarly in most countries, while there are large country differences between the values assigned to severe states. In other words, utilities for severe states are dependent on the country they were measured in, while utilities for mild states are not. When utilities are used, interactions are needed in the analysis, such as country $\times$ utilities or age $\times$ countries $\times$ utilities. This makes the interpretation of the results difficult and, given the relatively low number of countries, introduces a risk of overfitting the model. The proposed method using Δu does not require interactions to be modeled.

Another reason for using Δu and not utilities is that utilities are bounded upward by the value for perfect health equal to 1, for which our linear models cannot account. Also, for decision making, at least interval scale properties are necessary, so the use of differences is natural. In other words, using Δu as the dependent variable allows us to treat utilities for mild states as an anchor

with respect to the utilities for severe states and provides a feasible method to answer our research question.

Classification of Mild, Moderate, and Severe States

As we focus on the difference between mild and severe states, it was necessary to make a classification of which states are considered mild, moderate, or severe. There is, to our knowledge, no universal protocol to do so. For EQ-5D-3L health states, we followed the procedure of Luo et al.[31] Mild states had at most "moderate problems" or a "2" on 2 domains. Severe states had "extreme problems" or "3s" on at least 2 of the health domains. All other health states were considered moderate. For the EQ-5D-5L, we employed a similar procedure. Mild states contained at most 2 "3s" (moderate problems on a maximum of 2 domains), severe states contained at least two "5s" (extreme problems on at least 2 domains), and all other states were considered moderate.

Dependent Variable: Utility Differences between Mild and Severe States

Δu was constructed as follows. Each respondent had valued a number of health states, usually around 10, which may include mild, moderate, and severe states. Each of these health states was assigned utility by the respondent. For each respondent, the average utility for the severe states was subtracted from the average utility for mild states, which provided an indicator for the difference in valuation between mild and severe states. For example, if a respondent had valued 10 states, 3 mild, 4 moderate, and 3 severe states, only the 3 mild and 3 severe states were used in our analysis. As health utilities have values between [–1,1], Δu could in principle take any value between [–2,2] but in practice took values between [0,2]. Respondents who did not value mild and severe states were excluded, as Δu was undefined.

Independent Variables

Δm : *correcting for stimulus differences.* Health states are the stimuli presented to respondents in valuation tasks. The sets of health states differ between countries and respondents, which made it necessary to control for these differences. Consequently, Δm was included, based on what is often called the severity (or misery) index. The severity index is the sum of the score on the 5 domains of the EQ-5D. For 3L health states, it ranges from 5 (no problems on any domain) to 15 (extreme problems on every domain), while for 5L health states, it ranges from 5 (no problems on any domain) to 25 (extreme problems on every domain). These were rescaled to a common severity index that ranged [0, 1] to analyze both 3L and 5L data. As above, the average severity index for mild states was subtracted from the average severity index for severe states to create Δm. Δm can be seen as the average difference in deviation from full health between mild and severe states and should be treated as a city-block metric, which varied at the respondent level.

Sociodemographic, methodological, and cultural variables. Age was standardized to have a mean of 0 and a standard deviation of 1, using the overall mean of all included respondents, while sex was coded as male (0) and female (1) and education was coded low (1), middle (2), and high (3). The respondent's own EQ-5D self-description was transformed to a single variable and rescaled to [0,1] by summing up the levels on each dimension, subtracting 5, and dividing by 10 (in the case of 3L) or 20 (in the case of 5L). For each study, a dummy indicating whether a study was 3L (0) or 5L (1) was included.

The 2 cultural variables, traditional/rational-secular and survival/self-expression, ranged between –2 and 2, and were taken from the World Values Survey. The World Values Survey currently has data available for 6 study waves, conducted in different time periods. Means for the 2 cultural dimensions can be calculated by wave. Cultural data were matched on year of collection of the EQ-5D data. If this was not possible, the wave that is closest to the date of collection of the EQ-5D data was used.

Data Structure and Models

In our analyses, we were interested in Δu and whether Δu varied between countries. Furthermore, we were interested in whether sociodemographic background, methodological factors, and cultural variables could explain this variation. As respondents were nested in valuation studies, 2-level mixed-effects models were employed, which could account for this nested structure. (Some countries have both a 3L and a 5L data set [e.g., Spain, the Netherlands, Japan, Singapore, Thailand] for which different cultural values can be used. Respondents are nested in studies, which are nested in countries.) The lowest level was in this case the respondents, while the highest level was the study. As EQ-5D valuation studies are based on nationally representative samples, we assume that studies represent countries.

To establish baseline variations of Δu for countries, we started with an empty model, which means that only the dependent variable and a country-specific intercept were included in the model. This model is presented in equation (1).

