

Valuing EQ-5D using Time Trade-Off in France

Julie Chevalier · Gérard de Pouvourville

Received: 10 May 2011 / Accepted: 5 September 2011 / Published online: 21 September 2011
© Springer-Verlag 2011

Abstract

Objectives While a French language version of the EQ-5D exists, to date, there has been no French value set to accompany it. The objective of our study was then to derive the French TTO value set of the EQ-5D.

Methods A total of 452 respondents aged over 18 were recruited who were representative of the French population with regard to age, gender, and socio-professional group. The direct valuation of 24 health states was first obtained by Time Trade-Off (TTO), and the negative TTO values were bounded using the monotonic transformation. Several alternative model specifications were investigated to estimate the values for all 243 states in the EQ-5D descriptive system. Only the best fitting model is presented in this paper. The analysis was conducted at an individual level to make the maximum use of the available data, and we estimated mixed models with random intercept. Models were compared through the Akaike information criterion (AIC), the mean absolute error (MAE), and the Pearson correlation coefficient between the observed and the predicted values of each model.

Results After exclusion, 443 respondents took part in the study. The best fitting model included the same variables as the N3-model used in UK.

Conclusion This study provides the French value set of the EQ-5D based on the stated preferences of the French general public facilitating cost-effectiveness analysis.

Keywords Health-related quality of life · EQ-5D · Utility · Preferences

JEL Classification I19

Introduction

Quality-adjusted life years (QALYs) gained have become the most frequently used outcome measures in cost-effectiveness analysis. This indicator measures both quantity and quality of life: the duration of a given health state is weighted with a coefficient calculated based on its associated quality of life. This weight is based on the individual preferences for each possible health states. It reflects the relative desirability of the health states and is measured on an interval scale where 1 represents “Full Health” and 0 represents “dead.” Some severe health states may be considered as being worse than being dead and take a negative value.

Some preference-based measures of health status have been developed to facilitate the use of utilities in economical analyses, as for example, the EQ-5D [1], the Health Utility Index (HUI) [2], and the SF-6D [3]. They describe a number of health states through a generic descriptive system (questionnaire), and a value is associated to each health state. These instruments differ mainly in the dimensions used to describe the health states (the descriptive system). Our paper focuses on the EQ-5D, which is one of the most widely used health states descriptive system and its valuation in France.

An earlier version of this manuscript was presented at the 26th Plenary Meeting of the EuroQol Group, Paris, France, September, 3–5, 2009.

J. Chevalier (✉) · G. de Pouvourville
Chair of Health Economics and Management,
ESSEC Business School, Avenue Bernard Hirsch,
BP 50105, 95021 Cergy-Pontoise Cedex, France
e-mail: chevalier@essec.fr

France has always stood some what apart from research work on the valuation of health states through stated preference methods. A first pioneering work was published [4] on the development of a French value set for the HUI 3, but to our knowledge was not currently used for applications. In the academic community, there was and still is a fierce controversy about the QALY paradigm and its application for decisions related to the adoption of innovative technologies [5]. The French pricing and reimbursement process still leaves today little space for cost-effectiveness analysis. It is strongly egalitarian: the system should warrant that whatever the disease they suffer from, patients will have equal chances of access to innovation.

Because of that, the academic community did not take a chance to learn and practice methods to explicit and measure patients' preferences, which at least should yield important information for decision makers, who are in principle mandated to reflect the population's perspective.

But, due to the high prices of new treatment, there is growing concern voiced mainly by members of parliament and the National Accounting Court (Cour des comptes) that the opportunity cost of new treatments should at least be considered, if not used similarly to the NICE. In a first draft of recommendations for the use of economic assessment in P&R made public in December 2010, the Commission for Economic and Public Health Assessment, (CEESP), from the French HTA body, the HAS, did encourage to use QALYs as an outcome measure when the technology at study had a significant impact on longevity and quality of life [6].

In a country context where people and institutions give great value to national specificities, we assumed that a locally developed value set was needed to warrant acceptance in France. Moreover, there are good a priori reasons to think that stated preferences may differ from one country to another. There is abundant evidence provided by sociologists and anthropologists on cultural differences in the perception of diseases and suffering, see for example [7]. Second, differences in the availability of services among countries can also lead to different valuation of specific dimensions. For example, the abundant provision of affordable services for disabled people in one country compared to another may lead respondents to give less importance to a problem in mobility in the first country than in the other. The hypothesis of differences between stated preferences across countries is reinforced by the results of the different national value sets and their comparisons [8–11]. In this paper, we proceed to a graphic comparison of the French value set with the initial MVH 1993 set.

