ORIGINAL RESEARCH ARTICLE

EQ-5D-5L Valuation for the Malaysian Population

Asrul Akmal Shafie[1] · Annushiah Vasan Thakumar[1] · Ching Jou Lim[1] · Nan Luo[2] · Kim Rand-Hendriksen[3] ·
Faridah Aryani Md Yusof[4]

Abstract

Objectives The aim of this study was to develop an EQ-5D-5L value set reflecting the health preferences of the Malaysian adult population.

Methods Respondents were sampled with quotas for urbanicity, gender, age, and ethnicity to ensure representativeness of the Malaysian population. The study was conducted using a standardized protocol involving the EuroQol Valuation Technology (EQ-VT) computer-assisted interview system. Respondents were administered ten composite time trade-off (C-TTO) tasks and seven discrete choice experiment (DCE) tasks. Both linear main effects and constrained non-linear regression models of C-TTO-only data and hybrid models combining C-TTO and DCE data were explored to determine an efficient and informative model for value set prediction.

Results Data from 1125 respondents representative of the Malaysian population were included in the analysis. Logical consistency was present in all models tested. Using cross-validation, eight-parameter models for C-TTO only and C-TTO + DCE hybrid data displayed greater out-of-sample predictive accuracy than their 20-parameter, main-effect counterparts. The hybrid eight-parameter model was chosen to represent the Malaysian value set, as it displayed greater out-of-sample predictive accuracy over C-TTO data than the C-TTO-only model, and produced more precise estimates. The estimated value set ranged from -0.442 to 1.

Conclusions The constrained eight-parameter hybrid model demonstrated the best potential in representing the Malaysian value set. The presence of the Malaysian EQ-5D-5L value set will facilitate its application in research and health technology assessment activities.

✉ Asrul Akmal Shafie
  aakmal@usm.my

1 Discipline of Social and Administrative Pharmacy,
   Universiti Sains Malaysia, Penang, Malaysia

2 Saw Swee Hock School of Public Health, National University
   of Singapore, Singapore, Singapore

3 Health Services Research Centre, Akershus University
   Hospital, Lørenskog, Norway

4 Pharmacy Practice and Development Division, Ministry
   of Health Malaysia, Selangor, Malaysia

Key Points

A non-linear, constrained hybrid model was shown to be a viable statistical technique generating health state values from combined time trade-off and discrete choice experiment preference data.

The results are comparable to linear main-effects models, but with improved out-of-sample predictive accuracy.

The non-linear, constrained hybrid model was chosen to represent the Malaysian EQ-5D-5L value set, which has a range of -0.442 to 1.

1 Introduction

The EuroQol five-dimension inventory (EQ-5D), a preference-based instrument designed to capture health-related quality of life (HRQoL), is the most-used instrument worldwide in economic evaluations measuring health benefit in terms of quality-adjusted life-years (QALYs) [1, 2]. It is recommended by several health technology assessment agencies, the most prominent of which is the National Institute for Health and Care Excellence of England and Wales (NICE). The EQ-5D instrument describes health problems along five dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The original version [3], now generally referred to as the EQ-5D-3L, describes each dimension at three levels, roughly corresponding to no problems, some problems, and extreme problems. A newer version, known as the EQ-5D-5L, has been released and is gaining popularity [1]. As the name suggests, the EQ-5D-5L describes each dimension of health at five severity levels [4], corresponding to no, slight, moderate, severe, and unable/extreme problems.

The EQ-5D-5L instrument comprises two main components: the descriptive system, and value sets [5]. The descriptive system is a questionnaire allowing respondents to describe their health along the five dimensions, as well as presenting their overall health on a standardized, vertical visual analog scale, the EQ-VAS. Each unique combination of responses to the descriptive system (sans the EQ-VAS) is called as an *EQ-5D health state*. The health states are conventionally referred to using a five-digit short-hand, such that each digit represents the level of functioning for each dimension in the previously presented order. Thus, state 11111 reflects no problems on any dimension, and state 55555 reflects extreme problems on all five dimensions. For responses in terms of EQ-5D health states to be useful in QALY calculation, values reflecting the relevant populations' preferences for each health states must be assigned. By convention, such utility values are presented on a scale where 1 reflects the preference for full health, and 0 reflects the preference for not being alive. An EQ-5D value set comprises such values for all 3125 EQ-5D health states.

