ORIGINAL RESEARCH ARTICLE

A French Value Set for the EQ-5D-5L

Luiz Flavio Andrade[1] · Kristina Ludwig[2] · Juan Manuel Ramos Goni[3] · Mark Oppe[3] · Gérard de Pouvourville[1]

Abstract

Objective The objective of this study was to develop a French value set for the EQ-5D-5L, for academic and clinical research, and for regulatory requirements for price-setting of drugs and medical devices.

Method This study used the standardized valuation protocol developed by EuroQol, using computer-assisted personal interview software. A representative sample of 1048 French residents were interviewed by a market research company, under the supervision of the research team. Health states were valued using composite time trade-off and a discrete choice experiment. Modeling was used to create values for the 3125 possible health states. The composite time trade-off data were modeled using a Tobit model with censored observations at −1 and correcting for heteroscedasticity. A conditional logit model was used for the discrete choice results, and both models were combined using a hybrid model. An adjusted hybrid model was tested to correct for imbalance in the sample on age and sex compared with the general population. A comparison with the 3-level (3L) value set was performed.

Results The adjusted model was preferred to comply with the representativeness of the general population. It provided a value set for which all coefficients were logically consistent. Values ranged from − 0.525 to 1. The distribution of values presented a shift towards higher values versus the 3L value set. Ranking of dimensions changed. Pain and discomfort and mobility were the dimensions with the highest potential for disutility compared with mobility and self-care for the 3L instrument.

Conclusions This study provides a value set based on societal preferences of the French population, using an improved descriptive instrument of health-related quality-of-life health states. It will contribute to improve the quality of cost-effectiveness analysis in the French context and help stimulate disease-specific quality-of-life references for academic-, institutional-, and industry-promoted studies.

✉ Gérard de Pouvourville
pouvourville@essec.edu

Luiz Flavio Andrade
luizflaro81@gmail.com

Kristina Ludwig
ludwig@euroqol.org

Juan Manuel Ramos Goni
jramos@axentiva.com

Mark Oppe
moppe@axentiva.com

[1] Ecole Supérieure de Sciences Economiques et Commerciales (ESSEC Business School), 3, Avenue Bernard Hirsch, CS 50105, Cergy Pontoise, France

[2] Euroqol Research Foundation, Marten Meesweg 107, 3068 AV Rotterdam, The Netherlands

[3] Axentiva Solutions, Calle Calvario, 271-1 B, 38340 Tacoronte, Tenerife, Spain

Key Points for Decision Makers

This study aimed at providing health care decision makers with an up-to-date metric to measure the health benefit of health care products and programs.

In many studies, the 5-level EQ-5D questionnaire is considered a major improvement versus the 3-level version, to capture incremental changes in quality of life.

The availability of this value set will facilitate the publication of robust cost-effectiveness analyses, based on French surveys for different disease areas, whereas, up to now, the availability of such data has been recognized as one of the major limitations of the validity of such studies.

1 Introduction

Since October 2013, drug and medical device companies applying for coverage by the French National Sickness Fund and whose products are assessed by the French health technology assessment (HTA) body, the 'Haute Autorité de Santé' (HAS), must submit a cost-effectiveness analysis (CEA) under the following conditions. They claim for recognition of the status of 'innovative product' (i.e. an improvement in medical service rendered/expected of 1–3); they expect a revenue of €20 million after 2 years of sales; and/or an impact on the delivery or organization of health services is expected [1, 2]. Applications by companies are assessed by the Commission for Economic and Public Health Evaluation (CEESP), in parallel with the clinical assessment performed by the Transparency Commission (CT). In France, cost per quality-adjusted life-year (QALY) is the recommended type of analysis in most cases. Companies should provide QALY values related to treatments of interest using data collected in the French context and in French validated value sets [3]. Presently, two French population-based value sets are recommended: EuroQol 5-Dimension 3-Level (EQ-5D-3L) tariffs, and the Health Utility Index (HUI) 3 [4–6]. Other ad hoc utility values are accepted, if fully justified and published in peer review journals.

Hamers et al. [7] have published a review article focused on utility measures, employed in submissions to the HAS to the end of 2015. A total of 32 submissions were assessed. Two submissions covered two indications, and 34 CEAs were analyzed. The EQ-5D-3L was used in 24 CEAs, while HUI 3 was only used in one submission. Other methods were disease-specific instruments or time trade-off (TTO) applied to specific vignettes. Thus, EQ-5D was the most used instrument.

Chevalier and de Pouvourville [4] have previously published a French value set for the 3L version of the EQ-5D questionnaire and had participated in the linguistic validation of the French (for France) version of the EQ-5D 5-Level (EQ-5D-5L) questionnaire. The performance of the 5L version in terms of better descriptive and discriminative power than the 3L version has been shown in many countries [8–12]. The availability of an updated version of the standardized valuation protocol [13], with improvements for optimized data collection and its implementation in the EuroQol Valuation Technology (EQ-VT) software version 2.0 [14] by EuroQol, opened the opportunity of performing a valuation study for the EQ-5D-5L in France. This initiative was presented to the CEESP in June 2017, who provided strong support for it [15].

Thus, the primary objective of this study was to provide a value set reflecting societal values for the health states generated by the EQ-5D-5L, in the French population. Subsequently, we compared the 5L value set with the initial 3L tariff and with the crosswalk value set published in 2012 [16].

2 Methods

The study used the valuation protocol and its associated computer-assisted interview software (CAPI) EQ-VT version 2.0 developed by EuroQol [14]. Support from EuroQol included a training kit, a full script for interviewers, and a quality control module with specific criteria to adopt or reject interviews/interviewers. The French study team validated the French version of all documents, which were translated by a professional translation company. Interviews were performed by professional interviewers from a private market research company who had previously participated in the 3L valuation exercise.

2.1 Sample Selection

Following the protocol from EuroQol [17], a sample of 1000 respondents aged ≥ 18 years was targeted. The market research company targeted a sample size of 1100 to ensure a final sample of 1000 respondents with valid responses. Each week, interviewers received a quota sheet of 10 targets, with specific characteristics of respondents in terms of age, sex, and socioeconomic status. Sampling was based on national statistics. Geographical representativeness was not targeted, but interviewers were selected to provide reasonable coverage of the territory and population size of the residential location of respondents.

2.2 Eliciting Preferences

Respondents used EQ-VT interactively with interviewers. Each respondent was presented with a subset of the 3125 health states of the EQ-5D-5L, for which two preference elicitation tasks were required: composite TTO (cTTO) [18] and discrete choice experiment (DCE).

The EQ-VT design included a set of 86 EQ-5D-5L health states, divided into 10 blocks of 10 health states for the cTTO tasks (in which some states were present in multiple blocks), and 196 pairs of EQ-5D-5L health states, divided into 28 blocks of seven pairs for the DCE tasks [18]. Each respondent was randomly assigned by the software to one of the cTTO or DCE blocks. They were first presented with 10 health states using cTT0, and then proceeded to a DCE, where they were asked to choose one of two displayed health states, for seven pairs of health states.