Table 1 Health states set assignment

Set 1 Macran and Kind <i>n</i> = 150 respondents	Set 2 <i>n</i> = 150 respondents	Set 3 <i>n</i> = 150 respondents
<i>Group 1</i>	<i>Group 1</i>	<i>Group 2</i>
21111*	21111*	11211*
12111*	12111*	11121*
13311*	13311*	32211*
11113*	11113*	11112*
11131*	11131*	11312*
22222*	22222*	11133*
23232*	23232*	32223*
32313*	32313*	33323*
<i>Group 2</i>	<i>Group 3</i>	<i>Group 3</i>
11211*	22121	22121
11121*	21323	21323
32211*	22122	22122
11112*	22233	22233
11312*	33321	33321
11133*	13332	13332
32223*	23313	23313
33323*	33232	33232
<i>Additional health states</i>	<i>Additional health states</i>	<i>Additional health states</i>
33333*	33333*	33333*
11111*	11111*	11111*
Dead*	Dead*	Dead*

* Health states forming a part of the 17 states included in Macran and Kind

In this latter study, investigators collected the direct valuation of 42 from 243 EQ-5D health states using the Time Trade-Off (TTO) method [12]. Regression methods were used to estimate values for the whole health states (including the 42, which were directly valued). The estimated parameters of the regression model define an equation that gives a weight (or score) at each health state taking into account the public preferences. The resulting set of weights or tariff (the MVH A1 tariff) [12] is widely used to calculate utilities in cost-utility analyses of health care programs and treatments.

Lamers et al. [10] examined the effect of a reduction in the number of observations, both in terms of health states directly valued and respondents, on the estimated tariff. They concluded that using the Macran and Kind set of 17 health states [13] and a sample size of 200 respondents was appropriate for national EQ-5D valuation studies. The Macran and Kind set is a subset of the one used in the MVH study (cf. Table 1) and was used for the estimation of the Japanese [14] and Dutch [10] value sets.

Method

Sampling and selection of health states

Sampling

The sample size calculation was based on the following formula: $n = \frac{Z_{1-\frac{\alpha}{2}}^2 \sigma^2}{\delta^2}$, where $Z_{1-\frac{\alpha}{2}}^2$ is the percentile of the normal distribution used as the critical value in a two-tailed test of size α ($Z_{1-\frac{\alpha}{2}}^2 = 1.96$ for a 0.05 level test).

δ is the tolerated margin of error, i.e., the maximum allowable difference between the sample mean and the true mean in the French population. We chose $\delta = 0.05$.

σ is an estimate of the population standard deviation. We chose $\sigma = 0.4$, which is the mean standard deviation observed in studies valuing EQ-5D health states using TTO.

This formula provides the number of respondents necessary to obtain an estimation of the mean with a 95% ((100 - α)%) probability that the true mean falls in the interval: [observed mean \pm δ].

Three hundred respondents were then necessary to value each health state. This figure, although bigger, is accordant to the one recommended by Lamers et al. [10].

Members of the general public aged over 18 were recruited with national coverage by a marketing research company. They were selected to be representative of the French population with regard to age, gender, and socio-professional group. Thirty interviewers were trained by the researchers (JC, GdP) and performed the face-to-face interviews at the respondents' home during the month of December 2008. Respondents received a gift voucher worth €15 for participation.

Selection of health states

One of the changes made to the MVH protocol used in the United Kingdom [12] concerned the number and selection of the subset of EQ-5D states to be valued. First, unlike other valuation studies, the health state "Unconscious" was not valued in the French study as the TTO method is not adapted to its valuation. Then, only 24 health states were selected to be directly valued. This figure is in the range of proposed numbers in the MVH study (42) and by Lamers et al. (17) [10].

As respondents could not be expected to value all 24 health states using the TTO in a single interview, only 17 health states were valued by each respondent [12].

To allow the comparison with other valuation studies, we first chose the same 17 health states as Macran and Kind [13]. These states are presented with a "*" in Table 1. We completed the sets of states with 7 health

states randomly selected from the 42 health states of the MVH study.

The 24 health states were divided into three groups of 8 health states (Group 1, 2, and 3). Three sets of health states were then constituted with 2 groups each (Set 1 = Group 1 + Group 2, Set 2 = Group 1 + Group 3, Set 3 = Group 2 + Group 3), 33333, 11111, and "Dead." For example, people selected to value the second set of health states (set 2) had to value the health states of the group 1, the ones of the Group 2, the health states 33333, 11111, and "Dead." One set of health states (Set 1) contains all 17 of the states used by Macran and Kind. Table 1 also presents these different sets of health states. One hundred and fifty respondents were selected to value each sets of states leading to a sample size of $N = 450$. Each health state was then valued by at least 300 respondents.

