TITLE: EQ-5D value set based scaling factor models for estimating value sets for EQ-PSO and relationship bolt-ons

AUTHORS: Zhihao Yang[1], Aureliano Finch[2], Kim Rand[3], Nan Luo[4]

1. Health Services Management Department, Guizhou Medical University, China.
2. EuroQol office, the Netherlands
3. Health Services Research Center, Akershus University Hospital, Lørenskog, Norway
4. Saw Swee Hock School of Public Health, National University of Singapore.

ABSTRACT:

Introduction: Bolt-on dimensions can enhance the descriptive system of the EQ-5D by capturing additional treatment benefits not covered by the core five dimensions. Previous studies have shown improved psychometric properties with bolt-ons, but conventional valuation studies for bolt-ons are cost-prohibitive. The partially-fixed "scaling factor" model, which uses existing value set parameters adjusted with a scale parameter, offers a cost-effective alternative. This study aimed to evaluate the feasibility and performance of the scaling factor model in valuing bolt-on dimensions for the EQ-5D in China and Italy, using both cTTO and hybrid designs.

Methods: Two bolt-ons, EQ-PSO (itching and self-confidence) and social relationships, were selected for valuation in China and Italy, respectively. In China, 401 participants valued EQ-7D health states using cTTO. In Italy, 200 participants completed cTTO tasks, and 1000 participants completed DCE tasks for EQ-6D states. Data was modeled using standard main-effects and scaling factor approaches in which parameters from existing national value sets were used. Model performance was evaluated by predicting observed, out-of-sample health state cTTO values.

Results: The scaling factor model showed high predictive accuracy in both countries. In China, the scaling factor model using the new value set achieved similar accuracy to the standalone model. In Italy, the scaling factor model performed slightly worse than the standalone model, with marginal differences between cTTO-only and hybrid models. The scale parameters indicated a reduction in the size of coefficients for core dimensions, confirming bolt-on's compression effect observed in previous studies.

Discussions: The scaling factor model appears to be a viable and robust approach for developing bolt-on value sets. This method could facilitate cost-efficient application of bolt-ons in health economics and outcomes research when EQ-5D is considered to be inadequate due to its lack of specificity.

Introduction:

Bolt-on may be useful extensions to the EQ-5D instruments when the descriptive system is not adequate for capturing important treatment benefits, as relevant dimensions of health are not described. Many bolt-on items have been proposed (1-4) and there is evidence for improved psychometric properties of EQ-5D with the addition of bolt-on dimensions (2, 3, 5). For formal use of bolt-on items in QALY calculation, value sets are needed. While the EQ-VT protocol could technically be extended to allow bolt-on valuation, this would likely be prohibitively costly. Furthermore, fully independent bolt-on value sets could likely create discrepancies in the values assigned to the core five dimensions, creating a wide range of practical problems, and opening up to gaming. If the rank orders of values for health states defined by the core five dimensions were altered by the addition of a bolt-on dimension, different conclusions could be reached regarding the cost-effectiveness of treatment options. This would be a serious issue because resource allocation decisions might not be consistent if switching from EQ-5D without bolt-on to EQ-5D with bolt-on.

Studies exploring valuation of bolt-on were focusing on the effect of bolt-on on the valuation of EQ-5D dimensions (6, 7). This line of research assumes that interactions between EQ-5D dimensions and bolt-ons exist and hopes to characterize how the dimensions interact with each other. We have explored options using a different core assumption: considering the fact that most of the existing EQ-5D-5L value sets use a main-effects model without any interaction terms between dimensions, we assumed that interactions between bolt-on and EQ-5D dimensions, if any, have minimal effects on overall health-state values and therefore can be ignored.

Our previous studies supported this assumption and found that the most prominent effect on assigned values of adding a bolt-on was a compression of the utility decrements associated with the existing EQ-5D dimensions. Furthermore, we observed that imposing the coefficients from an existing EQ-5D value set (adjusted with a scaling parameter) improves out-of-sample predictive accuracy for bolt-on health states over a corresponding model fitted entirely to assigned bolt-on values. In the first project we elicited cTTO values from a university student sample divided into three arms, namely 29 EQ-5D-5L self-care bolt-off states, 30 EQ-5D-5L states and 31 EQ-5D-5L vision bolt-on states. Overall, the scaling factor mode demonstrated better performance in predicting bolt-on values compared with the conventional models which were estimated from the bolt-on values. In the second project we interviewed 600 members of the general public living in Guiyang, China, and asked them to value 31 EQ-5D-5L states, 32 EQ-5D-5L + cognition bolt-on states, or 32 EQ-5D-5L + vision bolt-on states. Similar to the previous study results, this study confirmed the good performance of the scaling factor mode in predicting values of two different bolt-on health states. Another Japanese study evaluated the valuation of 3 different cognition bolt-on items (cognition, remembering things, thinking clearly) respectively and found the scaling factor models outperformed the standalone models in both cognition and remembering things arms, but not in the thinking clearly arm.