2.3 Interview Process

Before starting the elicitation tasks, respondents were asked to declare their present health state, using the EQ Visual Analog Scale (EQ VAS) and the EQ-5D-5L questionnaire. Supplementary questions related to age, sex, direct or indirect experience of disease, plus other background questions on level of education and professional activity were included. In the typical cTTO task that respondents had to perform, they were asked to choose their point of equivalence between living 10 years in a given health state and x years in full health. When respondents considered that a given health state was 'worse than being dead', they were shifted to a 'lead time' TTO [19], for which they had to find a point of equivalence between two different 'lives', one lasting 20 years with 10 years in full health followed by 10 years in the health state worse than being dead, and x years of full health. The interview process included an explanation of the TTO, using 'being in a wheelchair' as an example and three practice health states to familiarize the respondents with the cTTO task and to prepare them with health states they might consider as 'worse than being dead'. After completion of the cTTO task, respondents were presented with a feedback module [20] presenting an overview of their valuations; health states were ranked from the less severe to the most severe, with states valued as 'equal' placed side by side. They were then asked to confirm their choices or to identify which of the health states they considered to be incorrectly ranked. The number of inconsistencies were reported and were used in the quality control process [21].

In the DCE tasks, respondents were presented with a pair of EQ-5D-5L health states, A and B, and were asked to state which one was better. Indifference was not an option.

2.4 Interviewers Monitoring

The French study team received a 2-day training session. The first group of 15 interviewers received an initial 1.5-day training session, before proceeding to a pilot test of 10 interviews per person. Interviews were conducted at the interviewees' homes.

Standard quality control criteria were predefined as follows [21]:

- The time spent on explaining the TTO task using the wheelchair example was too short (< 3 min).
- No explanation of the 'worse than dead' task ('lead time' TTO) was given in the wheelchair example.
- Inconsistencies in the cTTO ratings (a value of 55555 was not the lowest and was at least higher than the state with the lowest value, by 0.5).
- Time spent for the 10 TTO tasks was < 5 min.

If any of the criteria were met, the interview was flagged as being of suspect quality. Any batch of 10 interviews from a single interviewer with 40% of flags or more was rejected. Figure 1 presents the flowchart of the interview process (the full final quality control report is shown in electronic supplementary material [ESM] 1).

2.5 Modeling Methods

The study provided two different types of data to be modeled: the cTTO values and the DCE choices. The dependent variable for the cTTO values was obtained by subtracting the value of the cTTO values from 1, allowing the data to take only positive values. The DCE-dependent variable was a dummy variable with a value of 1 if the health state was chosen, and 0 if not. Dummies for the increments between consecutive levels were used to capture the disutility associated when moving from one level of the health dimension to another. Since the cTTO responses cannot take values lower than −1, a Tobit modeling approach was used to deal with the censored nature of the dependent variable. The cTTO values flagged during the feedback module were also excluded from the analysis. The DCE data were modeled using a conditional logit model.

Following the study by Ludwig et al. [22], a hybrid model approach was performed in order to maximize the information of the whole data set if the cTTO and DCE data were found to be in close agreement. The special feature of the hybrid model is that it estimates a single set of coefficients based on the two different types of data. A scaling function, theta, is introduced in order to rescale the estimates between TTO and DCE. If TTO and DCE scales are proportional, theta can be a single scaling parameter. Thus, coefficients can be easily compared and can take a value of 1 for full health and 0 for the health state 'being dead'. We used a 20-parameter model where the explanatory variables are incremental dummies for the five dimensions of the EQ-5D-5L, with level 1 considered as the reference. Incremental dummies allow to interpret the coefficients as being the variation in the disutility of health when moving from one level to the next.

The French value set was calculated using the specific hyreg STATA command created by EuroQol, which computes utility values using a hybrid specification model [23]. The hyreg command includes distributional modifiers, allowing heteroscedasticity to be taken into account.

In the cTTO-only models that were tested, the intercept term was close to zero and was non-significant ($p = 0.341$). By definition, the DCE model has no intercept. Consequently, the final models were estimated as follows, with no intercept term:

Fig. 1 Data collection process and quality control of interviewers
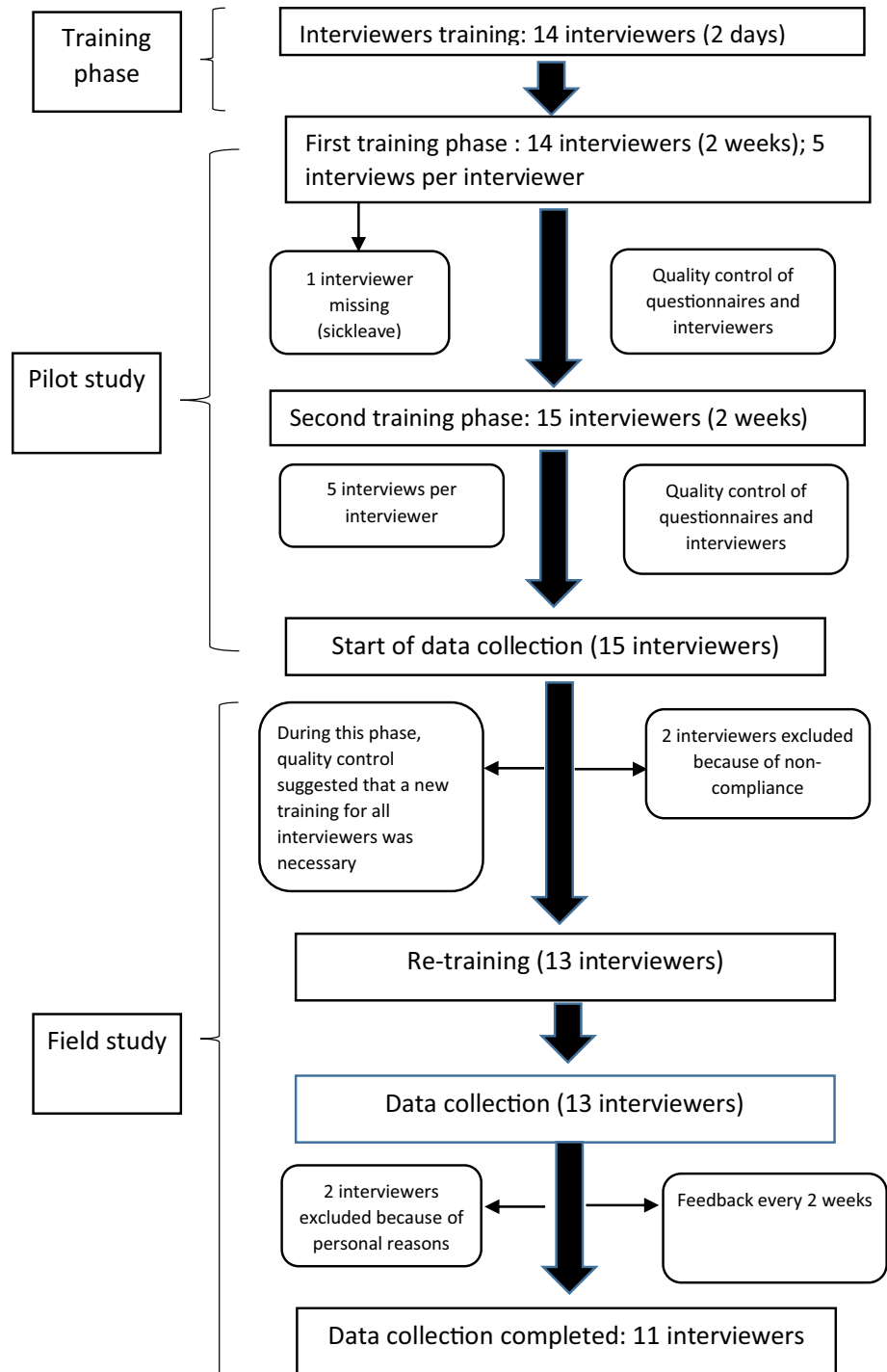
$$
\begin{aligned}
Y = {}& \beta 1 \times \mathrm{MO2} + \beta 2 \times \mathrm{MO3} + \beta 3 \times \mathrm{MO4} + \beta 4 \times \mathrm{MO5} \\
& + \beta 5 \times SC2 + \beta 6 \times \mathrm{SC3} + \beta 7 \times \mathrm{SC4} + \beta 8 \times \mathrm{SC5} \\
& + \beta 9 \times \mathrm{UA2} + \beta 10 \times \mathrm{UA3} + \beta 11 \times \mathrm{UA4} + \beta 12 \times \mathrm{UA5} \\
& + \beta 13 \times \mathrm{PD2} + \beta 14 \times \mathrm{PD3} + \beta 15 \times \mathrm{PD4} + \beta 16 \times \mathrm{PD5} \\
& + \beta 17 \times \mathrm{AD2} + \beta 18 \times \mathrm{AD3} + \beta 19 \times \mathrm{AD4} + \beta 20 \times \mathrm{AD5}
\end{aligned}
$$