Value sets for EQ-5D instruments have conventionally been generated through population valuation studies in which representative samples of the general public have been administered different tasks to value a subset of the possible EQ-5D health states, after which statistical modeling is used to generate a full set of values. Previous valuation studies [6, 7] for EQ-5D-3L suffered from substantial variation in elicitation methods. To improve comparability and ensure greater quality, the EuroQol Group

has developed a standardized protocol for conducting EQ-5D-5L valuation studies, involving the use of a computer-assisted interview system called EuroQol Valuation Technology (EQ-VT) [8]. EQ-VT and the standardized protocol are both updated to reflect best practice based on available evidence. For the first time, this valuation tool was utilized in the Malaysian setting to systematically derive a value set for the EQ-5D-5L instrument. This value set should have high comparability to other EQ-VT-based value sets due to the similar protocols employed.

The standardized protocol has been used in EQ-5D-5L valuation studies involving various countries, including Canada [9], Uruguay [10], England [11], the Netherlands [12], Korea [13], Japan [14], Spain [15], China [16], Hong Kong [17], Indonesia [18], Germany [19], Thailand [20], and Ireland [21], while studies are underway in countries including the United States and Singapore. As implemented methodologies are becoming more standardized, value set differences among countries, even those belonging to similar continents, are apparent and might be attributed, in part, to cultural values shared by people of a country. This is especially evident in the variability of preference weights set on different dimensions in different settings, stressing the need for a localized value set for economic evaluation use.

The EQ-5D is commonly used in the Malaysian healthcare setting [22], but the lack of a local EQ-5D-5L value set may limit its scope of usage. Having a local value set ensures that health effects or outcomes are valued according to the preferences of the concerned population. The use of a standardized and thoroughly tested valuation protocol and standard technology improves quality control and ensures greater comparability with other studies. The objective of this study was to derive a Malaysian EQ-5D-5L value set using the standardized protocol and the EQ-VT.

2 Methods

2.1 Sampling and Recruitment

A sample size of 1000 is recommended by the EuroQol Group for valuation studies aiming to establish an EQ-5D-5L value set [23]. To take into account possible study attrition, a total of 1200 Malaysians were targeted in this study. The eleven states located in the Peninsular Malaysia were divided into four regions (Northern, Central, Eastern, Southern), with each region considered as a cluster. One urban and one rural area from each cluster was then randomly selected using simple random sampling. The clusters included in the study were Penang (Northern), Selangor (Central), Kelantan (Eastern), and Malacca (Southern).

Respondents were recruited from the respective areas using quota-based sampling with quotas set for

urbanicity (urban/rural), gender (male/female), age (18–39/40–64/> 64 years), and ethnicity (Indigenous/Non_Indigenous) based on the Malaysian National Census [24]. Individuals aged > 18 years who understood English or Malay were recruited to participate in this study while those with low cognitive levels (visible signs of mental impairment) or who were unable to understand the instructions were excluded. Informed consent was obtained from all eligible respondents.

2.2 Data Collection

Eighteen trained interviewers (final-year pharmacy or medical undergraduate students) administered the valuation instrument. They moved from one region to the next after completing the required number of interviews. The interviewers were stationed at selected gathering points (e.g., shopping complexes, markets, food courts, or community halls) in targeted areas. Eligible subjects encountered in nearby public places were approached for study consent based on convenience sampling and pre-set quotas.

Prior to the field work, each interviewer attended a 3-day training workshop conducted by experienced EuroQol representatives and the Malaysian investigator team. Components of the training were co-designed by EuroQol Group and the Malaysian team. These included an introduction to the background and rationale of the study; detailed review of interview tasks using specifically designed visual materials, giving exposure to health preference elicitation methods; practical interviewing techniques; project-specific interviewer's tasks; demonstration of the interview procedures; practice interview with fellow participants, and a 'mock interview' with the general public. In addition, the interviewers performed a minimum of five practice interviews as part of their training. Only interviewers that satisfied the training requirements were allowed in the field.