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \varepsilon_{ik}. \qquad (1)$$

Δu_{ik} represents the utility difference variable for each respondent i in country k. β_0 is the fixed intercept, which can be interpreted as the average Δu across countries. γ_{0k} is the random intercept for country variation for country k. If γ_{0k} is significant, the average Δu varies significantly between countries. ε_{ik} is the residual variation term at the respondent level. We assume that γ_{0k} is distributed as $\gamma_{0k} \sim N(0, \sigma_{\gamma0}^2)$ and that ε_{ik} is distributed as $\varepsilon_{ik} \sim N(0, \sigma_\varepsilon^2)$.

Δm and sociodemographic variables were added subsequently, followed by random slopes for the sociodemographic variables. These models are presented in equations (2) to (4):

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \beta_1 \Delta m_{ik} + \varepsilon_{ik}. \qquad (2)$$

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \beta_1 \Delta m_{ik} + \sum_j \beta_{2j} Socdem_{ij} + \varepsilon_{ik}. \qquad (3)$$

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \beta_1 \Delta m_{ik} + \sum_j (\beta_{2j} + \gamma_{2kj}) Socdem_{ij} + \varepsilon_{ik}. \qquad (4)$$

$\beta_1 \Delta m_{ik}$ represents Δm and its coefficient, while $\sum_j \beta_{2j} Socdem_{ij}$ represents the j sociodemographic variables (age, sex, education, and EQ-5D self-description) and their respective coefficients. $\beta_1 \Delta m_{ik}$ and $\beta_{2j} Socdem_{ij}$ are both fixed effects, which means that they can be interpreted as the average slope across countries for Δm and the average slope across countries for the sociodemographic variables on Δu. In other words, these effects are the same for all countries. In model 4, random slopes for sociodemographic variables are added. If the random-effects parameter γ_{2kj} is significant, this means that the slopes of the sociodemographic variables j on Δu vary between countries. We assume that γ_{2kj} is distributed as $\gamma_{2kj} \sim N(0, \sigma_{\gamma2j}^2)$.

Last, a dummy indicating whether a study is 3L or 5L was added, followed by the cultural variables. These models are presented in equations (5) and (6).

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \beta_1 \Delta m_{ik} + \sum_j (\beta_{2j} + \gamma_{2kj}) Socdem_{ij} + \beta_3 Fivelevel_k + \varepsilon_{ik}. \qquad (5)$$

$$\Delta u_{ik} = \beta_0 + \gamma_{0k} + \beta_1 \Delta m_{ik} + \sum_j (\beta_{2j} + \gamma_{2kj}) Socdem_{ij} + \beta_3 Fivelevel_k + \sum_l \beta_{4l} Cult_{kl} + \varepsilon_{ik}. \qquad (6)$$

$Fivelevel_k$ is the dummy variable indicating that a study is 3L (0) or 5L (1), with its respective coefficient β_3. $\sum_l \beta_{4l} Cult_{kl}$ represents both cultural dimensions and their coefficients ($l = 2$). $\beta_{4l} Cult_{kl}$ and $\beta_3 Fivelevel_k$ are fixed effects and can be interpreted as the average slopes across countries for these variables on Δu.

Assessing Explained Variation between Countries with Intraclass Correlation

We were interested in the variation of Δu between countries and whether this variation was reduced when correcting for sociodemographic background, methodological factors, and cultural values. Intraclass correlation coefficients (ICCs) are suited to assess the systematic variation between countries and served as the main outcome variable of this study. The ICC measures the variation $\sigma_{\gamma0}^2$ in Δu between countries relative to the total variation, the latter being the sum of country variation $\sigma_{\gamma0}^2$ and respondent variation σ_ε^2; see equation (7). For instance, if the ICC decreased by adding a new variable while the residual variation remained constant, the variation between countries was reduced by adding that variable. (Country-level variables cannot affect the residual variance, which is at the respondent level, since their value is the same for each respondent within that country. However, respondent level variables can affect both the country and the residual variance.) This indicates that the added variable explains differences between countries. In general, the ICC can take values between 0 and 1.

$$ICC = \frac{\sigma_{\gamma0}^2}{\sigma_{\gamma0}^2 + \sigma_\varepsilon^2} \text{ for models 1 to 3.} \qquad (7)$$