We analyzed whether the sample was representative of the French general population. In case of not being representative, we re-estimated the model to test whether the factors of non-representativeness had an impact on the estimated values. If there was a significant impact, then a weighted model was estimated.

In order to estimate the weighted model, we first calculated the specific weights associated with each respondent

to force the sample to be representative of the total French population. Once weights were available, the likelihood function of each observation was multiplied for the respondent weight.

The full analysis process was monitored by one co-author, who was, at that time, a member of the EUROQOL consortium and a team member of the valuation studies. All initial models were run by the first author, and were subsequently checked by the EuroQol support team.

2.6 Comparing Value Sets

We compared the 5L value set with the French 3L value set and with values derived for France by the crosswalk interim scoring. Comparison was performed graphically using the Kernel distribution of value sets. The range of values, and ranking of dimensions and values for a selection of health states, were compared between the 5L and 3L values.

3 Results

A total of 1143 individuals were interviewed between March 2018 and November 2018, of whom 95 were excluded due to the poor quality of data. The final sample of respondents was 1048. The exclusion criteria were related to interviewers not complying to instructions or when serious inconsistencies in valuating health states were observed. Regarding the noncompliance in protocol rules, we excluded interviews for each interviewer who had not shown the 'worse than dead' configuration in the training part of the survey. The inconsistencies were related to conditions in which the respondent gave the worse state of 55555 a value that was higher than the value given to the mildest health state presented in the TTO task.

After exclusions, the average number of interviews per interviewer was 71.4 (standard deviation [SD] 30.9; minimum 10, median 82, maximum 132). The average time of interviews was 39.2 min (SD 9.4; minimum 17, median 37.8, maximum 95.4). The average time of a single TTO task was 60 s (SD 43.1; minimum 1.7, median 48.5, maximum 1081) and a single DCE task took an average time of 38.8 s to be completed (SD 30.0; minimum 4.8, median 30.2, maximum 725.9).

3.1 Sample Characteristics

Table 1 displays descriptive statistics of the sample. The final sample after exclusions ($n = 1048$) did not present major differences from the total sample ($N = 1143$). The average age was 49.4 years, while women represented 55.44% of respondents. The market research company used a three-level standard description of socioprofessional status

(higher retired and active socioprofessional status; lower retired and active socioprofessional status; no professional activity). With this classification, final distribution of the sample was consistent with stratification goals (see ESM 3 for description of socioprofessional classes).

The final sample presented with a difference in age and sex in the French general population [24]. An overrepresentation of females versus males in the sample was observed when compared with the planned stratification and with the general population. A breakdown of age groups per sex (Fig. 2) shows that there is an imbalance in favor of the 25–34 years age group for both sexes, and an imbalance for women, with a deficit in the number of women respondents in the older age group (75 years and older) versus women in the 55–74 years age group. An extra quota of 20 women aged 65 years and over was surveyed to reduce this imbalance but was not sufficient for a full correction. According to the market research company, acceptance of interviews was lower in this age group.

Figure 3 represents the geographical distribution of respondents. Compared with national statistics, rural areas were well represented, whereas there was an underrepresentation in population size of residences of 2,000–100,000 inhabitants, and an overrepresentation of people living in residential units of over 100,000 inhabitants, and also the Paris 'Petite Couronne' (i.e. Paris + 4 adjacent departments). Supplementary data on the sample, including reporting on the personal experiences of diseases are presented in ESM 3.

3.2 Data Characteristics

Respondents declared 181 health states out of 3125. The list of declared health states is presented in ESM 4. Of 181 health states, 5 represented 50% of the sample health states declared by respondents (11111, 11112, 11113, 11114, and 11121).

Overall, 20.2% of cTTO values were negative, with 2.3% elicited at − 1 (Fig. 4). An unwillingness to trade-off full health (value 1) was observed in 13.7% of responses. In addition, values of 0.5 and − 0.5 were often observed (9.29% and 5.2%, respectively) but were not interviewer-dependent. The proportion of values around 0 ($\pm - 0.05$) was 3.7%.