Quality control (QC) was carried out following the official guidelines using the EQ-VT QC tool [25] to closely and continuously monitor interviewers throughout the whole data collection process. In addition, all interviewers were required to report to the project supervisors on a daily basis.

2.3 Valuation Protocol

The valuation tasks were administered using the EQ-VT application, adhering to the valuation protocol set forth by the EuroQol Group [8]. Briefly, the interview had three parts: (1) self-reported health using the descriptive EQ-5D-5L and warm-up tasks leading to the valuation tasks; (2) valuation tasks involving completing ten composite time trade-off (C-TTO) tasks and subsequently seven discrete choice experiments (DCEs); and (3) country-specific questions including socio-demographics of respondents.

The C-TTO method has been described in detail elsewhere [26]. C-TTO applies the classic approach for health states deemed better than dead, whereby the respondent would find an indifference point between x years in full health (Life A) and 10 years in a specific health state (Life B), with the value of the impaired health state being $x/10$. If the respondent is faced with a health state where he or she perceives that dying immediately is preferable to being in the health state for 10 years (worse than dead, WTD), a slightly different scenario, called lead-time TTO, is used instead. In this case, the respondent must still find an indifference point between two lives: Life A consisting of x years in full health, and Life B consisting of 10 years in full health followed by 10 years in the health state. The health state is then valued as $(x-10)/10$. The current EQ-VT protocol employs a total of 86 health states for direct elicitation with C-TTO in ten blocks of ten health states each. One of the five very mild health states (i.e., having only slight problems in one of the five dimensions) and the pits state (55555) were present in every block, leaving 80 states (eight health states × ten blocks) chosen by Monte Carlo simulation for inclusion in the standardized EQ-VT C-TTO design. These 80 unique states were distributed over the ten blocks ensuring mean utility is approximately equal across blocks [23].

After valuing ten health states using C-TTO, respondents were administered a set of seven DCE tasks, in which they were asked to choose between two different EQ-5D-5L health states with no duration specified. DCEs measure health state preference indirectly, on a latent scale of relative attractiveness of one health state over another, without undergoing an iterative process to reach a point of indifference, as in C-TTO [27, 28]. As both C-TTO and DCE aim to measure the same construct (health-state preference), the variability in both methods may offer information regarding the utility of the measured health states. The current EQ-VT employs 196 different DCE pairs, organized in 28 blocks of seven pairs each, with the blocks relatively balanced in terms of severity. Further information about the selected DCE pairs can be found elsewhere [23]. Each respondent was randomly assigned one C-TTO block and one DCE block, and for both, the order of presentation was randomized.

2.4 Data Analysis and Modeling

Four main candidate models were extensively tested for performance: 20-parameter linear main-effect models; and eight-parameter non-linear constrained models; both fitted to C-TTO data only (ADD20_TTO, MULT8_TTO), or to C-TTO and DCE data in a hybrid model design (ADD20_HYBRID, MULT8_HYBRID).

The 20-parameter model employs one parameter for each of levels 2 through 5 for each dimension. Let α represent the intercept, x_{dl} the dummy variable indicating the presence of

problems on dimension d at level l, and β_{dl} the coefficient representing the estimated disutility of having problems on dimension d at level l (e.g., βMO3 representing the disutility of having moderate problems [level 3] on mobility). The mathematical function of the 20-parameter model is as follows:

$$
\begin{aligned}
y = \alpha + \Sigma_l \Sigma_d \beta_{dl} x_{dl} + e = \alpha \\
+ \beta_{MO2} x_{MO2} + \beta_{SC2} x_{SC2} + \beta_{UA2} x_{UA2} + \beta_{PD2} x_{PD2} + \beta_{AD2} x_{AD2} \\
+ \beta_{MO3} x_{MO3} + \beta_{SC3} x_{SC3} + \beta_{UA3} x_{UA3} + \beta_{PD3} x_{PD3} + \beta_{AD3} x_{AD3} \\
+ \beta_{MO4} x_{MO4} + \beta_{SC4} x_{SC4} + \beta_{UA4} x_{UA4} + \beta_{PD4} x_{PD4} + \beta_{AD4} x_{AD4} \\
+ \beta_{MO5} x_{MO5} + \beta_{SC5} x_{SC5} + \beta_{UA5} x_{UA5} + \beta_{PD5} x_{PD5} + \beta_{AD5} x_{AD5} + e.
\end{aligned}
$$

The constrained eight-parameter models fit a single parameter per dimension (β_{MO}, β_{SC}, β_{UA}, β_{PD}, β_{AD}), taking a value representing level 5; and one parameter for each of levels 2, 3, and 4 (L_2, L_3, L_4), which are multiplied by the respective dimension parameters. For example, the disutility of having moderate problems (level 3) on mobility is $\beta_{MO} \times L_3$. The mathematical function is as described below, where x_{dl} still represents the dummy variable representing the presence of problems on dimension d at level l.

$$
\begin{aligned}
y = \alpha + \Sigma_l (\Sigma_d \beta_d x_{dl}) L_l + e = \alpha \\
+ (\beta_{MO} x_{MO2} + \beta_{SC} x_{SC2} + \beta_{UA} x_{UA2} + \beta_{PD} x_{PD2} + \beta_{AD} x_{AD2}) L_2 \\
+ (\beta_{MO} x_{MO3} + \beta_{SC} x_{SC3} + \beta_{UA} x_{UA3} + \beta_{PD} x_{PD3} + \beta_{AD} x_{AD3}) L_3 \\
+ (\beta_{MO} x_{MO4} + \beta_{SC} x_{SC4} + \beta_{UA} x_{UA4} + \beta_{PD} x_{PD4} + \beta_{AD} x_{AD4}) L_4 \\
+ \beta_{MO} x_{MO5} + \beta_{SC} x_{SC5} + \beta_{UA} x_{UA5} + \beta_{PD} x_{PD5} + \beta_{AD} x_{AD5} + e.
\end{aligned}
$$

The constrained models were tested based on previous studies demonstrating that these may produce improved out-of-sample predictions [16, 29]. C-TTO-only models were fitted using mixed-effects models with random intercepts at the level of individual study participants.

Hybrid models were also explored. In these models, C-TTO and DCE data were combined as one data file to find a common vector of coefficients. This modeling technique was originally proposed by Oppe and van Hout [30] and has been successful in producing logically consistent predictions in a number of studies [15, 17–19, 31]. The hybrid modeling method uses a joint maximum likelihood procedure, in which the same set of parameters is used to obtain likelihood estimates over C-TTO data assuming a normal distribution, and DCE data using conditional logit. As DCE produces relative distances between health states on a latent scale, a scaling parameter between DCE and C-TTO is introduced. Since standard errors (SEs) and confidence intervals (CIs) are not readily available from non-linear models, bootstrapping was used. All models were fitted to each of the 10,000 individual respondent-level bootstrap samples.

In addition to the eight-parameter and 20-parameter models, various other models and extensions were tested

but discarded due to lack of monotonicity or greatly reduced model performance. All models were tested with intercepts (separate intercepts for C-TTO and DCE in the hybrid cases). Intercepts were removed if they were statistically non-significant.

All data were analyzed as disutilities (1 − utility score) using statistical software STATA version 12 (StataCorp, College Station, TX, USA) and R statistical package, version 3.3.0 (R Foundation for Statistical Computing, Vienna, Austria) [32].