Interview procedure

The face-to-face interview consisted of the following stages:

- (1) Self-reported health on the five dimensions descriptive system (EQ-5D questionnaire).
- (2) Self-reported health:
 - On a visual analog scale (EQ-VAS) for half of the respondents.
 - Using a new experimental scoring method developed by Parkin et al. [15] for the others.
- (3) Valuation of hypothetical health states (warm-up exercise):
 - Ranking and valuation of 19 (17 + 11111 + Dead) health states using a Visual Analog Scale (VAS) for half of the respondents.
 - Valuation of 17 health states using a new experimental scoring method [15] for the others.
- (4) TTO valuation of 17 hypothetical health states.
- (5) Socioeconomic background questions.

The whole questionnaire was replicated on the computer's screen of the interviewer.

We took the opportunity of this large national study to test some properties of the experimental scoring method. Its purpose was also, as with the VAS, to be used as a warm-up exercise aiming to familiarize the respondents with the presentation of the different health states.

Each health state was presented to the respondent on separate cards. The different sets of health states presented in Table 1 were alternatively used by the interviewers. By design, Dead and 11111 were not required to be valued in the scoring method and in the TTO method. Health states cards were shuffled before the TTO exercise and presented in a random order to the respondents.

During the warm-up exercise (EQ-VAS or scoring), respondents were told that each health state was to be considered as lasting 10 years without change, followed by death.

The TTO exercise used the same visual prop as used in United Kingdom. This prop is a two-sided board allowing for both positive and negative values for the health states, respectively, considered as better than and worse than death. Through an iterative procedure, the respondents were asked to select a length of time (t) in Full Health (11111) that they considered equivalent to 10 years in the targeted health states. The shorter the length of time t is, the worse the targeted health state is. Respondents were allowed to trade time in month and weeks instead of years when they are still indifferent between 9 years in full health and 10 years in the targeted health states. For health states considered as being worse than death, the trade-off is between “death” and spending a length of time ($10 - t$) in the targeted health state followed by t years in Full Health (11111). The longer the length of time t is, the worse the targeted health state is.

This paper focuses on the 4th element of the interview, the TTO valuation of the EQ-5D. Results of the scoring method and VAS are reported in a separate paper.

Statistical analyses

Exclusion criteria of data

Questionnaires from respondents considered to have misunderstood the task were removed. These respondents were identified according to the following exclusion criteria that have been used in other similar studies [16]:

- TTO data missing for all health states.
- Only 1 or 2 states valued.
- All states given the same value.
- All states valued as worse than dead.

A health state H1 with one higher level in at least one dimension, and with no dimensions at lower levels, compared to a health state H2, will be considered as inconsistent if its utility value is greater than that of H2. For example, the utility value of 11123 is defined as inconsistent if it is greater than the utility value of 11122 because the latter is a logically better health state. For each respondent, the number of inconsistencies in their responses was examined as a data quality measure but was not an *ex ante* exclusion criterion.

Transformation of health states

For states better than death, the TTO utility value is $v = \frac{t}{10}$ where t represents the number of years in full health.

As in the most of other countries having computing their own value set, values for states worse than death were calculated by $v = \frac{-t}{10-t}$. The lowest possible value is -39 . This value occurs when the respondent prefers immediate death to 3 months in the targeted health state followed by 9.75 years in 11111. Following the convention adopted by other studies using the MVH protocol, we chose to transform negative value using a monotonic transformation [17]: $v' = \frac{v}{1-v} = \frac{-t}{10}$ so that they were bounded by a maximum negative value of -1 .

Regression analysis

Regression analyses were used to interpolate TTO values for the whole health states (included the 24, which were directly valued). Several model specifications were investigated including different sets of variables. We tried a number of specifications as for example the Dolan N3-model [12] or the D1 model used by the US researchers [18]. The different model specifications were compared through the Akaike information criterion (AIC), the mean absolute error (MAE), and the Pearson correlation coefficient between the observed and the predicted values of each model. Only the best fitting model is presented in this paper.

Variables

For each respondent and each health state, the dependent variable is 1 minus the TTO score given to that health state. It represents the loss of utility associated with the health state.

The model that was selected to estimate the French value set used the same variables as the ones used in the MVH model (the N3-model). The shifts between levels within a dimension were modeled using dummy variables: one measuring the difference between level 1 and level 2 and another measuring the difference between level 1 and level 3. Two dummy variables were then generated for each dimension: mobility (MO2, MO3), self-care (SC2, SC3), usual activities (UA2, UA3), pain/discomfort (PD2, PD3), and anxiety/depression (AD2, AD3). An eleventh dummy variable (N3) was introduced to control for the severity of the health states. It indicates if at least one dimension is scored at level 3. As the aim of the study was to estimate one preference-based EuroQol tariff for the whole French population, respondents' characteristics such as age or sex were not included in the model. An overview of the independent variables used in our model is given in Table 4. The parameter estimates were considered significant at an alpha level of 5%.