Models 4 to 6 included variation from random slopes for the sociodemographic variables, while models 1 to 3 did not. Therefore, the variability of the random slopes y_{2j} had to be included to calculate the ICC in models 4 to 6, as shown in equation (8).[32] (The ICC for our random slopes models is defined as
ICC =

$$\frac{\sigma_{\gamma0}^2 + \overline{Socdem_j} T_{Socdem} \overline{Socdem_j}' + 2\overline{Socdem_j}^* Cov(\gamma_{2j}, \gamma_0) + trace(T_{Socdem} S_{Socdem})}{\sigma_{\gamma0}^2 + \overline{Socdem_j} T_{Socdem} \overline{Socdem_j}' + 2\overline{Socdem_j}^* Cov(\gamma_{2j}, \gamma_0) + trace(T_{Socdem} S_{Socdem}) + \sigma_\varepsilon^2}$$

where $\overline{Socdem_j}$ is the vector of means of each

sociodemographic variable j that has a random slope and $\overline{Socdem_j}'$ is its transpose. S_{Socdem} is the covariance matrix of the sociodemographic variables that have random slopes, and T_{Socdem} is the covariance matrix of the random slopes themselves. Since we have standardized age to have mean zero and unit variance, and age is the only variable that is included with a random slope, this expression reduces to equation (8), as T_{Socdem} reduces to $\sigma_{\gamma 2j}^2$ and Σ_{Socdem} reduces to 1, while $\overline{Socdem_j}$ equals 0.) In equations (7) and (8), $\sigma_{\gamma 0}^2$ is the country variation, $\sigma_{\gamma 2j}^2$ is the variation for sociodemographic variables, and σ_ε^2 the residual variation at the respondent level.

$$\text{ICC} = \frac{\sigma_{\gamma 0}^2 + \sigma_{\gamma 2j}^2}{\sigma_{\gamma 0}^2 + \sigma_{\gamma 2j}^2 + \sigma_\varepsilon^2} \quad \text{for models 4 to 6.} \tag{8}$$

Additional analyses included a jackknife analysis to assess whether a country was considered an influential point. If a country was influential, associations found for Δu may not be representative for the remainder of the sample. To detect influential points, the model of equation (6) was constructed using the original sample, each time excluding another country from the original sample of countries. If the ICC was different for the subsample, the country was considered for exclusion. Furthermore, an analysis using Hofstede's cultural dimensions in model 6 was performed to compare the results with those found in the literature. Last, 4 additional analyses were performed with stricter or less strict definitions of mild and severe states to test the robustness of our definition of mild and severe states. An example of such a definition would be to define mild 3L states as having at most two 2s, compared to having at most three 2s in our current definition.

Results

Data Collection and Descriptive Statistics

The collection of data sets started in January 2016 and ended in August 2017. Forty-four studies were initially identified as currently completed or ongoing EQ-5D valuation studies. PIs were contacted through email and at the EuroQol meetings. Out of these 44 studies, 4 had not collected TTO data, 3 studies were not published yet, and 6 PIs were difficult to contact, leaving 31 studies. One of the data sets could not be shared with us for contractual reasons, leaving 30 studies. Data of 30 studies were obtained, of which 19 were EQ-5D-3L data sets and 11 were EQ-5D-5L data sets.[9,31,33–59] Two studies were excluded, as sociodemographic data or cultural values (in the case of the United Arab Emirates) were not available. Therefore, 28 studies remained. The jackknife

analysis did not identify any influential points. Thus, 28 countries remained for the final analysis. For 21 of these studies, information about educational level was also available, while in 26 of these studies, the EQ-5D self-description was included.

In total, 690 respondents did not value both mild and severe states, making it impossible to calculate Δu, and were excluded. Of these exclusions, 592 came from the Brazilian study, which is a saturation study with a balanced incomplete block design. Half of the total amount of respondents in that study did not value at least 1 mild, 1 moderate, and 1 severe state.[33] The remaining 98 exclusions came from 12 different studies. In total, the remaining sample included about 29,140 respondents.

Table 1 provides information on the studies that were obtained and their characteristics. The scores on the 2 cultural dimensions for each country are shown in Figure 1 and show a wide spread of cultural values. Dotplots were computed to illustrate the variation in average Δu by country (Figure 2). Δu varied by country, as shown in Figure 2. The smallest Δu was about 0.4, which means that severe states were valued only 0.4 lower than mild states. The highest Δu was around 1.2. Correlations across countries between Δu, age, the cultural variables, and the 3L or 5L dummy are reported in Table 2.

Multilevel Models

Preliminary analyses that included education and a rescaled EQ-5D self-description showed that education was not significant. As 6 studies had no measure of education, education was excluded from analysis to avoid losing data. The EQ-5D self-description was a significant predictor of Δu but could not explain variation in Δu between studies. As 2 studies did not include EQ-5D self-description, self-description was also excluded from analysis.

The results from the multilevel analyses are reported in Table 3. The columns represent the 6 different models described in equations (1) to (6). The first 7 rows present the coefficients of the fixed intercept and fixed effects for the included variable: the βs for the sociodemographic, methodological, and cultural variables. The next 3 rows present the random-effect parameters and residual variation: the σ_γs and σ_εs. In the last row, the ICC is described.