2.5 Model Performance

Model performance was evaluated based on two criteria. The primary criterion evaluated is monotonicity; logical consistency of the parameters with respect to the hierarchical structure of levels within the five dimensions. Parameters of increasing severity (of a dimension) should have higher disutility values (e.g., value$_{MO2}$ < value$_{MO3}$). The second criterion was out-of-sample predictive accuracy over mean C-TTO health state values. This was tested using a cross-validation approach [33, 34]; sequentially splitting the dataset into two subsets, fitting the models to one set, using the fitted models to predict the other set, and comparing the predicted and observed values. The splitting method tested here is known as the leave-out-by-block method [29]. Respondents were excluded from both C-TTO and DCE data depending on their assigned C-TTO block. One block was sequentially left out, serving to evaluate prediction accuracy of the fitted model using the remaining nine blocks of health states. Cross-validation as demonstrated previously [16, 29] has shown to be useful in revealing cases in which improved fit on observed data does not correspond to improved out-of-sample predictive accuracy. Mean square error (MSE), mean absolute error (MAE), Lin concordance correlation coefficient (CCC), intra-correlation coefficient (ICC), and Pearson product-moment correlation (R) were used to assess how similar predicted mean health state values are compared with mean observed values. Better predictive accuracy is indicated by lower MSE and MAE values, and higher CCC, ICC, and R values.

2.6 Exclusion Criteria

Respondents whose C-TTO value increased with health-state severity and those who valued −1 on all health states were excluded on the basis that these choices were not rational, and the respondents may not have understood the time trade-off process.

3 Results

3.1 Respondents

A total of 1137 out of the 1172 (97.01%) consenting respondents successfully completed interviews in eight locations in the four randomly selected Malaysian states: Penang (Bayan Baru, Balik Pulau), Kelantan (Kubang Kerian, Rantau Panjang), Melaka (Bandar Melaka, Jasin), and Selangor (Subang Jaya, Sungai Buloh). The remaining 35 respondents asked to discontinue the study as they either did not understand the valuation process or were not comfortable with the time trade-off process, or the forced choice presentation of the DCE. Twelve interviews met the exclusion criteria, leaving data from 1125 respondents to be included in the data analysis process. Table 1 outlines the demographic characteristics of included respondents.

While respondents were largely representative of the general population in terms of gender, age, ethnicity, and residential area, those categorized as being employed full time/self-employed were slightly under-represented.

3.2 Data Characteristics

The constant term for the 20-parameter model fitted to C-TTO data was statistically significantly different from 0. The constant was therefore retained in the C-TTO models. For the hybrid, the constant terms were non-significant, and were therefore dropped. Table 2 presents the parameter estimates for the four models tested in this study. In terms of the characteristics of values generated by the models (Table 3), hybrid models predicted slightly lower 55555 scores than their C-TTO-only model counterparts. Among the 3125 EQ-5D-5L health states, the proportion for which negative values were predicted was roughly equal between the tested models, with

Table 1 Study sample characteristics ($N=1125$)

	n (%)	Malaysian general population (%)	Differences (%)	p value
Gender[a]				0.903
Male	576 (51.2)	51.5	−0.3	
Female	549 (48.8)	48.5	+0.3	
Age, year; mean (SD)[a]				0.536
18–39	632 (56.2)	53.9	+2.3	
40–64	405 (36.0)	37.5	−1.5	
≥65	88 (7.8)	8.6	−0.7	
Ethnicity[a]				0.596
Indigenous	770 (68.44)	67.4	+1.04	
Non-indigenous	355 (31.56)	32.6	−1.04	
Residential area[a]				0.685
Urban	790 (70.2)	71.0	−0.8	
Rural	355 (29.8)	29.0	+0.8	
Employment status[b]				<0.001
Employed full time/self-employed	629 (55.9)	64.3	−8.4	
Unemployed (able to work)	48 (4.3)	3.4	+0.9	
Outside labor force[c]	313 (37.8)	32.0	+5.8	
Others	14 (1.2)			
Missing	9 (0.8)			
Level of education		N/A	N/A	
Primary or lower	68 (6.0)			
Secondary	398 (35.4)			
College diploma	229 (20.4)			
Degree	352 (31.3)			
Postgraduate	78 (7.0)			