Model specification

The analysis was conducted at an individual level to make the maximum use of the available data. Since each respondent valued several states, it was expected that a relationship existed between responses. For example, a respondent offering higher or lower value than the average for a particular health state is likely to do that consistently across health states. The variance of the error term in the model would be partially determined by each respondent, which violates one of the key assumptions of ordinary least square (OLS) regression. A random effect (RE) model was used as an estimation method to address this problem.

It was specified as

$$y_{ij} = \beta_1 + \beta_2 x_{2,ij} + \dots + \beta_k x_{k,ij} + \varepsilon_{ij}$$

$$\beta_1 = \alpha_1 + u_{1i}$$

where α_1 is an overall intercept, ε_{ij} is the traditional error term that represents the deviation between the observed value of the state j for the respondent i and the predicted one, and u_i is an error term representing the deviation between the intercept for the i th respondent and the overall intercept. A random term could be applied to any of the parameters by defining $\beta_k = \alpha_k + u_{ki}$. Following Dolan [12] and others, we estimated a mixed model with random intercept.

Statistical analyses were performed using SAS software.

Results

Characteristics of the sample

In total, 452 respondents were participated in the survey. Nine (9) were excluded since they met at least one exclusion criteria, 8 gave all states the same value, and 1 valued all states worse than death. Characteristics of the sample after exclusions are presented in Table 2.

The mean number of inconsistencies among the full sample, as defined in the method section, is 5.2 (SD: 4.9). There is no statistical difference in this number according to the set of states valued. Ninety percent of the sample exhibits at least one logically inconsistent valuation. Ten percent of the respondents had more than 10 logical inconsistencies but were not excluded from the analysis as inconsistencies did not constitute an *ex ante* exclusion criterion.

The analyzed sample was the representative of the French population in terms of sex, age, and socio-professional group. The majority of problems reported in the EQ-5D descriptive system were pain/discomfort (49%) and anxiety/depression (36%). The mean EQ-VAS health state score was 77 (SD 18.2).

Direct valuation of health states

Each respondent was asked to value 17 health states using the TTO procedure. After the transformation to the lower bound of -1 , the mean values for the 24 health states directly valued ranged from 0.88 for state 11121 to -0.50 for state 33333; the median values from 0.99 to -0.62 for states 11112 and 33333, respectively (see Table 3).

Regression analysis: results

Parameter estimates of the model are presented in Table 4. This model presented the best fit statistics, with all parameters being statistically significant and was then selected to compute the French societal tariff for the EQ-5D. Estimates for dummies representing the difference between level 1 and level 2 (MO2, SC2, UA2, PD2, AD2) are lower than ones representing the difference between level 1 and level 3. The respondents give the greatest importance to the “Mobility” and “Self-Care” dimensions: the disutility of being in level 3 in one of these two dimensions is greater than the disutility of being in any level of the other dimension.

The analysis on only the 17 health states from Macran and Kind (instead of the 24—17 + 7—health states included in our study) resulted in the selection of the same model. No major difference (greater than 0.05) was reported in the predicted values of the 243 health states when the value set is estimated from a subgroup of 17 health states instead of the 24 health states.

Final valuation model

The utility value of each health state can be computed using the following formula:

$$\begin{aligned} U(\text{Hi}) = & 1 - 0.155 \text{MO2} - 0.372 \text{MO3} - 0.212 \text{SC2} \\ & - 0.326 \text{SC3} - 0.156 \text{UA2} - 0.189 \text{UA3} \\ & - 0.112 \text{PD2} - 0.265 \text{PD3} - 0.090 \text{AD2} \\ & - 0.204 \text{AD3} - 0.174 \text{N3} \end{aligned}$$

For example, the utility value associated with the health state 11233 will be

$$\begin{aligned} U(11233) = & 1 - (0.155 * 0) - (0.372 * 0) - (0.212 * 0) \\ & - (0.326 * 0) - (0.156 * 1) - (0.189 * 0) \\ & - (0.112 * 0) - (0.265 * 1) - (0.090 * 0) \\ & - (0.204 * 1) - (0.174 * 1) \\ = & 1 - 0.156 - 0.265 - 0.204 - 0.174 \\ = & 0.201 \end{aligned}$$