ICC

In every model, the random intercepts for country variation were significant, indicated by the row "RE country" in Table 3. This indicated that Δu varied reliably between studies. In the second to last row of Table 3, the ICCs

Table 1 Obtained Studies and Their Characteristics[a]

Country	3L/5L	Year	No. of Respondents	HS	Mode of Administration	Elicitation Method
Spain	3L	1997	972	12	Interview	TTO
Germany	3L	1997	339	12	Interview	TTO
Great Britain	3L	1993	3378	12	Interview	TTO
Netherlands	3L	2003	298	17	Interview	TTO
Italy	3L	2012	439	17	Interview	TTO
Portugal	3L	2012	450	7	Interview	TTO
Poland	3L	2008	321	23	Interview	TTO
Singapore	3L	2013	455	10	Interview	TTO
Japan	3L	1998	543	17	Interview	TTO
Taiwan	3L	2007	741	13	Interview	TTO
Australia	3L	2011	417	12	Online	TTO
France	3L	2008	452	17	Interview	TTO
Thailand	3L	2007	1388	10	Interview	TTO
Denmark	3L	2000	1332	14	Interview	TTO
Brazil	3L	2012	1146	7	Interview	TTO
Argentina	3L	2004	611	13	Interview	TTO
Zimbabwe	3L	2000	2348	7	Interview	TTO
United States	3L	2002	4043	9	Interview	TTO
Slovenia	3L	2005	225	13	Interview	TTO
Spain	5L	2012	1000	11	Interview	cTTO
Canada	5L	2012	1230	10	Interview	cTTO
Uruguay	5L	2014	805	13	Interview	cTTO
Korea	5L	2013	1080	13	Interview	cTTO
Japan	5L	2013	1026	13	Interview	cTTO
United Arab Emirates	5L	2013	200	10	Interview	cTTO
China	5L	2011	1302	10	Interview	cTTO
Netherlands	5L	2012	983	11	Interview	cTTO
Singapore	5L	2016	1000	13	Interview	cTTO
Thailand	5L	2013	1263	13	Interview	cTTO
Indonesia	5L	2015	1054	10	Interview	cTTO

cTTO, composite time tradeoff; HS, amount of health states valued by each respondent; TTO, time tradeoff.
a. The mode of administration shows us whether interviewers were present for the TTO or cTTO task, and the elicitation method provides information on whether TTO or cTTO was used in the study.
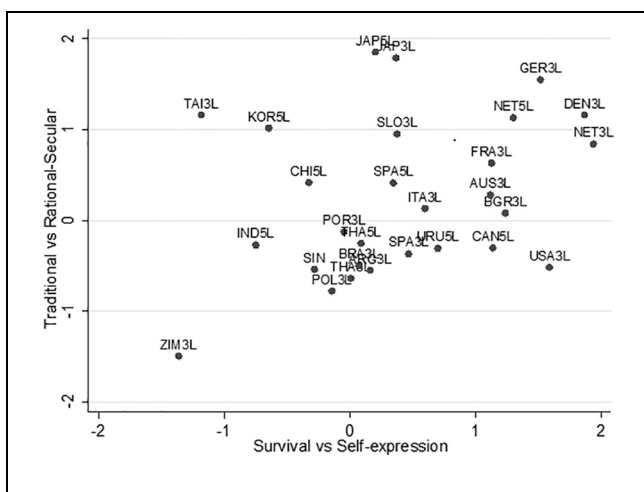

Figure 1 Scores on the 2 cultural dimensions, by country.

show the amount of variation attributed to country differences ($\sigma_{\gamma0}^2$) or to residual variation (σ_{ε}^2). The empty model (model 1) has an ICC of 13.1%, indicating that 13.1% of total variation (i.e., variation due to differences between countries and differences between respondents) is caused by country differences. Adding Δm and the sociodemographic variables (models 2–4) yielded a small reduction in ICC from 13.1% to 12.4%, caused by a slightly lower variation for the residual and a lower variation for country effects. Adding a random slope for age increased the ICC from 12.4% to 13.2%, and adding the 3L/5L dummy (model 5) increased the between-country variation and the ICC, from 13.2% to 14%. Adding the cultural variables (model 6) resulted in a further increase in ICC from 14% to 14.7%, caused by an increase in country-level variation. The respondent variation σ_{ε}^2 (Table 3, "Residual") remained stable because

Table 2 Correlations between Average Values per Country[a]