[a]Data are based on results of the 2010 Malaysian National Census

[b]Data are based on the 2016 Malaysian Labour Force Survey Report

[c]The category of 'outside labor force' includes those with sickness or disability, caretakers of households, students, and the retired

[d]Chi-square test

Table 2 Parameter estimates of the tested models

| Model | TTO | | | | HYBRID | | | |
| | ADD20_TTO | | MULT8_TTO | | ADD20_HYBRID | | MULT8_HYBRID | |
	Coefficient	SE*	Coefficient	SE*	Coefficient	SE*	Coefficient	SE*
MO			0.327	0.011			0.340	0.008
SC			0.248	0.011			0.261	0.008
UA			0.205	0.011			0.202	0.007
PD			0.349	0.011			0.338	0.009
AD			0.287	0.011			0.300	0.009
L2			0.214	0.024			0.239	0.010
L3			0.315	0.023			0.316	0.012
L4			0.789	0.019			0.766	0.012
MO2	0.067	0.011	0.070	0.009	0.086	0.007	0.081	0.004
MO3	0.084	0.012	0.103	0.009	0.113	0.008	0.108	0.005
MO4	0.235	0.013	0.258	0.011	0.250	0.009	0.261	0.008
MO5	0.325	0.012	0.327	0.011	0.350	0.010	0.340	0.008
SC2	0.069	0.011	0.053	0.007	0.070	0.008	0.062	0.003
SC3	0.105	0.012	0.078	0.006	0.092	0.009	0.083	0.004
SC4	0.224	0.013	0.196	0.010	0.220	0.010	0.200	0.007
SC5	0.266	0.012	0.248	0.011	0.257	0.008	0.261	0.008
UA2	0.044	0.012	0.044	0.006	0.049	0.007	0.048	0.003
UA3	0.069	0.012	0.065	0.006	0.068	0.009	0.064	0.003
UA4	0.168	0.013	0.162	0.009	0.160	0.008	0.155	0.006
UA5	0.202	0.012	0.205	0.011	0.199	0.008	0.202	0.007
PD2	0.048	0.010	0.075	0.009	0.070	0.007	0.081	0.004
PD3	0.080	0.012	0.110	0.009	0.090	0.008	0.107	0.005
PD4	0.259	0.011	0.275	0.011	0.243	0.009	0.259	0.008
PD5	0.332	0.013	0.349	0.011	0.335	0.009	0.338	0.009
AD2	0.040	0.012	0.061	0.008	0.068	0.008	0.072	0.004
AD3	0.077	0.013	0.090	0.008	0.097	0.009	0.095	0.005
AD4	0.210	0.012	0.226	0.010	0.230	0.010	0.230	0.007
AD5	0.282	0.011	0.287	0.011	0.298	0.009	0.300	0.009
Intercept	0.031	0.012	0.016	0.012				

*Bootstrapped-based SEs with 10,000 resamples performed at the level of individual study respondents

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

ADD20_HYBRID additive 20-parameter hybrid model, *ADD20_TTO* additive 20-parameter C-TTO only model, *C-TTO* composite time trade-off, *MULT8_HYBRID* multiplicative eight-parameter hybrid model, *MULT8_TTO* multiplicative eight-parameter C-TTO only model, *SE* standard error

Table 3 Comparison of score characteristics among models

Characteristics of scores	ADD20_TTO	MULT8_TTO	ADD20_HYBRID	MULT8_HYBRID
Lowest score	−0.439 (55555)	−0.432 (55555)	−0.440 (55555)	−0.442 (55555)
Median score	0.337	0.332	0.338	0.336
Second highest score	0.929 (11112)	0.940 (11211)	0.951 (11211)	0.952 (11211)
No. of negative scores	282 (9.02%)	286 (9.15%)	283 (9.06%)	281 (8.99%)
Ranking of dimensions	PD-MO-AD-SC-UA	PD-MO-AD-SC-UA	MO-PD-AD-SC-UA	MO-PD-AD-SC-UA

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

ADD20_HYBRID additive 20-parameter hybrid model, *ADD20_TTO* additive 20-parameter C-TTO only model, *C-TTO* composite time trade-off, *MULT8_HYBRID* multiplicative eight-parameter hybrid model, *MULT8_TTO* multiplicative eight-parameter C-TTO only model

MULT8_TTO and ADD20_TTO having the highest (9.15%) and lowest (9.02%), respectively. Dimension ranking differed slightly in the preference of the top two, with C-TTO-based models favoring pain/discomfort, whilst the hybrid models tilted towards mobility.