We describe this approach as "partially-fixed", as the majority of the modelled parameters are taken from an existing EQ-5D value set (this approach has also been described as a 'scaling factor' model because of the scale factor). This approach has two major advantages. First, a full valuation study is not needed for developing a bolt-on value set; considering that we only need power to estimate the bolt-on-specific parameters and a scaling parameter, a small additional study would be sufficient. Furthermore, as the parameters from a national value set can be imposed, the resulting bolt-on value sets are consistent with the national value set in question. Therefore, the scaling factor mode may represent a cost-efficient approach to developing bolt-on value sets.

So far, evidence for scaling factor model is still limited. First, the two completed projects were conducted in China. It is well-known that the health preferences and cTTO data characteristics differ across countries. Second, in the two previous studies we used a control arm eliciting values for the EQ-5D without bolt-ons, which served as a proxy for the national value sets. The extent to which this approach works with national value sets remains to be determined. Third, in previous projects we only tested the scaling factor mode using vision and cognition bolt-ons. Considering different bolt-on item may have different relative importance compared with the five EQ-5D dimensions and some bolt-on dimensions may be partially overlapped with EQ-5D dimensions, it is therefore important to examine how this model could work with other bolt-on items. This is important if we want to use a standardized protocol to estimate all bolt-on value sets. Fourth, following the EQ-VT protocol, we wish to explore the performance of this approach based on cTTO alone, DCE alone, and using a hybrid approach. Hence, it would be useful to test the scaling factor mode using both cTTO and DCE data. If the scaling factor model also works with DCE data, the cost of conducting such studies would be greatly reduced.

However, the DCE data alone does not work with the scaling factor model as the DCE data does not contain the scaling information. As an alternative, the DCE data maybe used in the hybrid model.

In this study, we aimed to investigate the scaling factor model in China and Italy using both cTTO only design and hybrid design for estimating their value sets. This two-country two-bolt-on design allows us to further evaluate the performance of the scaling factor mode and its generatability.

Methods:

<u>Selection of countries</u>

The two selected countries were China and Italy. The two countries were chosen based on several considerations. First, most of the existing EQ-5D-5L value sets are from Europe and Asia and the two continents represent two main different cultures. Second, the two countries are selected to represent value sets developed using different models and valuation methods and allows to test and compare the performance of the partially-fixed model. Third, the two countries have experienced EQ-VT interviewers because of recently completed or ongoing EuroQol projects.

<u>Selection of bolt-ons</u>

Two bolt-ons were selected for this study, after discussion with the Bolt-on Special Squad and the Valuation Working Group, the EQ-PSO (8) and social relationships (9). The bolt-ons were selected as they differed on a number of methodological aspects. The EQ-PSO is a two-item bolt-on i.e., itching and self-confidence, while social relationships is a single item bolt-on. The EQ-PSO was originally developed to be used in a specific health condition and population i.e., psoriasis, while social relationships to be used across disease areas and health conditions. Both bolt-ons may partially overlap with the core EQ-5D-5L descriptive system, and therefore may interact with some of its core dimensions e.g., itching with pain/discomfort and social relationships with usual activities, differing substantially from previous bolt-ons tested i.e., vision and hearing. Finally, both bolt-ons were developed using mixed methods research and aimed at closely resemble the format of the EQ-5D-5L, meeting some of the criteria reported in Mulhern et al (10). This study was intended as a first assessment of the feasibility of using the scaling factor model in presence of these differences but was not designed to isolate and test each of them individually.

Experimental design

This study tested the scaling factor model for a two-item bolt-on i.e., itching and self-confidence in China, and social relationships in Italy. The health states valued in China had 7 dimensions and was referred as EQ-7D hereafter; the health states valued in Italy had 6 dimensions and was referred as EQ-6D hereafter. These modified versions were approved by the EuroQol office, for methodological research only. Valuation of these bolt-on items was conducted using both cTTO and DCE methods in general public samples of China and Italy respectively.

Health states selection

In China, only cTTO method was used to value the EQ-5D-5L health states with psoriasis bolt-on items. The health states were selected. Both cTTO designs were derived from the orthogonal arrays, with the EQ-6D health states selected following the design of a previous study and had 25 health states plus 5 mild states and the pits state (555555); the EQ-7D states selected from an orthogonal array of 11 dimensions. We kept the first 7 dimensions and further added 7 mild states and the pits state (5555555). EQ-6D states were divided into 2 blocks and the EQ-7D states were divided into 4 blocks, with each block contained 15 or 16 health states.