3.3 Value Set

Altogether, seven models were tested: (1) a cTTO tobit model unadjusted for age and sex; (2) a DCE logit model unadjusted for age and sex; (3) a hybrid model unadjusted for age and sex; (4) a hybrid model adjusted for age and sex; (5) a hybrid model adjusted for age only; (6) a hybrid model adjusted for sex only; (7) and a main effect adjusted hybrid model.
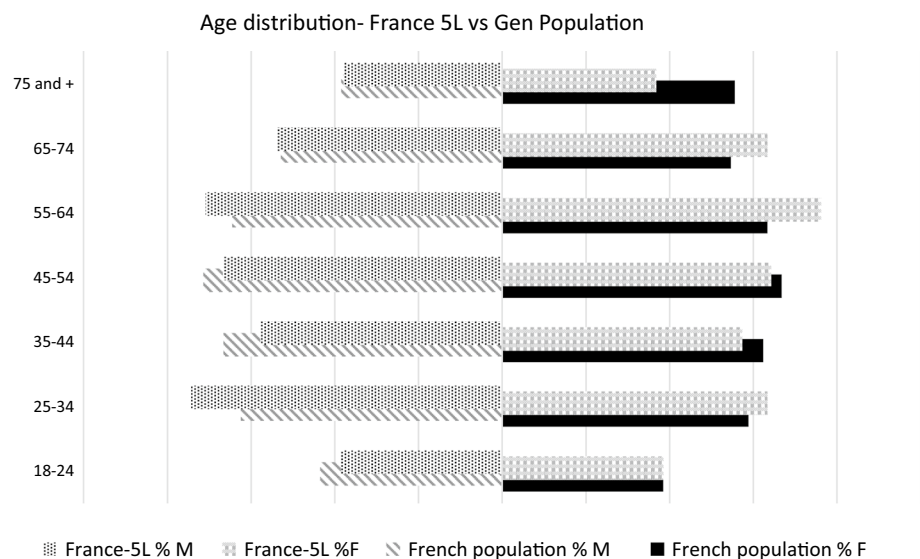
Table 1 Characteristics of respondents

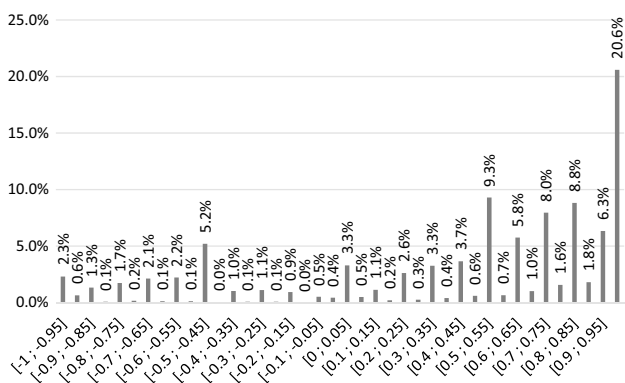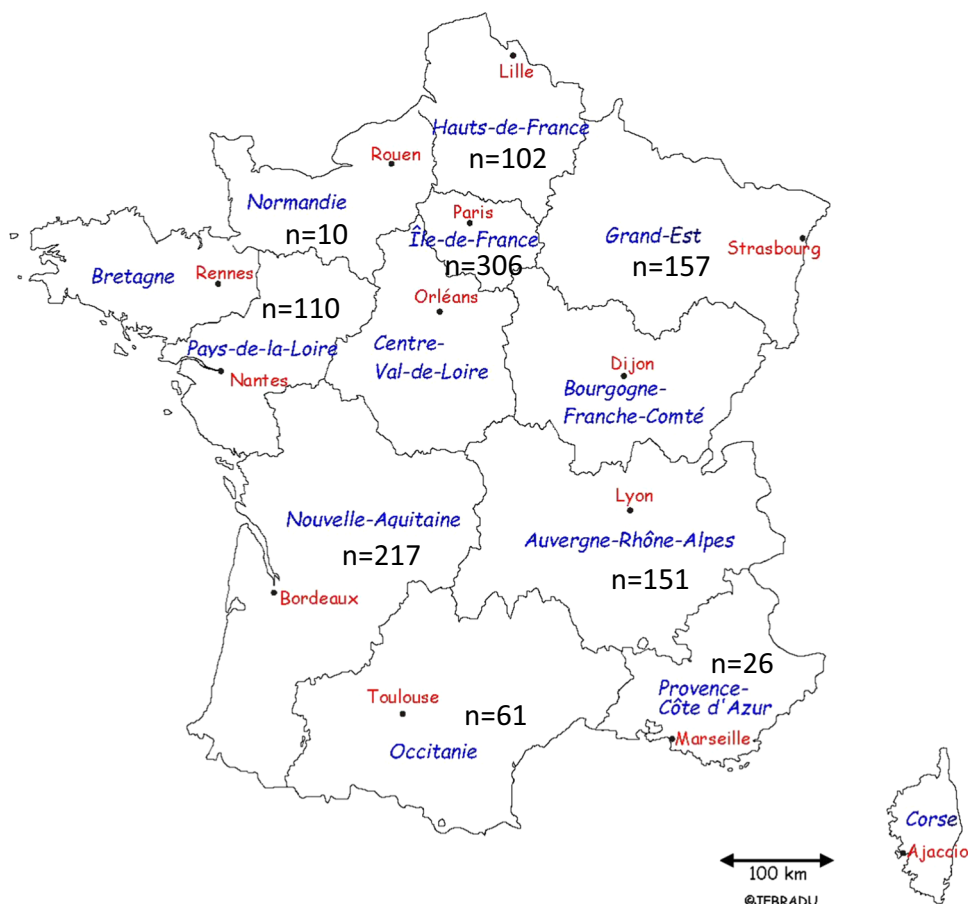|	Full sample (N = 1143)		Selected sample (n = 1048)		Excluded sample (n = 95)	
	Mean	Range	Mean	Range	Mean	Range
Age	49.6	18–93	49.4	18–93	51.5	19–87
	Frequency	%	Frequency	%	Frequency	%
Sex						
Female	630	55.12	581	55.44	49	51.58
Male	513	44.88	467	44.56	46	48.42
Level of education[a]						
Missing	5	0.4	5	0.48	0	0
Level 1	216	18.9	197	18.8	19	20.0
Level 2	561	49.2	516	49.2	45	47.3
Level 3	361	31.5	330	31.49	31	32.6
Occupation						
Missing	5	0.44	5	0.48		
Farmers	4	0.35	3	0.29	1	1.05
Craft and trade, Managers	37	3.24	35	3.34	2	2.11
Managerial staff	104	9.1	96	9.16	8	8.42
Employees	253	22.13	233	22.23	20	21.05
Students/retired/domestic activities	422	36.92	385	36.74	37	38.95
Workers	146	12.77	133	12.69	13	13.68
Technicians and Clerks	172	15.05	158	15.08	14	14.74
	Mean	Range	Mean	Range	Mean	Range
VAS 5L	78.9	0–100	79.0	0–100	78.6	25–100
Distribution of values (N = 1048)	1	90–99	80–99	≤ 80		
Frequency, %	5.2	35.0	27.3	32.5		

VAS visual analog scale, *5L* 5-level

[a]Level 1: 8 years of education after preschool; Level 2: 14 years of education after preschool; Level 3: 16 years or more of education after pre-school

Fig. 2 Comparative distribution of age in the study and the general population [24]. *5L* 5-level, *M* male, *F* female

Age distribution- France 5L vs Gen Population

▨ France-5L % M   ▧ France-5L %F   ◿ French population % M   ■ French population % F

Fig. 3 Geographical distribution of respondents (N = 1140)

Fig. 4 Observed distribution of composite time trade-off values

When including age and sex in the hybrid model, only age was highly significant (p = 0.023), but its coefficient was small (0.00250). When including age alone, it was no more significant (p = 0.066). Nevertheless, because the initial objective of the study was to provide a value set that reflects preferences of the general population, correction for sample biases was essential. Thus, a hybrid main effect model adjusted for age and sex was performed and is the preferred value set. This model was compared with an unadjusted hybrid model to measure the effect of adjusting. In Table 2, we present the cTTO and DCE models, followed by the unadjusted and adjusted main effects hybrid model. Coefficients are incremental utility variations when moving from one level to the next. Using the sum of levels across dimensions as a proxy for health state severity, the higher the severity, the lower the mean cTTO values but the higher the SD, indicating heteroscedasticity in the cTTO data (Fig. 5). Heteroscedasticity was thus taken into account by modeling the variance. The theta rescaling coefficient was 5.226 (the full data of the preferred value set, including Sigma statistics, are shown in ESM 2, and the full value set is shown in ESM 5).