Table 2 Characteristics of the sample

	French general population*	Sample (n = 443)
Women (%)	51.6	51.7
Years (%)		
18–24	11.5	12.0
25–34	16.3	16.5
35–44	18.1	19.2
45–54	17.4	16.9
55–64	15.4	18.1
65–74	10.2	12.6
>75	11.1	4.7
Mean age (std dev)		46.1
Socio-professional group (SPG) (The unemployed having already worked are classified according to their last trade)		
Craftsman trading and heads of undertakings	5.8	5.9
Executive and intellectual professions	11.5	12.9
Intermediate occupations	18.3	16.7
Employees	26.1	27.1
Working	21.3	21.2
Farmer owners	3.0	0.7
People without community activity	14.0	15.6
Educational level (%)		
For respondents aged 25–65 (n = 309)		
Low	56.8	51.8
Middle	28.4	29.8
High	14.0	18.5
In couple (%)		
For respondents aged up to 25 (n = 389)	68.1	67.1
EQ-5D		
Respondents reporting problems on (%)		
Mobility	Not available	13
Self-care	Not available	2
Usual activities	Not available	12
Pain/discomfort	Not available	49
Anxiety/depression	Not available	36
EQ-VAS own health (n = 216)		
Mean (SD)		77.0 (18.2)
Respondents (%)		
Considers himself religious (n = 430)	Not available	53.9
Believes in life after death (n = 386)	Not available	38.6
Perceived health (%)		
Excellent/(very) good	Not available	87
Fair/poor	Not available	13

* Source French National Institute for Statistics and Economic Studies. Situation in 2008

Comparison with the MVH A1 tariff

Figure 1 presents a comparison between the French observed TTO values and the MVH ones [12]. It highlights some major differences in the intermediate health states' scores, while the extreme health states (i.e., closer to full

health and 33333) are valued in the same way in both countries.

Some major differences (greater or smaller score) are observed on the value set (see Figs. 1 and 2), especially for the intermediate health states in terms of severity. Table 4 shows the results of the French model and the UK

Table 3 Mean, median, and standard deviation for observed values (after exclusion) and percentage of negative values per state

EQ-5D health state	<i>N</i>	Mean	Median	STD	Negative values (%)
11121	296	0.88	0.95	0.26	1.3
11112	296	0.88	0.99	0.26	1.3
11211	295	0.86	0.93	0.23	0.3
21111	293	0.82	0.93	0.35	3.7
12111	292	0.78	0.93	0.39	3.4
11312	295	0.63	0.80	0.43	7.8
11113	289	0.58	0.83	0.54	11.1
22121	291	0.56	0.70	0.50	12.0
11131	289	0.47	0.70	0.59	19.0
22122	291	0.45	0.63	0.54	16.1
11133	294	0.38	0.53	0.58	20.7
13311	291	0.36	0.50	0.57	19.6
21323	292	0.19	0.34	0.61	31.5
22222	286	0.18	0.38	0.65	32.5
32211	292	0.08	0.20	0.64	40.1
13332	288	-0.10	-0.18	0.63	52.1
23313	286	-0.11	-0.03	0.61	50.3
32223	293	-0.17	-0.30	0.60	57.7
23232	286	-0.19	-0.31	0.60	57.3
22233	284	-0.19	-0.30	0.61	57.4
32313	290	-0.23	-0.38	0.63	61.7
33321	283	-0.25	-0.38	0.58	62.2
33232	290	-0.35	-0.50	0.55	72.8
33323	290	-0.39	-0.50	0.54	75.9
33333	430	-0.50	-0.63	0.50	81.9

Table 4 Comparison of the French and the MVH estimated coefficients

		MVH	France
	Constant	0.081	
MO2	1 if mobility is level 2; 0 otherwise	0.069	0.155
MO3	1 if mobility is level 3; 0 otherwise	0.314	0.372
SC2	1 if self-care is level 2; 0 otherwise	0.104	0.212
SC3	1 if self-care is level 3; 0 otherwise	0.214	0.326
UA2	1 if usual activities is level 2; 0 otherwise	0.036	0.156
UA3	1 if usual activities is level 3; 0 otherwise	0.094	0.189
PD2	1 if pain/discomfort is level 2; 0 otherwise	0.123	0.112
PD3	1 if pain/discomfort is level 3; 0 otherwise	0.386	0.265
AD2	1 if anxiety/depression is level 2; 0 otherwise	0.071	0.090
AD3	1 if anxiety/depression is level 3; 0 otherwise	0.236	0.204
N3	1 if any dimension is at level 3; 0 otherwise	0.269	0.174
AIC			9,484
MAE			0.043
Pearson <i>r</i>			0.993

one. In summary, French people ascribed more weight than the UK respondents to some or severe problems in mobility, self-care, and usual activities. They ascribed less weight to problems in the two other dimensions. The most

striking difference was observed for some problems in usual activities: the loss of utility is five times more important for the French people than for the UK respondents.