In Italy, the DCE design was constructed using a D-efficient Bayesian design in Ngene. We first created a candidate set of 2500 pairs with 2-dimension overlap and used the Federov algorithm to select 288 pairs using the coefficients of the Italian value set as priors. The priors of the social relationship bolt-on were set as the means of the five dimensions.

Sampling and the interviews

In China, we used quota sampling to achieve a representation sample of China following the 7th National population census. Interviews were conducted in Guizhou, Guangdong, Shandong, Heilongjiang and Shanghai. All interviewers participated in the Chinese EQ-5D-5L valuation study and had experience in conducting cTTO interview. The interviewees were recruited through a snowball sampling method by rolling out from the acquaintance of the interviewers. Upon receiving consent, the respondent was assigned to a random block of health state for a face-to-face interviews using EQ-VT (11).

A team of ten interviewers who had experience of conducting cTTO interviews were recruited and re-trained for China. EQ-VT was used (12) and the standard quality control (QC) procedure of the EuroQol Group (13) was followed. To ensure protocol compliance and identify any interviewer effects, QC reports were generated every 10 interviews and sent to the interviewers. If necessary, individual suggestions were provided to improve interview performance.

The interview procedure was consistent with the EQ-VT studies, in brief, respondents first self-reported their health using EQ-5D and then valued a randomly selected block of 15 or 16 health states. Following the EQ-VT protocol, respondents were given 2 wheelchair examples, 3 practice states (e.g., 2112112, 4422213, 4433341 for EQ-7D arm) and a block of health states for formal valuation task.

In Italy, the target sample for the cTTO data collection was 200 participants, representative of the Italian adult population. Age, gender and geographical distribution for macro-areas (north-east, north-west, centre, south and islands) were used as quotas. Recruitment and interviews scheduling was organized by Pepe Research s.r.l., a survey agency experienced in quantitative and qualitative healthcare research. Possible responders were identified through a panel and a network of local recruiters. Phone reminders were sent the day prior to the interview to each participant.

Data were collected between October 2022 and November 2022 by three interviewers, using computer assisted personal interviews (CAPI) administered through Zoom. Each interviewer participated in a VC administered training. All interviewers were MSc students. The training familiarized interviewers with the EQ-5D, the EQ-VT protocol, the elicitation techniques, and the content of the QC reports. After the training, each interviewer conducted 10 practice interviews with family and friends. The fieldwork was implemented in batches of 10 interviews per interviewer, with interviewer performance being assessed using EQ quality control procedures. In the interviews, responders first completed the EQ-5D-5L, the EQ VAS and a familiarization session. Subsequently, 17 formal cTTO questions for the EQ-5D-5L and relationships were presented.

For the DCE data collection, the target sample size was 1000 respondents, representative of the Italian general adult population. Quotas were used for age, gender and distribution of geographic macro-areas. Each respondent was presented with 14 choice tasks comparing two health states, A and B, both being EQ-6D health states.

Data were collected through an online survey programmed by Maths in Health. Respondents were recruited online via the same survey company used for scheduling the cTTO data collection, Pepe Research s.r.l.. Respondents were presented an information sheet and asked to indicate informed consent. Subsequently, they responded to socio-demographic questions and were then presented with the choice tasks. The layout of the DCE task was the same as the one used in the EQ-VT data collection. Respondents received a gift voucher upon survey completion.

Data quality was assessed to exclude respondents who completed the DCE task in less than 8.3 seconds per task on average, as done in previous DCE studies of the EQ-5D-5L and often in DCE studies more in general (14), and by assessment of unusual patterns.

Model Construction and Data Analysis

By country, we estimated a standard main-effective model and a scaling factor model using the cTTO data only. For EQ-6D, we further estimated these two models using hybrid approach. To estimate the scaling factor model, the existing value sets were used. For EQ-6D, the existing Italian value set was published in 2022 (15). For EQ-7D, we used both the published Chinese value set (16) and the new Chinese value set (not published yet). The data collection for the new Chinese value set was completed in 2022, using a more representative sample of Chinese general population. The same quotas were used in the new Chinese value set study, whereas the old Chinese value set study was conducted only in urban areas. The new Chinese value set differed with the published Chinese value set in terms of the dimension ranking, suggesting a preference change over the last decade in China. For EQ-6D, we used the additive model (Model equation not shown) following the model choice of the published Italian value set. Since both DCE data and cTTO data were collected for EQ-6D, we tested both the cTTO only model, DCE only model and hybrid model. For EQ-7D, we used the CALE (Formula 1) model following the model choice of the published Chinese value set.