Variable 1	Variable 2	Correlation	95% Confidence Interval
Δu	Tradrat	−0.233	−0.563 to 0.161
Δu	Survself	−0.160	−0.509 to 0.235
Δu	Fivelevel	0.327	−0.060 to 0.629
Δu	Age	−0.119	−0.447 to 0.274
Tradrat	Survself	0.233	−0.161 to 0.563
Tradrat	Survself	0.099	−0.292 to 0.462
Tradrat	Survself	0.353	−0.031 to 0.646
Survself	Fivelevel	−0.260	−0.582 to 0.133
Survself	Age	0.418	0.046 to 0.689[b]
Fivelevel	Age	−0.227	−0.559 to 0.168

Survself, survival v. self-expression cultural variable; Tradrat, traditional v. rational-secular cultural variable.
a. One country was excluded, as it was identified as an outlier. Age was standardized before calculating these correlation coefficients.
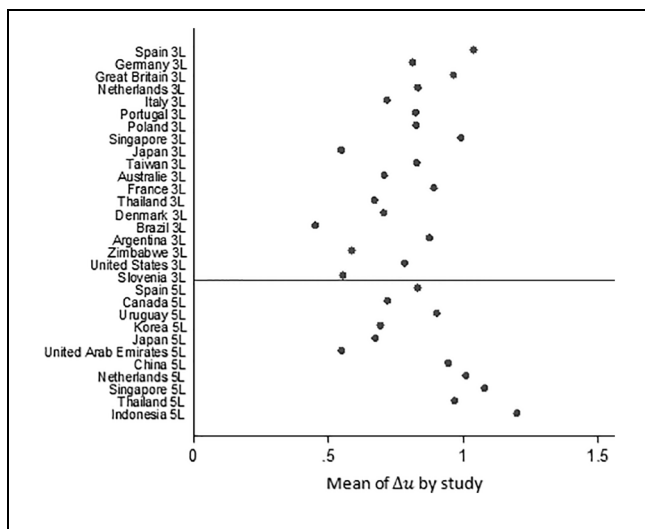b. Significant at the 5% level.

Figure 2 Dotplot of average Δu scores by country.

respondent variation cannot be affected by adding country-level variables.

Fixed Effects

The first 7 rows of Table 3 show that in model 3, the fixed effect of age was only weakly related to Δu, while the fixed effect of sex was not related to Δu. In addition, model 4 shows that the slope for age on Δu differed between studies, shown by "RE age," the variation in slope for age. The 3L/5L dummy, indicated by "Fivelevel" in models 5 and 6, was a significant negative predictor of Δu, implying that the utilities for severe states in 5L studies were raised by 0.118. The fixed

effects for the cultural dimensions of traditional/ rational-secular values and survival/self-expression, indicated by the "tradrat" and "survself" rows in model 6, were not related to Δu.

Additional Analyses

Additional analyses were performed using Hofstede's cultural dimensions.[22] The same model was used as in equation (6), now including Hofstede's 5 cultural dimensions instead of Inglehart's 2 cultural dimensions. The results showed that none of the 5 cultural dimensions of Hofstede was significantly related to Δu, and the ICC decreased slightly by 0.2%.

The analyses with different definitions for mild and severe states produced similar results. Although the ICC varied slightly for each model, depending on the definition of mild and severe states, the same pattern of reduction in country-level variation was found. The ICC did not decrease when the Fivelevel dummy was added or when the cultural variables were added to the models.

Discussion

Main Findings

We aimed to examine the effect of sociodemographic background, methodological factors, and cultural values on differences in health utilities, Δu, between countries. We did not find a relation between cultural values and Δu, as neither a linear relation could be found from the multilevel models, nor did cultural values explain variation in Δu between countries. Δm, the average difference in severity index between mild and severe states, was related to Δu, as were differences in using a 3L or a 5L

Table 3 Results from Multilevel Analyses for 27 Countries[a]

Variable/Analysis	1	2	3	4	5	6
Constant	0.825[b]	0.212[b]	0.202[b]	0.205[b]	0.245[b]	0.253[b]
Δm_i		0.978[b]	0.978[b]	0.974[b]	0.978[b]	0.978[b]
Age			0.0142[b]	0.004	0.004	0.004
Sex			0.006	0.006	0.006	0.006
Fivelevel					−0.118[c]	−0.118[c]
Tradrat						−0.023
Survself						−0.011
RE country	0.168[b]	0.161[b]	0.160[b]	0.162[b]	0.168[b]	0.173[b]
RE age				0.037[b]	0.037[b]	0.0366[b]
Residual	0.432[b]	0.426[b]	0.427[b]	0.426[b]	0.426[b]	0.426[b]
ICC, %	13.1	12.5	12.4	13.2	14.0	14.7

Fivelevel is a dummy variable indicating whether 0) 3L or 1) 5L was used. ICC, intraclass correlation; RE, random effect; Survself, survival v. self-expression cultural variable; Tradrat, traditional v. rational-secular cultural variable.