Fig. 1 Mean French and MVH TTO observed values for 24 health states

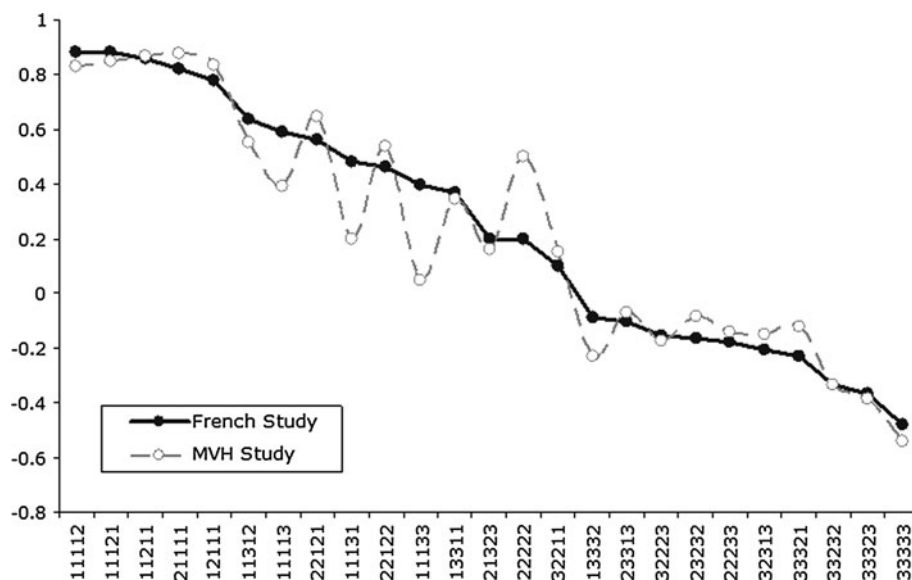
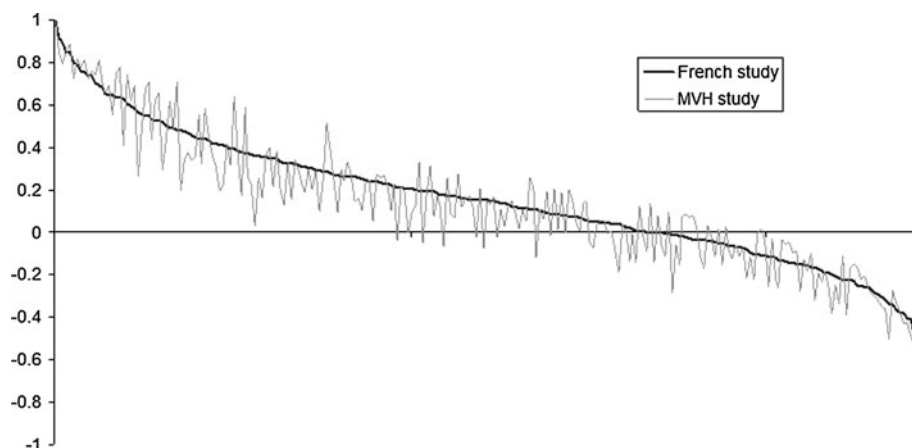


Fig. 2 French and MVH predicted values for the 243 health states (comparison of the 2 value sets)



Discussion

The study aimed to compute the French value set of the EQ-5D applying a modified version of the protocol used in the MVH study. Based on a representative sample of the general French population aged over 18, we tested several models. The best according to the fit statistics (AIC, MAE, and Pearson correlation coefficient) has been selected to assess the French tariff for the EQ-5D. The direct valuation of 24 health states was first obtained by TTO, and the negative TTO values were bounded using the monotonic transformation. The main modification in the MVH protocol concerned the number of respondents and the number of health states directly valued. Their determination was based on the analysis of other published valuation studies. According to Chuang and Kind [19], the minimal numbers of respondents and health states are, respectively, 100 and 29. A function (obtained with an estimated regression model) based on a smaller number of respondents or health

states should yield a MAE greater than 0.05, which is the threshold over which the difference is considered to be important [12, 20]. With the same threshold, Lamers' results [10], based on values obtained with TTO (instead of Visual Analog Scale as in Chuang and Kind), suggest a direct valuation of 22 health states by 300 respondents each. The present study was based on the data from 24 health states valued by at least 300 respondents each, which is in accordance with the conclusion of the authors previously mentioned.

Like the MVH A1 tariff, the French EQ-5D tariff was based on the N3 model and estimated using a random effects regression model. In absence of a French tariff, the A1 tariff has previously been widely used in economic analyses in France. To assess the effect of using the new French tariff in these analyses, both value sets were compared. Some major differences (greater or smaller scores) were observed especially for the intermediate health states in terms of severity. Although the significance of the

differences in coefficients should be further tested by computing a model on the combined UK and French data, we can assume that the difference in coefficients will probably lead to the differences when cost-utility ratios are calculated with one or the other value set.