$$
\begin{aligned}
\text{EQ-7D disutility} = \alpha + {} & (\beta_{MO}x_{MO2}+\beta_{SC}x_{SC2}+\beta_{UA}x_{UA2}+\beta_{PD}x_{PD2}+\beta_{AD}x_{AD2}+\beta_{IT}x_{IT2}+\beta_{CO}x_{CO2})\ \beta_{L2}+ \\
& (\beta_{MO}x_{MO3}+\beta_{SC}x_{SC3}+\beta_{UA}x_{UA3}+\beta_{PD}x_{PD3}+\beta_{AD}x_{AD3}+\beta_{IT}x_{IT3}+\beta_{CO}x_{CO3})\ \beta_{L3}+ \\
& (\beta_{MO}x_{MO4}+\beta_{SC}x_{SC4}+\beta_{UA}x_{UA4}+\beta_{PD}x_{PD4}+\beta_{AD}x_{AD4}+\beta_{IT}x_{IT4}+\beta_{CO}x_{CO4})\ \beta_{L4}+ \\
& (\beta_{MO}x_{MO5}+\beta_{SC}x_{SC5}+\beta_{UA}x_{UA5}+\beta_{PD}x_{PD5}+\beta_{AD}x_{AD5}+\beta_{IT}x_{IT5}+\beta_{CO}x_{CO5}) + \varepsilon \qquad (1)
\end{aligned}
$$

For the estimation of the scaling factor model, we used the parameters of the five core dimensions from the existing value sets and fitted the data with the following parameters, i.e., β_{IT}, β_{CO}, β_{SR}, α (INTERCEPT), and λ (scale) (See Formula 2 for the scaling factor model of EQ-7D health states). We used a λ parameter to capture the scaling effects of adding psoriasis and social relationship bolt-on items, respectively, to the five core dimensions of Chinese and Italian value sets. The formulas for the additive models are not shown. Similarly, the EQ-6D DCE scaling factor was modelled using the core dimension coefficients of the published Italian EQ-5D value set. Note this model does not have the scaling parameter λ. We applied an additional step to rescale the DCE coefficients onto the EQ-6D scale since the DCE data does not contain the scaling information. To rescale the DCE coefficients, since we are looking at predictions for bolt-on cTTO data, but for the DCE only models, there are not inherent scaling to match. The DCE scaling factor model produces coefficients scaled to match the original EQ-5D value set, but not the EQ-6D cTTO data. So, the DCE only models cannot inherently predict on the EQ-6D cTTO scale. For comparison purposes, we fitted a linear transform to the DCE models, i.e. fitting cTTO value of EQ-6D = ax+b such where x is DCE predictions. Note this is just done for comparison and one cannot generate bolt-on value set using DCE data alone.

Both cTTO and hybrid models were specified to have a random-effects intercept at the level of individual study participants, using incremental dummies and to censor at -1. The coefficients in all the models are box- constrained to [0, Inf], so where non-monotonic estimates would fit better, the best available estimate will be 0.

$$\begin{aligned}\text{EQ-6D disutility} = \alpha + \lambda\,(&(\beta_{MO}x_{MO2}+\beta_{SC}x_{SC2}+\beta_{UA}x_{UA2}+\beta_{PD}x_{PD2}+\beta_{AD}x_{AD2}+\beta_{IT}x_{IT2}+\beta_{CO}x_{CO2})\,\beta_{L2}+\\ &(\beta_{MO}x_{MO3}+\beta_{SC}x_{SC3}+\beta_{UA}x_{UA3}+\beta_{PD}x_{PD3}+\beta_{AD}x_{AD3}+\beta_{IT}x_{IT3}+\beta_{CO}x_{CO3})\,\beta_{L3}+\\ &(\beta_{MO}x_{MO4}+\beta_{SC}x_{SC4}+\beta_{UA}x_{UA4}+\beta_{PD}x_{PD4}+\beta_{AD}x_{AD4}+\beta_{IT}x_{IT4}+\beta_{CO}x_{CO4})\,\beta_{L4}+\\ &(\beta_{MO}x_{MO5}+\beta_{SC}x_{SC5}+\beta_{UA}x_{UA5}+\beta_{PD}x_{PD5}+\beta_{AD}x_{AD5}+\beta_{IT}x_{IT5}+\beta_{CO}x_{CO5})) + \varepsilon \qquad (2)\end{aligned}$$

Next, we employed a by-state cross-validation approach (17, 18) to compare the performance of the scaling factor models and the standard models in predicting observed, out-of-sample health state cTTO values. The hybrid models and DCE models were evaluated in the same way. As the distribution of c-TTO data suggested that negative values were censored at -1, we used likelihood-based mean values predicted using a tobit model. We computed Pearson product-moment correlation (R), Lin concordance correlation coefficient (CCC), mean absolute errors (MAE), and root mean squared error (RMSE) between the likelihood-based values and the predicted values of each model. Since the cTTO models have the advantage of predicting the values that are used in the modelling process, we hypothesize that the cTTO alone models would perform better. We considered that the scaling factor performed acceptable if the prediction accuracy of the scaling factor model is no more than 20% of the prediction accuracy of the standalone models.