a. Country-level variables are written in italics. Residual indicates respondent-level variation. Both of these are presented as standard deviations. The ICC for each model was calculated using Table 2 and equations (7) and (8). For example, in model 1, only a random intercept for country variation was included. Therefore, the ICC equals $ICC = \frac{\sigma_{\gamma0}^2}{\sigma_{\gamma0}^2 + \sigma_\varepsilon^2} = \frac{0.168^2}{0.168^2 + 0.432^2} = 0.131$. This indicated that 13.1% of the total variation in Δu could be attributed to differences between countries.

b. Significant at the 1% level.

c. Significant at the 5% level.

protocol. Sociodemographic variables such as age and sex were only weakly related to Δu. Despite these findings, a large variation between countries remained.

Interpretation

Although cultural values were hypothesized to be related to variation in utilities for health states, we did not find a relation between cultural values and Δu.[5,6] The cultural variables were not significantly associated with Δu, health utility differences, and they did not decrease variation in Δu between countries. In addition, correlations between average Δu and the 2 cultural variables were nearly zero. Thus, we conclude that cultural values cannot account for differences in valuations between countries. Although we did not find a relation between cultural values and Δu, it was not unreasonable to hypothesize an association. Findings of previous studies by Augestad et al.[60] and Jakubczyk et al.[61] suggested a possible role for cultural values in explaining differences in TTO valuations. Jacubczyk et al.[61] showed that religious people assign higher utilities to health states in TTO valuations. Augestadt et al.[60] showed that attitudes toward euthanasia are related to TTO valuations. Religion and attitudes toward euthanasia are also related to our cultural values. For instance, the cultural dimension "traditional values" is related to a higher importance of religion. Also, the cultural dimension "survival values" is related to low tolerance for abortion, which is likely to be related to "prolife stances," entailing lower tolerance for euthanasia. Since both religion and "prolife stances" seem to be related to Inglehart's cultural dimensions,[23] cultural values are a promising candidate to explain utility differences between countries.

Our results are contrasted by Bailey and Kind,[24] who looked at the mean TTO value for 7 mild health states in 10 countries, correlated those with 5 Hofstede cultural dimensions for each country, and found a relation between Hofstede's dimensions and the TTO scores.[22] The relation found by Bailey and Kind was the strongest for Hofstede's "Power Distance" and "Uncertainty Avoidance" dimensions; these dimensions were also strongest in our own analysis of Hofstede's cultural dimensions. However, there are some differences between Bailey and Kind's study[24] and our study. Our study included more countries and considered respondent-level data, whereas Bailey and Kind[24] used mean TTO data for some specific health states and correlated those with Hofstede's cultural dimensions. Furthermore, our study is on differences in utilities, Δu, not on utilities given to specific health states themselves.

We found 2 predictors of Δu. Δm, the average difference in severity index between mild and severe states, corrects for the selection of health states, whose composition differed between studies and respondents. As expected, Δm was related to Δu; an increase of 1 in Δm would cause a 0.978 increase in Δu. Furthermore, after correcting for the selection of health states, differences between 3L and

5L studies remained; that is, Δu was smaller for 5L studies than for 3L studies. This implies that in 5L studies, values of severe states are raised by 0.118. One possible explanation could be an upward shift of the values in the cTTO task, which is used in 5L studies. This shift may arise for negative states in the cTTO, as the state to be valued is preceded by 10 years in good health, effectively changing and improving the state to be valued. Indeed, Xie et al.[62] found that severe states were valued higher in the cTTO task compared to the TTO task, with average differences as large as 0.213 for some health states. These observations corroborate our results.

Limitations and Strengths

This study has some limitations. First, our analyses were done on existing data, and also, desired data was not collected in all countries. Second, only a small number of sociodemographic variables could be considered for analysis, since more could not be analyzed while preserving a sufficient sample of included countries. Third, cultural data were collected at the national level instead of at the respondent level, which reduces the chance of finding a relation between cultural values and Δu. Fourth, the methods of data collection differ between 3L and 5L studies. To account for this, we collapsed methodological differences into a single variable: the 3L/5L dummy. Fifth, the valuation data were not cleaned, which might affect our findings, although for preference-based methods, removing inconsistent responses hardly affects valuations.[63,64] Sixth, we assume that EQ-5D valuation studies use a representative sample for their respective country. However, since designs and sample sizes differ between 225 and 4043 respondents, we cannot be sure about this, which is a limitation. Seventh, correlation between the independent variables made it harder to interpret the results, which is a limitation of the data. Last, moderate states were not included in our analyses, so not the whole spectrum of EQ-5D health states was analyzed.