Finally we found that computing the model on 17 or 24 health states did not produce major differences in value sets.

This study is the first to assess values to the EQ-5D health states in France; however, some limitations have to be mentioned.

First, as in most of other countries having computed their own value set, a monotonic transformation was applied to the values for states worse than death. This decision has a great impact on the scores of the health states valued directly and consequently on the estimated values. For example, consider a person which is indifferent between 8 years in a worse than death health state i and 2 years in full health followed by death. The utility value associated with the health state i (before transformation) is -0.25 ($-2/(10-2)$) (cf. methodology of the TTO). Applying the monotonic transformation results in a utility value of -0.2 ($-0.25/(1-0.25)$). It is still equal to -0.25 when negative values are bounded by -1 and equal to -0.00641 ($-0.25/39$) when they are divided by 39. Concerning the results of the model, Lamers [17] showed that the smallest MAE occurred when negative values were linearly transformed (i.e., $v' = \frac{v}{39}$ instead of $v' = \frac{v}{1-v} = \frac{v}{10}$, see Transformation of health states in the Methods section earlier). She also underlined the fact that modifying the bounding method for negative values at -1 results in different social tariffs for EQ-5D. Using the linear transformation as in the US study [18] should result in smaller numbers of QALYs and probably in smaller QALY gains, especially for more severe diseases. The method in which all the negative values are divided by 39 conducts to a model with best fit statistics than when other transformation methods are applied. In that case, in fact, the MAE between observed and predicted values by the “N3-model” (retained to compute the French value set) is smaller than when using the monotonic transformation as in our study. Thus, the comparability of national value sets requires at least the use of the same transformation method for the health states considered as worse than death. For this reason, the monotonic transformation was the method used in our study: all but one (the United States) of the countries who have EQ-5D TTO value sets used this method. When applied to the French data, even if it did not produce the best results, this transformation nevertheless yields a high goodness of fit between the estimated values and the observed ones ($MAE < 0.05$).

Second, we strictly used the same exclusion criteria as in other national valuation studies (number of non-valued

health states, values of 11111, and “Dead”...), apart from the treatment for inconsistencies. To our mind, inconsistent valuations may have two sources. First, respondents may be themselves inconsistent in their responses. That could be the case if they did not understand the task. But, inconsistencies could also be due to the difficulty of the TTO exercise. In both cases, inconsistent values influence the parameter estimates of the regression equation as they can conduct to product contrary effects to the ones expected. Thus, the exclusion of the respondents presenting some inconsistencies could have allowed a better quality data set and should improve the validity of the resulting value set.

Only a few authors have taken this criterion into account. In the New Zealand study, Devlin et al. [21] showed that computing a tariff admitting all inconsistencies or admitting none or just one resulted in different regression equations. Under a restricted budget, valuation studies include a minimal number of respondents. Thus, excluding inconsistencies could lead to the exclusion of an important number of respondents and may compromise the representativeness of the sample. This results in a reduction in the accuracy of the estimation of the parameters, as most of the respondents in our sample presented at least one inconsistency.

In our study, we tested the effect of the inconsistencies on the coefficient parameters. As in the New Zealand study, two tariffs were compared: the one presented in the paper (the « official » French value set) calculated on the whole data and another calculated on the data of the respondents presenting less than 5 inconsistencies in their answers. The two tariffs were very different (46% of the health states had a difference in their valuation greater than 0.05). Actually, computing a model on data presenting a great number of inconsistencies conducted to an underestimation of the difference of disutilities between levels 2 and 3 in each of the dimension. Results obtained in New Zealand presented the same feature. More investigations could be made to determine the impact of inconsistencies on the tariff and to confirm or not that tendency.

There is no simple solution to the issue of inconsistencies. It is intrinsic in the valuation exercise (TTO exercise but also Visual Analog Scale or Standard Gamble): even if a respondent tried to be consistent, he/she never has a global vision of his/her responses. This would be difficult to achieve, since each health state is considered one by one. Respondents could attempt to memorize their answers but the number of health states directly valued makes that task difficult. We could imagine modifying the TTO protocol so that the respondents could have a feedback on their whole responses but this solution would pose two problems. On one hand, entitling the respondents to change the value of the health states at the end of the exercise should lead to the

presence of non-TTO-based values. On the other hand, it will force the consistency among respondent who are not necessarily consistent.