Results:

Between January and March 2024, 401 participants completed the interview in China. Table 1 summarizes the demographic information of the Chinese sample, which was representative in terms of sex, education, age, and residency. Table 2 shows the demographic information of the Italian samples, which were representative and resembled the sample of the Italian valuation study (15).

Table 1. Demographic information of the Chinese sample

TTO sample, N=401, n, %			
Sex	Male	206	51.37
	Female	195	48.63
Residency	Urban	255	63.59
	Rural	146	36.41
Age group	20-29	63	15.71
	30-39	82	20.45
	40-49	76	18.95
	50-59	83	20.7
	>60	97	24.19
Education	Low	102	25.44
	Middle	224	55.86
	High	75	18.7
Ethnicity	Han	377	94.01
	Minority	24	5.99
Working status	Employed	223	55.61
	Retired	43	10.72
	Student	44	10.97
	Farmer	57	14.21
	Others	34	8.48
Health Insurance status	Employee insurance	132	32.92
	Resident insurance	270	67.33
	Commercial insurance	61	15.21
	Other insurance	11	2.74
	No insurance	1	0.25
Health condition	No	279	69.58
	1 condition	81	20.2
	more than 2 conditions	41	10.21

Marriage status	Single	72	17.96
	Married or live together	306	76.31
	Divorced or separated	14	3.49
	Widowed	9	2.24
Household income (monthly, RMB)	<=5000	43	10.72
	5001-8000	116	28.93
	8001-12000	92	22.94
	12001-20000	63	15.71
	>20000	29	7.23
	Prefer not to say	58	14.46

Table 2. Demographic information of the Italian samples

Characteristics	TTO Sample n=202	DCE Sample n=1001
Age groups, n (%)		
18-24	15 (7.43%)	83 (8.29%)
25-34	20 (9.90%)	125 (12.49%)
35-44	34 (16.83%)	144 (14.39%)
45-54	34 (16.83%)	190 (18.98%)
55-64	30 (14.85%)	176 (17.58%)
65+	69 (34.22%)	283 (28.27%)
Gender, n (%)		
Male	97 (48.02%)	484 (48.35%)
Female	105 (51.98%)	517 (51.65%)
Geographical distribution (a)		
North-West	36 (17.82%)	257 (25.67%)
North-East	59 (29.21%)	198 (19.78%)
Centre	58 (28.71%)	201 (20.08%)
South and Islands	49 (24.26%)	345 (34.46%)
Education (b)		
Elementary	0 (0.00%)	6(0.60%)
Middle inferior	6 (2.97%)	77(7.69%)
High school	104 (51.49%)	547(54.65%)
BSc	19 (9.41%)	121(12.09%)
MSc	57 (28.22%)	197(19.68%)
Second level Masters	15 (7.43%)	36(3.6%)
PhD	1 (0.5%)	17(1.7%)
Profession		
Employed	82 (40.59%)	362(36.16%)
Self-employed	28 (13.86%)	118(11.79%)
Student	14 (6.93%)	98(9.79%)
Pensioner	56 (27.72%)	267(26.67%)
Unemployed	9 (4.46%)	74(7.39%)

Housewife	10 (4.95%)	70(6.99%)
Other	3 (1.49%)	12(1.2%)
Marital status (c)		
Single	53 (26.24%)	306(30.57%)
Married or living with partner	133 (65.84%)	591(59.04%)
Separated or divorced	10 (4.95%)	70(6.99%)
Widow	6 (2.97%)	34(3.4%)
Children (d)		
Yes	126 (62.38%)	555 (55.44%)
No	76 (37.72%)	446 (44.56%)
Number of people living in the same household (e)		
Nobody	25 (12.38%)	141(14.09%)
One	68 (33.66%)	340(33.97%)
Two	56 (27.72%)	224(22.38%)
Three	36 (17.82%)	192(19.18%)
Four or more	17 (8.42%)	104(10.39%)
Chronic conditions (f)		
No	84 (41.58%)	388 (38.66%)
Yes	118 (58.42%)	613 (61.24%)

Table 3 shows the modelling coefficients of the standalone model and two scaling factor models using the old and new Chinese value sets, respectively. The rank order of the core dimensions based on their estimated coefficients in different models differed (Table 3). For example, the rank order of the five core dimensions based on the new value set is PD, MO, AD, UA, and SC, while in the standalone model, it is PD, MO, UA, SC, and AD. The dimension weight of itching (IT) is larger than the coefficient of anxiety (AD), and self-confidence (SC) has the smallest coefficient. The predicted value for the pits state 5555555 is -0.808, -0.787, and -0.698 using the standalone model, the scaling factor model with the new value set, and the scaling factor model with the old value set, respectively.