First, a major strength of this study is its methodology; we have combined the largest number of EQ-5D valuation data sets to date. Second, our method of analysis takes into account the multilevel structure of the data. Third, our method of analysis is well suited to correct for disturbing variables. Lastly, the results of our study are robust as different definitions of mild and severe states produced similar results.

Practical Implications

Countries use their own EQ-5D tariffs for the calculation of QALYs in cost-utility analyses. This is reasonable, as our findings reveal a large amount of variation in Δu between countries. Some protocols aimed to collect valuation data in many countries with the aim to derive a common tariff, such as the BIOMED project, which generated a common VAS value set for European countries.[65] In a similar vein, some countries may rely on value sets from other countries for the calculation of QALYs in cost-utility analyses. As we found that utility differences differ strongly between countries, a multinational tariff or tariffs from a neighboring country would likely misrepresent the tariff of individual countries. This strengthens the case for national tariffs for instruments such as the EQ-5D.

Conclusion

Health utilities differ between countries, as shown, for example, by the varying amounts of health states worse than dead reported by EQ-5D valuation studies. The aim of this article was to assess these differences and to test whether these differences were related to the sociodemographic background of the respondents, methodological differences, and cultural values. Cultural values did not explain Δu variation between countries. Despite correction for various variables, differences in Δu between countries remain substantial.

ORCID iD

Bram Roudijk (iD) https://orcid.org/0000-0001-5000-0875

References

1. Froberg DG, Kane RL. Methodology for measuring health-state preferences –I: Measurement strategies. *Journal of Clinical Epidemiology*. 1989;42(4):345–54.
2. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ*. 2002;21(2):271–92.
3. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. York, the United Kingdom. 1995.
4. Torrance GW, et al. Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2. *Med Care*. 1996;34(7):702–22.
5. Norman R, et al. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health*. 2009;12(8):1194–200.

6. Wang P, et al. Do Chinese have similar health-state preferences? A comparison of mainland Chinese and Singaporean Chinese. *Eur J Health Econ*. 2015;16(8):857–63.

7. Olsen JA, Lamu AN, Cairns J. In search of a common currency: a comparison of seven EQ-5D-5L value sets. *Health Econ*. 2018;27(1):39–49.

8. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5 D-5 L value set for E ngland. *Health Economics*. 2018;27(1):7–22.

9. Purba FD, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics*. 2017;35(11):1153–65.

10. Al Sayah F, et al. Determinants of time trade-off valuations for EQ-5D-5L health states: data from the Canadian EQ-5D-5L valuation study. *Qual Life Res*. 2016;25(7):1679–85.

11. Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? *Soc Sci Med*. 2002;54(6):919–29.

12. Kind P, Dolan P. The effect of past and present illness experience on the valuations of health states. *Med Care*. 1995;33:AS255–63.

13. Dolan P. The effect of experience of illness on health state valuations. *J Clin Epidemiol*. 1996;49(5):551–64.

14. Jonker MF, et al. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Econ*. 2017;26(12):1534–47.

15. Froberg DG, Kane RL. Methodology for measuring health-state preferences—II: scaling methods. *J Clin Epidemiol*. 1989;42(5):459–71.

16. Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metrics*. 2003;1(1):12.

17. Shah KK, et al. One-to-one versus group setting for conducting computer-assisted TTO studies: findings from pilot studies in England and the Netherlands. *Eur J Health Econ*. 2013;14(1):65–73.

18. Stalmeier PF, et al. What should be reported in a methods section on utility assessment? *Med Decis Making*. 2001; 21(3):200–7.

19. Rokeach M. *The Nature of Human Values*. New York: Free Press; 1973.

20. Inglehart R. *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies*. Princeton, NJ: Princeton University Press; 1997.

21. Inglehart R, Baker WE. Modernization, cultural change, and the persistence of traditional values. *Am Sociol Rev*. 2000;65:19–51.

22. Hofstede G, Hofstede GJ, Minkov M. *Cultures and Organizations: Software of the Mind*. Vol. 2. Maidenhead, UK: McGraw-Hill; 1991.

23. Roudijk B, Donders R, Stalmeier P. Cultural values: can they explain self-reported health? *Qual Life Res*. 2017; 26(6):1531–9.

24. Bailey H, Kind P. Preliminary findings of an investigation into the relationship between national culture and EQ-5D value sets. *Qual Life Res*. 2010;19(8):1145–54.

25. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.

26. Herdman M, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.

27. Janssen BM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(1):5–13.

28. Dolan P, et al. The time trade-off method: results from a general population study. *Health Econ*. 1996;5(2):141–54.

29. Oppe M, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–53.

30. Inglehart R, Welzel C. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge, UK: Cambridge University Press; 2005.

31. Luo N, et al. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. *Pharmacoeconomics*. 2014;32(5): 495–507.

32. Snijders T, Bosker R. *Multilevel Analysis: An Introduction to Basic and Applied Multilevel Analysis*. London, UK: Sage; 1999.