The appropriateness of the models can be assessed by identifying the inconsistencies in each specification. We expect disutility to increase as we move to worse health conditions. Both the cTTO and DCE models present one illogically ordered coefficient (MO3 and UA3, respectively), which is corrected for in all hybrid models. The agreement between models can also be assessed by comparing the ordering on the most impacted dimensions of health-related quality of life. UA was the dimension with the lowest cumulative decrement in all models, but models differ in the relative position of mobility, anxiety/

Table 2 Value set

	Model 1		Model 2		Model 3		Model 4	
	Tobit cTTO		Logit DCE		Hybrid		Hybrid	
					Unadjusted		Adjusted	
	β (SE)	p > t	β (SE)	p > t	β (SE)	p > t	β (SE)	p > t
MO2	0.018 (0.014)	0.180	0.379 (0.058)	0.000	0.033 (0.005)	0.000	0.037 (0.005)	0.000
MO3	**−0.0002 (0.015)**	0.988	0.022 (0.067)	0.739	0.009 (0.009)	0.314	0.010 (0.009)	0.280
MO4	0.141 (0.017)	0.000	0.738 (0.066)	0.000	0.134 (0.010)	0.000	0.131 (0.010)	0.000
M05	0.145 (0.017)	0.000	0.613 (0.070)	0.000	0.140 (0.010)	0.000	0.145 (0.010)	0.000
SC2	0.035 (0.013)	0.006	0.288 (0.062)	0.000	0.037 (0.005)	0.000	0.036 (0.005)	0.000
SC3	0.018 (0.017)	0.277	0.022 (0.069)	0.758	0.012 (0.008)	0.178	0.014 (0.008)	0.110
SC4	0.137 (0.017	0.000	0.684 (0.073)	0.000	0.122 (0.010)	0.000	0.121 (0.010)	0.000
SC5	0.076 (0.014)	0.000	0.480 (0.067)	0.000	0.086 (0.009)	0.000	0.086 (0.009)	0.000
UA2	0.037 (0.014)	0.007	0.351 (0.059)	0.000	0.031 (0.005)	0.000	0.033 (0.005)	0.000
UA3	0.025 (0.015)	0.101	**−0.070 (0.07)**	0.299	0.005 (0.008)	0.504	0.006 (0.008)	0.437
UA4	0.098 (0.017)	0.000	0.657 (0.068)	0.000	0.119 (0.009)	0.000	0.117 (0.009)	0.000
UA5	0.068 (0.017)	0.000	0.450 (0.070)	0.000	0.082 (0.009)	0.000	0.083 (0.009)	0.000
PD2	0.021 (0.012)	0.076	0.353 (0.062)	0.000	0.023 (0.004)	0.000	0.021 (0.004)	0.000
PD3	0.004 (0.017)	0.834	0.033 (0.068)	0.625	0.025 (0.009)	0.005	0.025 (0.009)	0.005
PD4	0.215 (0.016)	0.000	1.136 (0.070)	0.000	0.217 (0.010)	0.000	0.216 (0.010)	0.000
PD5	0.222 (0.017)	0.000	0.756 0.074)	0.000	0.179 0.011)	0.000	0.180 (0.010)	0.000
AD2	0.032 (0.013)	0.013	0.225 (0.066)	0.001	0.019 (0.004)	0.000	0.020 (0.005)	0.000
AD3	0.013 (0.016)	0.423	0.181 (0.068)	0.008	0.029 (0.008)	0.001	0.026 (0.008)	0.003
AD4	0.166 (0.016)	0.000	0.826 (0.075)	0.000	0.154 (0.009)	0.000	0.153 (0.009)	0.000
AD5	0.060 (0.015)	0.000	0.218 (0.070)	0.002	0.058 (0.009)	0.000	0.057 (0.009)	0.000
Observations	9803		7336		17,139		17,139	
cTTO CU	9562				9562		9562	
cTTO CLS	241				241		241	
DCE DIC			7336		7336		7336	
U (11111)	1		1		1		1	
U (21111)	0.982		0.927		0.966		0.962	
U (12111)	0.965		0.944		0.963		0.963	
U (11211)	0.963		0.932		0.969		0.967	
U (11121)	0.979		0.932		0.977		0.978	
U (11112)	0.968		0.957		0.981		0.980	
U (55555)	− 0.531		− 0.596		− 0.522		− 0.525	
#illogically ordered	1		1		0		0	
#non-signific	7		4		3		3	
Ranking of dimensions	PD, MO, AD, SC, UA		PD, MO, SC, AD, UA		PD, MO, AD, SC, UA		PD, MO, SC, AD, UA	

DCE discrete choice experiment, *cTTO* composite time trade-off, *SE* standard error, *CU* continuous uncensored, *CLS* continuous left censored, *DIC* dichotomous observations, *PD* pain and discomfort, *MO* mobility, *AD* anxiety/depression, *SC* self-care, *UA* usual activities

Bolded values represent illogical coefficients

depression, and self-care; however, in all models, cumulative decrements of anxiety/depression and self-care are very close. In the hybrid non-adjusted model, anxiety/depression ranks third, and also ranks third in the hybrid adjusted self-care model (Table 3). For 2402 health states, utility values were higher in the unadjusted model versus the adjusted model, which is consistent with what was expected by correcting for the imbalance in age. The value of the worst health state (55555) was − 0.5255 in the adjusted model versus − 0.5217 in the unadjusted model.
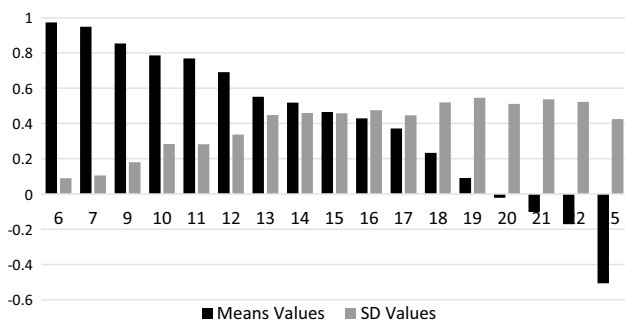
Fig. 5 Mean TTO value, by level sum score. *TTO* time trade-off, *SD* standard deviation

Table 3 allows for calculation of the utility of any given health states, using cumulative decrements. For example, the utility for the health state 54321 from the adjusted model is equal to $1 - 0.32509 - 0.172251 - 0.03979 - 0.02198 = 0.441$.