The standalone model (f10TTO) achieved the highest prediction accuracy among these three models. The scaling factor model (f10ScaleTTOnew) using the new value set achieved slightly lower prediction accuracy. The SCALE parameter differs between these two scaling factor models, with the one using the old value set having a SCALE larger than 1.

Table 3. Modelling results of the Chinese/EQ-7D data

	f10TTO	f10ScaleTTOnew	f10ScaleTTOold
INTERCEPT6D	**0.035**	-0.035	-0.060

Model coefficients	SCALE	/	0.948	1.046
	MO	0.309	*0.299*	*0.345*
	SC	0.242	*0.241*	*0.253*
	UA	0.243	*0.279*	*0.233*
	PD	0.447	*0.447*	*0.302*
	AD	0.200	*0.289*	*0.258*
	IT	0.225	0.238	0.211
	CO	0.108	0.110	0.093
	L2	0.090	*0.110*	*0.191*
	L3	0.307	*0.356*	*0.458*
	L4	0.773	*0.704*	*0.832*
Model performance	CCC	0.988	0.985	0.968
	ICC	0.988	0.985	0.968
	MAE	0.044	0.048	0.083
	Pearson's R	0.989	0.986	0.968
	RMSE	0.059	0.066	0.098

*Coefficients in Italic font are were taken from EQ-5D-5L value sets for China; bolded coefficients were not significant at 0.05 level.

Table 4 shows the estimated model coefficients and modelling performance of the Italian data. The scale parameters were 0.810 and 0.835 for the scaling factor cTTO-only model and scaling factor hybrid model, respectively, suggesting that the size of the coefficients for the five core dimensions shrinks by around 20%. Notably, when using the cTTO-only standalone data, many coefficients were not significant at the 0.05 level, and the third level of the bolt-on dimension was inconsistent. The DCE standalone model had only two non-significant coefficients. Overall, when using the scaling factor model, the prediction accuracy for the health states' cTTO values was lower than with the standalone models. The difference among three scaling factor models are marginal, with the DCE model demonstrating slightly better performance.

Using the relative importance of the dimensions (calculated by adding up the incremental coefficients of each dimension), we also noticed that the rank order of core dimensions changed. Specifically, adding the social relationship to the EQ-5D health states also led to change of rank orders, from PD, MO, AD, SC and UA to MO SC, PD, UA and AD. The social relationship has a relative weight larger than anxiety/depression in the standalone model. This is not the case for the scaling factor model, social relationship is the least important dimension in the scaling factor model.

Table 4. Modelling results of the Italian/EQ-6D data

		f24TTO	f24ScaleTTO	f24H	f24ScaleH	f24ScaleDCE	f24DCE
	INTERCEPT6D	**0.003**	**-0.017**	**-0.004**	**-0.012**	/	/
	MO2	**0.025**	*0.051*	0.059	*0.051*	*0.051*	0.309
	MO3	**0.044**	*0.013*	0.060	*0.013*	*0.013*	0.293
	MO4	0.110	*0.180*	0.147	*0.180*	*0.180*	0.711
	MO5	0.121	*0.085*	0.145	*0.085*	*0.085*	0.710
	SC2	0.047	*0.046*	0.051	*0.046*	*0.046*	0.242
	SC3	**0.000**	*0.010*	0.025	*0.010*	*0.010*	0.141
	SC4	0.146	*0.160*	0.136	*0.160*	*0.160*	0.642
	SC5	0.055	*0.041*	0.073	*0.041*	*0.041*	0.347
	UA2	**0.032**	*0.050*	0.027	*0.050*	*0.050*	0.120
	UA3	**0.000**	*0.014*	**0.016**	*0.014*	*0.014*	0.089
	UA4	0.146	*0.161*	0.137	*0.161*	*0.161*	0.642
	UA5	**0.014**	*0.030*	0.029	*0.030*	*0.030*	0.151
	PD2	**0.025**	*0.047*	0.040	*0.047*	*0.047*	0.198
	PD3	0.057	*0.041*	0.028	*0.041*	*0.041*	0.121
	PD4	0.249	*0.265*	0.143	*0.265*	*0.265*	0.616
	PD5	0.071	*0.055*	0.058	*0.055*	*0.055*	0.263
	AD2	**0.018**	*0.044*	0.023	*0.044*	*0.044*	0.119
	AD3	**0.006**	*0.065*	0.020	*0.065*	*0.065*	0.097
	AD4	0.115	*0.209*	0.116	*0.209*	*0.209*	0.543
	AD5	**0.000**	*0.004*	0.000	*0.004*	*0.004*	**0.000**
	SR2	**0.038**	**0.039**	**0.001**	**0.000**	**0.000**	**0.000**
	SR3	**0.000**	**0.000**	0.035	0.037	0.044	0.183
Model Coefficients	SR4	0.122	0.148	0.084	0.114	0.105	0.363
	SR5	0.059	0.085	0.038	0.046	0.038	0.175
	SCALE	/	0.810	/	0.835		/
	Theta	/	/	4.815	3.178	3.186	/
	Pearson's R	0.985	0.980	0.969	0.980	0.982	0.972
	ICC	0.985	0.980	0.969	0.979	0.982	0.972
Model performance	MAE	0.050	0.059	0.071	0.060	0.054	0.066
	RMSE	0.065	0.073	0.091	0.074	0.070	0.086