33. Andrade MV, et al. Societal preferences for EQ-5D health states from a Brazilian population survey. *Value Health Regional Issues*. 2013;2(3):405–12.

34. Augustovski F, et al. An EQ-5D-5L value set based on Uruguayan population preferences. *Qual Life Res*. 2016;25(2):323–33.

35. Augustovski FA, et al. Argentine valuation of the EQ-5D health states. *Value Health*. 2009;12(4):587–96.

36. Badia X, et al. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making*. 2001;21(1):7–16.

37. Chevalier J, de Pouvourville G. Valuing EQ-5D using time trade-off in France. *Eur J Health Econ*. 2013;14(1):57–66.

38. Claes C, et al. An interview-based comparison of the TTO and VAS values given to EuroQol states of health by the general German population. In: *Proceedings of the 15th Plenary Meeting of the EuroQol Group*. Hannover, Germany: Centre for Health Economics and Health Systems Research, University of Hannover; 1999.

39. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–108.

40. Ferreira LN, et al. The valuation of the EQ-5D in Portugal. *Qual Life Res*. 2014;23(2):413–23.

41. Golicki D, et al. Valuation of EQ-5D health states in Poland: first TTO-based social value set in Central and Eastern Europe. *Value Health*. 2010;13(2):289–97.

42. Ikeda S, et al. Developing a Japanese version of the EQ-5D-5L value set. *J Natl Inst Public Health*. 2015;64(1): 47–55.

43. Jelsma J, et al. How do Zimbabweans value health states? *Popul Health Metrics*. 2003;1(1):11.

44. Kim S-H, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res*. 2016;25(7):1845–52.

45. Lamers LM, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ*. 2006;15(10):1121–32.
46. Lee H-Y, et al. Estimating quality weights for EQ-5D (EuroQol-5 dimensions) health states with the time trade-off method in Taiwan. *J Formosan Med Assoc*. 2013; 112(11):699–706.
47. Luo N, et al. Estimating an EQ-5D-5L value set for China. *Value Health*. 2017;20(4):662–9.
48. Papadimitropoulos EA, et al. An investigation of the feasibility and cultural appropriateness of stated preference methods to generate health state values in the United Arab Emirates. *Value Health Regional Issues*. 2015;7:34–41.
49. Pattanaphesaj J, et al. *Health-Related Quality of Life Measure (EQ-5D-5L): Measurement Property Testing and Its Preference-Based Score in Thai Population*. Salaya, Thailand: Mahidol University; 2014.
50. Rupel VP, Rebolj M. The Slovenian VAS tariff based on valuations of EQ-5D health states from the general population. In: *Discussion Papers/17th Plenary Meeting of the Euroqol Group*. Tudela, Spain: Universidad Pública de Navarra; 2001.
51. Scalone L, et al. Italian population-based values of EQ-5D health states. *Value Health*. 2013;16(5):814–22.
52. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–20.
53. Tongsiri S, Cairns J. Estimating population-based values for EQ-5D health states in Thailand. *Value Health*. 2011;14(8):1142–5.
54. Tsuchiya A, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ*. 2002;11(4):341–53.
55. Versteegh MM, et al. Dutch tariff for the five-level version of EQ-5D. *Value Health*. 2016;19(4):343–52.
56. Viney R, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health*. 2011;14(6):928–36.
57. Wittrup-Jensen KU, et al. Generation of a Danish TTO value set for EQ-5D health states. *Scand J Public Health*. 2009;37(5):459–66.
58. Xie F, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98.
59. Gandhi M, et al. Sample size determination for EQ-5D-5L value set studies. *Qual Life Res*. 2017;26(12):3365–76.
60. Augestad LA, et al. Time trade-off and attitudes toward euthanasia: implications of using 'death' as an anchor in health state valuation. *Qual Life Res*. 2013;22(4):705–14.
61. Jakubczyk M, Golicki D, Niewada M. The impact of a belief in life after death on health-state preferences: true difference or artifact? *Qual Life Res*. 2016;25(12):2997–3008.
62. Xie F, et al. How different are composite and traditional TTO valuations of severe EQ-5D-5L states? *Qual Life Res*. 2016;25(8):2101–8.
63. Lamers LM, et al. Inconsistencies in TTO and VAS values for EQ-5D health states. *Med Decis Making*. 2006;26(2): 173–81.
64. Torrance GW, et al. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care*. 1996;34(7):702–22.
65. Brooks R, Rabin R, De Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQol BIOMED Research Programme*. New York: Springer Science & Business Media; 2013.