Figure 6a, b, and c represent the scatterplots of the predicted values of two by two models of the 86 health states assessed in the cTTO part of the study; Fig. 6d is the scatter plot of the predicted values versus the observed values of the same health states using the adjusted model. DCE coefficients have been rescaled using the theta parameter to facilitate the comparisons. The DCE model provides a better fit in terms of convergence with the adjusted hybrid model, than the cTTO model. This has also been the case when comparing each model's predicted versus observed values for the 86 health states from the cTTO experiment (see ESM 6). Thus, data support the assumption of proportionality between cTTO and DCE coefficients, and justify using a hybrid model, which brings together two different sources of stated preferences, with a larger number of health states than for each submodel alone.

3.4 Comparing Value Sets: 5-Level, 3-Level, and Crosswalk

Figure 7 provides the Kernel density distributions for the French 5L value set, the 3L, and the 5L crosswalk. It highlights a displacement of 5L utility values to the right side of the distribution, indicating a shift to higher values. The 5L crosswalk distribution curve is similar to the 3L value set.

In the 3L version, 78/243 (32%) health states had a negative value, while in the 5L version, this number was 401/3125 (12.8%), confirming that this shift to higher values also impacts negative values. However, this is mitigated by the fact that if 5 and 3 are considered as the worst levels in both sets, there are proportionally less health states including a 5 (67%) than those including a 3 (87%). The worst health state has a value of -0.52, and was 0.53 in the 3L and the crosswalk value sets.

The ranking of dimensions has changed. In the 3L version, the worst utility decrement (Level 3) was observed for mobility, followed by self-care and then pain/discomfort (followed by anxiety/depression and usual activities). In the 5L version, the ranking was pain/discomfort, mobility, self-care, anxiety/depression, and usual activities. The coefficients of self-care and anxiety/depression are very close. Maximum decrements are also lower in the 5L value set than in the 3L value set. In the 3L value set, the maximum decrement for MO was 0.5602, versus 0.3250 for 5L; the maximum decrement for PD was 0.4517 in the 3L value set and 0.4439 in the 5L value set. *Mutatis mutandis,* main differences with the 5L crosswalk, are quite similar to what was observed for the 3L value set.

Table 4 presents a selection of health states and values for both questionnaires, which confirm the shift to the right of the 5L value set. However, caution is recommended when comparing the 3L and 5L value sets, since formulation of the worse level for mobility is 'confined to bed' (3L version) versus 'unable to walk about' (5L version); intermediate level 2 labeling in the 3L version is classified as 'some

Table 3 Cumulative decrements of utilities per dimension

	MO	SC	UA	PD	AD
Unweighted model					
Level 2	0.0338811	0.0374599	0.0310561	0.0231809	0.0192951
Level 3	0.0433628	0.0494621	0.0367977	0.0487449	0.0484484
Level 4	0.1778412	0.1715431	0.1558802	0.2660137	0.2027724
Level 5	0.318512	0.2580992	0.2380036	0.4455017	0.261614
Weighted model					
Level 2	0.03759	0.03656	0.03313	0.02198	0.02046
Level 3	0.04774	0.050781	0.03979	0.04704	0.04683
Level 4	0.17949	0.172251	0.15689	0.26374	0.20005
Level 5	0.32509	0.258331	0.24005	0.44399	0.25803

PD pain and discomfort, *MO* mobility, *AD* anxiety/depression, *SC* self-care, *UA* usual activities

Fig. 6 Scatterplot of predicted values of the **a** adjusted hybrid model versus cTTO, **b** DCE model versus cTTO, **c** adjusted hybrid model versus DCE, and **d** adjusted hybrid model versus observed values. *cTTO* composite time trade-off, *DCE* discrete choice experiment
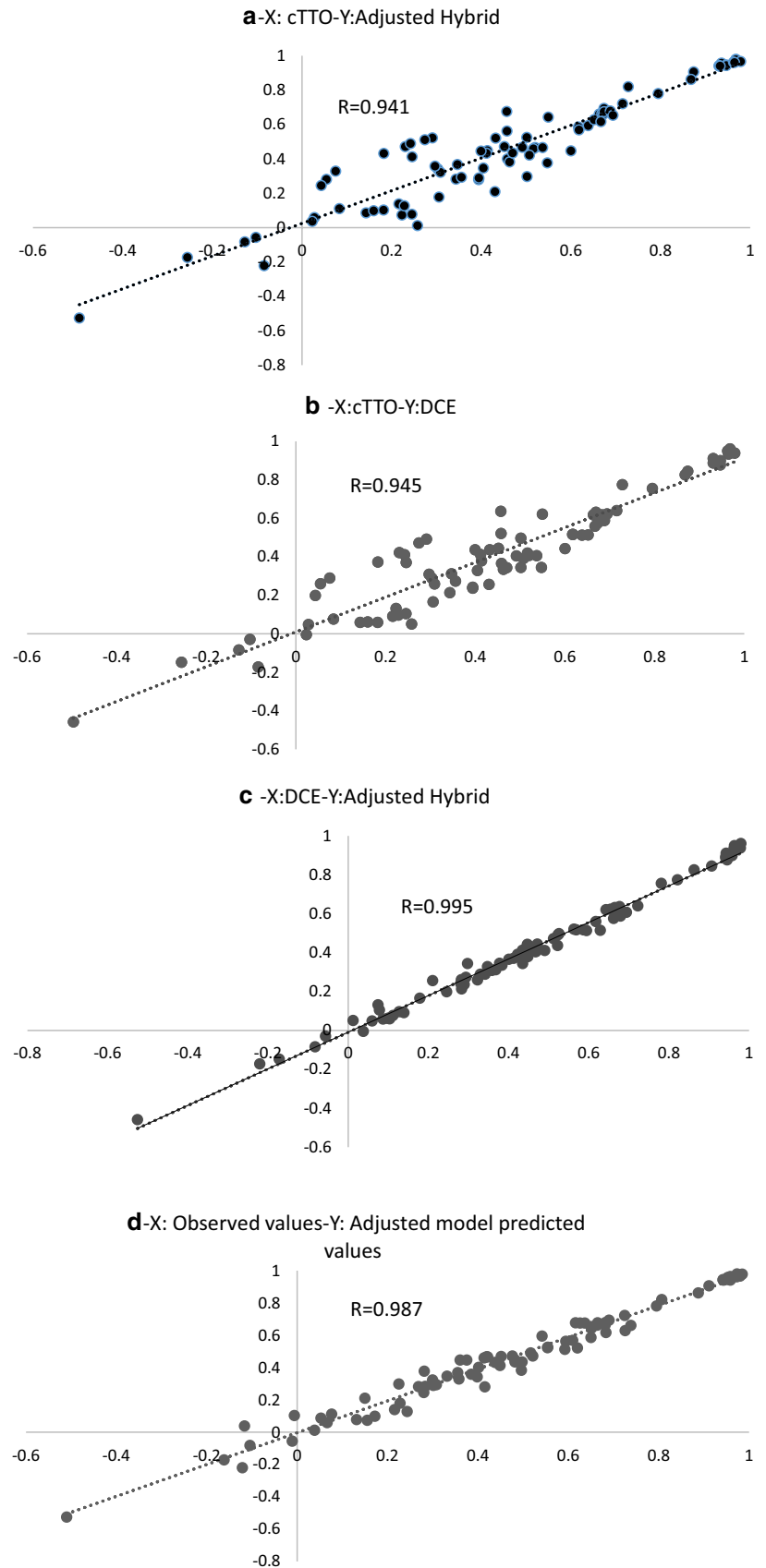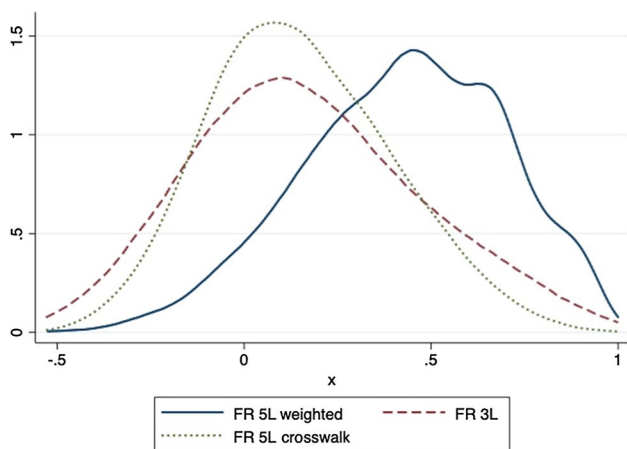
Fig. 7 Compared Kernel distribution of values. *5L* 5-level, *3L* 3-level, *FR* French, L, 3L, and 5L crosswalk