*Coefficients in Italic font were taken from the published Italian value set; bolded coefficients were not significant at 0.05 level.

For the coefficients DCE models, linear transformations were applied in order to compute meaningful prediction accuracy.

Discussions:

This study further validates the potential of the scaling factor model in estimating bolt-on value sets. In addition to existing evidence, this study shows that the scaling factor model: 1) can work with a two-item bolt-on valued using the cTTO method; 2) can work with existing value sets from both Asia and Europe; 3) can work with both additive and CALE model specifications; and

4) can work both cTTO only, and hybrid approaches. Overall, the scaling factor model performed slightly worse than the standalone model in both countries. However, it is worth noticing that criterion we used in the study is the prediction accuracy of the observed bolt-on health state cTTO values. The standalone models using only the cTTO bolt-on values have the advantage of predicting the same data that were used in the modelling process.

In China, even though the rank order of the five dimensions changed between the earlier value set (MO, PD, AD, SC, UA) and the new value set (PD, MO, AD, UA, SC), the scaling factor models using either value set as a core demonstrated acceptable performance. However, when using the new value set, the prediction indicators improved and were similar to those of the standalone model. This may be because the sample of this study is more similar to the sample of the new value set study, and the preferences of these two samples are more aligned. In addition, the scale parameter is larger than 1 when using the published Chinese value set, suggesting the size of five core dimensions increased even after the bolt-on item was added. This effect has been observed in the Japanese study too. A possible explanation is that in earlier valuation studies, the sizes of the coefficients were smaller than the coefficients from the valuation studies using more recent EQ-VT interviewer protocol (19) which lead to more extreme cTTO values especially towards the lower end.

In Italy, the scaling factor model outperformed the standalone model when the hybrid model was used. This highlights the value of the scaling factor model, as most of its parameters are based on existing value sets which were estimated using a much larger dataset (i.e. cTTO and DCE data collected from 1000 individuals). In contrast, three of the 20 coefficients were insignificant in the standalone hybrid model, which suggests that, to achieve a similarly robust parameter estimation, a standalone valuation study would require a larger design, which means a larger population sample or higher respondent burden, and increased cost. The worse and better performance of the scaling factor model compared to the standalone model when only the cTTO and DCE data was used suggests that DCE data was the driver of the better performance of the scaling factor model when both cTTO and DCE data was used. This finding seems supporting the use of both cTTO and DCE data when the scaling factor modelling approach is used to estimate bolt-on value sets. Such a design could further reduce study costs since the data collection cost for DCE is much lower than that for cTTO. As this is the only study reporting the use of scaling factor model using DCE data, future study is needed to further test the DCE method.

It can be observed that the coefficient of anxiety/depression decreased significantly after adding the EQ-PSO bolt-on. This violates the assumption of the scaling factor model that no interaction exists between the bolt-on item and the five core dimensions. A similar effect was observed in Italy: when the social relationship bolt-on was added, the coefficients for usual activities and anxiety/depression decreased, making anxiety/depression the least important dimension. These changes in the coefficients of core EQ-5D dimensions may suggest overlap in the health problems measured by EQ-5D and bolt-on items. Indeed, itch is one type of discomfort that has been covered by pain/discomfort. Despite the clear impact of the bolt-on on core dimensions, the scaling factor model still worked, suggesting its robustness to interactions introduced by bolt-ons. This is an important finding as interaction was a major concern in estimating bolt-on value sets. Our study suggests that it might not be a concern if the scale factor model and existing EQ-5D value sets are used, though more research is needed to test this approach with other bolt-ons.