The issue of inconsistencies is not a budget problem since increasing the number of respondents will not decrease the number of inconsistencies. It will just allow removing a subset of the respondents without damaging the accuracy of the estimates but it remains difficult to determine which respondents should be excluded. Their number differs between studies (90% of the respondents presented at least one inconsistency in our sample vs. 79% in the New Zealand one).

To conclude, our valuation study provided the French value set of the EQ-5D based on the stated preferences in France: a utility value is associated with each health state defined by this instrument. Although there are still some major debatable issues on methodological choices made, we have chosen to stick to the current methodology adopted up to now by other countries to ensure some standardization/comparability across countries, but limitations do call for further research. This research work has already received an institutional recognition, since it is recommended in the already quoted draft guidelines from the French HTA agency, the HAS [6].

Acknowledgments Financial support for this study was provided by a grant from the EuroQol Group, GlaxoSmithKline, and the ESSEC Chair of Health Economics and Management. We are grateful to the interviewers of TNS Healthcare for their help in data collection and our respondents for their participation. This study was not possible without the friendly and permanent support of Frank de Charro, exec director of EuroQol, and the guidance of the EUROQOL valuation task force, especially from Nancy Devlin, Paul Kind, Ben van Hout and David Parkin.

References

1. Brooks, R.: EuroQoL: the current state of play. *Health Policy* **1**, 53–72 (1996)
2. Torrance, G.W., Boyle, M.H., Horwood, S.P.: Application of multi-attribute theory to measure social preferences for health states. *Oper. Res.* **6**, 1043–1069 (1982)
3. Ware, J.E., Sherbourne, C.D.: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care* **6**, 473–483 (1992)
4. Le Galès, C., Buron, C., Costet, N., et al.: Développement d'un index d'états de santé pondéré par les utilités en population française: le health utilities index. *Econ. Prev.* **150**, 71–87 (2001)
5. Collège des Économistes de la Santé: Guide méthodologique pour l'évaluation économique des stratégies de santé. Recommandations méthodologiques. Paris (2003)
6. Haute Autorité de Santé: L'évaluation économique à la Haute Autorité de Santé Principes et méthodes. Paris (2010)
7. Adam, P., Herzlich, C.: Sociologie de la maladie et de la médecine. Armand Colin (2007)
8. Johnson, J.A., Ohinmaa, A., Murti, B., et al.: Comparison of Finnish and U.S.-based visual analog scale valuations of the EQ-5D measure. *Med. Decis. Making* **20**, 281–289 (2000)
9. Johnson, J.A., Luo, N., Shaw, J., et al.: Valuations of EQ-5D health states: are the United States and United Kingdom different? *Med. Care* **43**, 221–228 (2005)
10. Lamers, L.M., McDonnell, J., Stalmeier, P., et al.: The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* **15**, 1121–1132 (2006)
11. Golicki, D., Jakubczyk, M., Niewada, M., et al.: Valuations of EQ-5D health states in Poland: first TTO-based social value set in Central and Eastern Europe. *Value Health* **13**, 289–297 (2010)
12. Dolan, P.: Modeling valuations for EuroQol health states. *Med. Care* **11**, 1095–1108 (1997)
13. Macran, S., Kind, P.: Valuing EQ-5D health states using a modified MVH protocol: preliminary results. In: Badia, X., Herdman, M., Roset, M. (eds.) Proceedings of the 16th Plenary Meeting of the EuroQol Group, pp. 205–240. Institut de Salut Publica de Catalunya (2000)
14. Tsuchiya, A., Ikeda, S., Ikegami, N., et al.: Estimating an EQ-5D population value-set: the case of Japan. *Health Econ.* **4**, 341–353 (2002)
15. Parkin, D., Devlin, N., Sharma, P.: A deathless and VAS-free EQ-5D self-completion valuation questionnaire. In: Stavem, K. (ed.) Proceedings of the 22nd Plenary Meeting of the EuroQol Group, Oslo, pp. 170–189 (2005)
16. Szende, A., Oppe, M., Devlin, N.: EQ-5D value sets. Inventory, comparative review and user guide. Springer, Dordrecht (2007)
17. Lamers, L.M.: The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value-sets. *Med. Care* **3**, 238–244 (2007)
18. Shaw, J.W., Johnson, J.A., Coons, S.J.: US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med. Care* **3**, 203–220 (2005)
19. Chuang, L.H., Kind, P.: The effect of health state selection on the valuation of EQ-5D. *Med. Decis. Making* **31**, 186–194 (2010)
20. O'Brien, B.J., Drummond, M.F.: Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *Pharmacoeconomics* **5**, 389–398 (1994)
21. Devlin, N.J., Hansen, P., Kind, P., et al.: Logical inconsistencies in survey respondents' health state valuations—a methodological challenge for estimating social tariffs. *Health Econ.* **7**, 529–544 (2003)