Table 4 Comparing 5L–3L values

Health states, 5L	Value	Health states, 3L	Value
11111	1	11111	1
22222	0.850		
33333	0.768	22222	0.275
44444	0.027		
55555	− 0.525	33333	− 0.53

5L 5-level, *3L* 3-level

problems', whereas it is classified as 'moderate problems' for intermediate level 3 in the 5L version.

4 Discussion

The 5L version of the EQ-5D questionnaire was developed to meet critics regarding the lack of sensitivity of the 3L version to small changes in quality of life, leading, for example, to concentration effects of 11111 answers, and to difficulties in assessing intermediary levels between moderate and extreme problems [25]. Thus, in a country such as France, where cost-utility analysis is increasingly required or studied and is now mandatory to inform pricing decisions for innovative treatments, it was necessary to proceed to a valuation study.

In this study, the choice of combining cTTO and DCE was dictated by the results, as in other valuation studies [26]: the strong agreement between the cTTO and DCE data and the improvement in the fit between observed and predicted values. In addition, using the hybrid model led to compensating for non-logical findings for the estimation of the 'mobility' L3 utility decrement in the cTTO tobit model,

and the 'usual activity' L3 decrement in the conditional logit model for DCE.

Results show important changes versus the earlier EQ-5D-3L value set. Values shifted upwards, as was also found in Germany [26], the UK [27] and The Netherlands (cTTO model) [28]. There were also proportionally fewer negative values in the 5L value set, which may be related to the introduction of 'lead time'. However, more strikingly for France, the ranking of dimensions has changed. In the 3L version, mobility and self-care came first and second, followed by pain and discomfort, whereas the 5L version ranks pain and discomfort and mobility first, followed by self-care, anxiety/depression, and usual activities. Maximum utility decrements are also smaller. Finally, at the time of the first French valuation studies, quality monitoring of the interviewers was not routinely implemented and may have led to higher interviewer effects. Thus, there is more confidence in the stated preferences of respondents than in the 3L study.

A comparison of the French value set with other published sets in the European context confirms the validity of maintaining national tariffs. In Germany [26], 55555 is valued at − 0.661, in the UK [27] it is valued at − 0.285, in Spain [30] it is valued at − 0.416, and in Ireland [29] it reaches a minimum of − 0.974. There are also differences between France, Germany, the UK, and Spain in the ranking of dimensions: pain and discomfort ranks first in Germany, UK, Spain, and France, but comes second to anxiety and depression in Ireland.

The issue of the representativeness of the sample needs to be discussed. If distribution according to socioeconomic status was consistent with the initial stratification goals, our sample had a higher rate of female respondents than in the general population, with an excess in the 55–74 years age group and a deficit in the ≥ 75 years age group, and a higher relative rate of young (25–34 years) and mature (55–64 years) male respondents. It was felt necessary to correct for such differences to comply with the ground principles of the elicitation of preferences on a representative sample of the population. The adjusted model showed little changes in coefficients for all dimensions. Geographical distribution was not a stratification criterion. Nevertheless, we observed an overrepresentation of Paris and immediate surrounding 'departments' (Petite Couronne).

The availability of a standardized valuation protocol has facilitated the transition to the French 5L value set, expected by academics and promoters of health care products and programs. The 5L version appears to capture small changes in health-related quality of life. Nevertheless, according to Hernandez et al. [31], in the UK the shift to the right, and the higher concentration of high utility values, could lead to lower QALY increments and higher incremental cost-effectiveness ratios (ICERs), except for treatments with high life-year gains, which may raise issues of historical consistency

between past (with 3L) and future decisions. Continuing to use the crosswalk as an interim solution would not lead to major changes versus the 3L.

Contrary to the National Institutes for Health and Care Excellence in the UK, which uses cost-utility analysis as a major criterion for access to coverage, in France the results of the economic evaluations presented by companies only serve as additional information in the price negotiation. Regulations and price agreements between the payer and companies have rejected the setting of a threshold, be it a single value or a range. Thus, there is no historical backlog against which past decisions may be challenged by a change in the valuation system, even if one cannot underestimate the scaling effects when recent assessments have already provided ultra-high ICERs.

5 Conclusions

The availability of the French 5L value set will now facilitate the development of disease-specific studies, to document health-related quality of life in the French context, which are one of the weak points of the dossiers presented to health authorities. Indeed, such studies have been delayed in the recent past by investigators because of the unavailability of a French value set. The value set is also much expected by academic health economists and clinicians. There is indeed a growing interest of the latter to include reference quality-of-life questionnaires in clinical trials and other clinical epidemiology studies.

Author Contributions GP was the project leader of this research and the main investigator. He wrote the protocol, wrote the French version of the training material and trained interviewers, supervised data collection and the statistical analysis, and wrote all versions of this article. LFA managed the data collection, trained interviewers, performed the data analysis under the supervision of Gérard de Pouvourville and the EuroQol supervising team, and wrote the Methods and Results sections of this article. KL was the EuroQol-appointed supervisor of this study, in accordance with the Code of Conduct signed between the ESSEC and EuroQol, and was the lead for bi-monthly quality control reports. MO was part of the supervision team of EuroQol and performed interim analyses of the data to identify potential inconsistencies and suggest corrections in data collection. JMRG and MO performed the models to adjust for biases in the representativeness of the French sample on age and sex. KL, MO, and JMRG revised all versions of the article and contributed to answers to the reviewers' comments.