So far, all studies have shown that the scaling factor model results in similarly or more accurate estimations of bolt-on health state values compared to standalone models. These studies used a wide range of bolt-ons, including vision, cognition (various variants), EQ-PSO, and social relationships. Evidence has been collected from China, Japan, and Italy, with ongoing studies in the United Kingdom and Poland using OPUF and DCE with duration methods. Notably, both the completed and ongoing studies have utilized a full study design, meaning both the cTTO design and the DCE design allow for the estimation of a standalone main-effects model. The purpose of this is to enable comparison between the standalone model and the scaling factor model. For estimating values sets for a one-item bolt-on, we may actually use smaller designs, as only two additional parameters need to be estimated if the scaling factor model is used with the CALE specification. However, future studies are needed to determine the most cost-effective size of design for such studies.

Reference:

1. Finch AP, Brazier J, Mukuria C, et al. An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: the EQ-5D case study. Value Health. 2018; (
2. Wang P, Chong SL, Luo N. A hearing bolt-on item increased the sensitivity of the EQ-5D-5L in a community-based hearing loss screening program. 2017 EuroQol Academy discussion paper. 2017; (
3. Luo N, Wang X, Ang M, et al. A Vision "Bolt-On" Item Could Increase the Discriminatory Power of the EQ-5D Index Score. Value Health. 2015; 18(8): 1037-42.
4. Yang Y, Rowen D, Brazier J, et al. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. Value Health. 2015; 18(1): 52-60.

2024 EuroQol plenary discussion paper.

5. Arons A, Krabbe P. Considering Cognition as a Bolt-on Dimension for the EQ-5D. EuroQol Plenary Meeting 2011 Discussion Papers. 2011; (
6. Yang Y, Brazier J, Tsuchiya A. Effect of adding a sleep dimension to the EQ-5D descriptive system: a "bolt-on" experiment. Med Decis Making. 2014; 34(1): 42-53.
7. Yang Z, Rand K, Busschbach J, et al. Modelling TTO values of vision bolt-on and self-care bolt-off health states: can bolt-on and bolt-off value sets be built upon EQ-5D value set? , EuroQol Plenary Meeting 2019. Brussels, Belgium, 2019.
8. Swinburn P, Lloyd A, Boye KS, et al. Development of a disease-specific version of the EQ-5D-5L for use in patients suffering from psoriasis: lessons learned from a feasibility study in the UK. Value Health. 2013; 16(8): 1156-62.
9. Finch AP, Brazier J, Mukuria C. Selecting Bolt-on Dimensions for the EQ-5D: Testing the Impact of Hearing, Sleep, Cognition, Energy, and Relationships on Preferences Using Pairwise Choices. Med Decis Making. 2021; 41(1): 89-99.
10. Mulhern BJ, Sampson C, Haywood P, et al. Criteria for developing, assessing and selecting candidate EQ-5D bolt-ons. Qual Life Res. 2022; 31(10): 3041-48.
11. Stolk E, Ludwig K, Rand-Hendriksen K, et al. Overview, update and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. Manuscript Submitted for Publication, 2018.
12. Yang Z, Luo N, Oppe M, et al. Towards a smaller design for EQ-5D-5L valuation studies. Submitted for publication. 2019; (
13. Ramos-Goni JM, Oppe M, Slaap B, et al. Quality Control Process for EQ-5D-5L Valuation Studies. Value Health. 2017; 20(3): 466-73.
14. Jonker MF, Roudijk B, Maas M. The Sensitivity and Specificity of Repeated and Dominant Choice Tasks in Discrete Choice Experiments. Value in Health. 2022; 25(8): 1381-89.
15. Finch AP, Meregaglia M, Ciani O, et al. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. Soc Sci Med. 2022; 292(114519.
16. Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L Value Set for China. Value Health. 2017; 20(4): 662-69.
17. Rand-Hendriksen K, Ramos-Goni JM, Augestad LA, et al. Less Is More: Cross-Validation Testing of Simplified Nonlinear Regression Model Specifications for EQ-5D-5L Health State Values. Value Health. 2017; 20(7): 945-52.
18. Yang Z, Rand K, Busschbach J, et al. Cross-Attribute Level Effects Models for Modeling Modified 5-Level Version of EQ-5D Health State Values: Is Less Still More? Value in Health. 2023; 26(6): 865-72.
19. Stolk E, Ludwig K, Rand K, et al. 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.