Data Availability Statement The full data set will be made available from the corresponding author upon reasonable request, once all complementary analyses of data have been published.

Compliance with Ethical Standards

Conflicts of interest Gérard de Pouvourville was Full Professor at the ESSEC Business School at the time of this study. He has been, and is, a regular consultant for the funding companies, as a member of advisory boards on cost-effectiveness analyses. The ESSEC has also received funding to perform cost-effectiveness analysis. None of the existing funding has been related to developing value sets for quality-of-life questionnaires. The ESSEC is currently funded to perform a survey on the quality of life of French inflammatory bowel disease patients using EQ-5D-5L. The funding sources were Euroqol, Amgen, MSD, Janssen Cilag and Biogaran, through unrestricted research grants. Luiz Andrade was a research assistant at the time of the study and was employed by the ESSEC. He has no direct financial conflicts of interest. Kristina Ludwig worked as a freelancer for the EuroQol Research Foundation during the data collection phase of the study and has received research grants from the Foundation to conduct methodological research. She has received no direct payment from industrial sponsors for this study. As a collaborator of EuroQol and the main investigator of the German Value Set, she supports the EQ-5D-5L and the methodological choices made by the Foundation. Mark Oppe was an employee of the EuroQol Research Foundation at the time of the study, but now works for Axentiva Solutions, a consulting company based in Spain. He was a main contributor to the development of the EQ-5D-5L valuation methodology, specifically for the cTTO protocol, the Quality Control module, and the development of the hybrid methodology. Juan Manuel Ramos Goni is a member of EuroQol and founder of Axentiva Solutions. He has received no direct payment from industry sponsors for the French study. He also contributed to the development of the EQ-5D-5L valuation methodology, specifically for the development of the hybrid methodology.

Ethical standards This study has been declared to the Commission Nationale Informatique et Liberté (CNIL) as MR003 type research, under number 225070 v 0. The research does not come under the general field of "research organized to be applied to the human being with the aim of the development of biologic or medical knowledge" and did not require submission to a Committee for the Protection of Persons (CPP). Informed consent of respondents was requested before the interview started, and the interview process could not start if potential respondents refused to participate.

References

1. Decree n° 2012-1116, October 2, 2012, defining the health economics missions of the High Health Authority. French Republic Offical Gazette n° 0231, October 4 2012, page 15222, text n° 8. https://www.legifrance.gouv.fr/.

2. Decision n°2013.0111/DC/SEESP, September 18, 2013: Collge of the High Health Authority relative to the significant impact on expenditures for the National Sickness Fund triggering the need for an economic assessment of health products claiming an improvement/expected of Medical Serivce Rendered of I, II or III. https://www.has-sante.fr/portail/upload/docs/application/pdf/2013-09/c_2013_0111_definition_impact_significatif.pdf.

3. Haute Autorité de Santé (HAS). Choices in methods for economic evaluation. Department of Economics and Public Health Assessment. Haute Autorité de Santé; Oct 2012. https://www.has-sante.fr/portail/upload/docs/application/pdf/2012-10/choices_in_methods_for_economic_evaluation.pdf.

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

5. Costet N, Le Galès C, Buron C, Kinkor F, Mesbath M, et al. French cross-cultural adaptation of the Health Utilities Indexes mark 2 (HUI2) and 3 (HUI3) classification systems. Qual Life Res. 1998;7:245–56.

6. Le Gales C, Buron C, Costet N, Rosman C, Slama G. Développement d'un index d'états de santé par les utilités en population française: le Health Utilities Index. Economie et prévision. 2001;150–1(4–5):71–87.

7. Hamers FF, Ghabri S, Le Gales C. Health-state utility estimates for health technology assessment: a review of the manufacturers' submissions to the French National Authority for Health. Expert Rev Pharmacoecon Outcomes Res. 2017;17(5):489–94.

8. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. PharmacoEconomics. 2018;36:675–97.

9. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care. 2007;45:259–63.

10. Janssen MF, Simon Pickard A, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22:1717–27.

11. Agborsangaya BC, Lahtinen M, Cooke T, Johnson JA. Comparing the EQ-5D 3L and 5L: measurement properties and association with chronic conditions and multimorbidity in the general population. Health Qual Life Outcomes. 2014;12(74):4.

12. Feng Y, Devlin N, Herdman M. Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? Health Qual Life Outcomes. 2015;13:171.

13. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17(4):445–53.

14. Stolk E, Ludwig K, Rand K, van Hout B, Ramos Goñi JM. Overview, update and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. Value Health. 2019;22(1):23–30.

15. Commission for Health Economics and Public Health Assessments (CEESP). Minutes of Tuesday, June 13, 2017 session. https://www.has-sante.fr/portail/upload/docs/application/pdf/2017-09/compte-rendu_ceesp_13_juin_2017.pdf.

16. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708–15.

17. Oppe M, van Hout B. The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. EuroQol Working Paper Series Number 17003, October 2017. https://euroqol.org/wp-content/uploads/2016/10/EuroQol-Working-Paper-Series-Manuscript-17003-Mark-Oppe.pdf.

18. Janssen MF, Oppe M, Versteegh M, Stolk E. Introducing the composite time trade-off: a test of feasibility and face validity. Eur J Health Econ. 2013;14(Suppl 1):S5–S13.

19. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ. 2011;20:348–61.

20. Wong E, Shah K, Cheung AWL, Wong AYK, Visser M, Stolk E. Evaluation of split version and feedback module on the improvement of time trade-off data. Value Health. 2018;21(6):732–41.

21. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. Value Health. 2016;20:466–73.

22. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, et al. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med Care. 2017;55(7):e51–8.

23. Ramos-Goñi JM, Craig B, Oppe M, van Hout B. Combining continuous and dichotomous responses in a hybrid model. 2016. EuroQol Working Paper Series #16002. https://euroqol.org/

24. National Institute for Statistics and Economic Studies (INSSE). Total population age and sex, January 1st, 2018. https://www.insee.fr/fr/statistiques/1892086?sommaire=1912926.

25. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727.

26. Ludwig K, Graf vonder Schulenburg JM, Greiner W. German Value Set for the EQ-5D-5L. Pharmacoeconomics. 2018;36:663.

27. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ5D-5L value set for England. Health Econ. 2017;27(1):7–22.

28. Versteegh MM, Vermeulen KM, Evers S, Ardine de Wit, Prenger R, et al. Dutch tariff for the five-level version of EQ-5D. Value Health. 2016;19:343-352.

29. Hobbins A, Barry L, Kelleher D, Shah K, Devlin N, Ramos Goni JM, et al. Utility values for health states in Ireland: a value set for the EQ-5D-5L. PharmacoEconomics. 2018;36(11):1345–53.

30. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. Value Health. 2018;21(5):596–604.

31. Hernandez AM, Wailoo A, Grimm S, Pudney S, Gomes M, et al. EQ-5D-5L versus EQ-5D-3L: the impact on cost effectiveness in the United Kingdom. Value in Health. 2018;21(1):49–56.