3.3 Modeling Data

Logical consistency was present in all the four estimated models whereby higher severity parameters had higher disutility values. Subsequent cross-validation analysis revealed the multiplicative models exhibiting slightly better predictive accuracy (lower MSE and MAE; higher CCC, ICC, and R) when compared with their additive model counterparts (MULT8_TTO vs ADD20_TTO; MULT8_HYBRID vs ADD20_HYBRID). A similar trend of performance was observed in all predictive accuracy measures when comparing hybrid and C-TTO-only models, with the MULT8_HYBRID model slightly outperforming the other three models by all the prediction accuracy standards (Table 4). Thus, the MULT8_HYBRID model was chosen as it best represents the Malaysian EQ-5D-5L value set with a value range of -0.442 to 1. The predicted values of the 3125 health states, together with SEs and 95% CIs are available in CSV (comma-separated values) format in the Electronic Supplementary Material. Figure 1a illustrates the range of the predicted 3125 health state values while Figs. 1b and 2 display how similar the predictions of MULT8_HYBRID are compared with the observed C-TTO health state values and rescaled DCE health state values, respectively.

In calculating the predicted index value for the health state '12345' using the chosen MULT8_HYBRID model's value set, the following formula is used:

$$\text{Utility weight}('12345') = 1 - \text{MO1}(0) - \text{SC2}(0.062) - \text{UA3}(0.064)$$
$$- \text{PD4}(0.259) - \text{AD5}(0.300)$$
$$= 0.315.$$

4 Discussion

A Malaysian EQ-5D-5L value set was developed from the health preferences of 1125 Malaysians using a constrained eight-parameter hybrid model fitted to C-TTO and DCE data. The respondents were representative of the general population of Malaysia in terms of urbanicity, gender, age, and ethnicity.

Constrained, non-linear models were explored in this study as they have demonstrated better out-of-sample predictive accuracy of C-TTO values in a past study that compared linear main-effect models and constrained non-linear modeling using four-country datasets [29]. Conceptually, the eight-parameter non-linear model used in this study is a special case of the 20-parameter model, in which the relative distance between levels is constrained such that it is the same for all five dimensions. With fewer parameters and less complexity, the risk of overfitting is reduced, and parameters can be estimated with improved accuracy [33]. The observation that the constrained models produce better out-of-sample predictions suggests that any real deviation from the underlying assumption of shared relative level difference between dimensions is less substantial than the noise introduced by allowing the model to fit separate level parameters for each dimension. This was also demonstrated in our study, where cross-validation analysis revealed a slightly better fit of constrained models with the observed C-TTO values.

On the other hand, hybrid modeling has also gained a reputation for modeling utility values, as it overcomes logical inconsistencies [15, 19] encountered when C-TTO or DCE data are modeled individually. Although our models did not suffer from logical inconsistencies, hybrid modeling was also included as it incorporates utility responses from two distinct tasks, potentially resulting in a more complete view on the pattern of utility response. The two hybrid models tested in this study had on average a wider range of utility values and smaller bootstrapped SEs when compared with their C-TTO-only model counterparts. Importantly, the inclusion of DCE data in the modeling improved out-of-sample predictive accuracy over left-out

Table 4 Cross validation of various models

Statistic	ADD20_TTO	MULT8_TTO	ADD20_HYBRID	MULT8_HYBRID
Mean square error	0.0042	0.0037	0.0035	0.0034
Mean absolute error	0.0521	0.0484	0.0468	0.0455
Lin CCC	0.9860	0.9877	0.9883	0.9886
ICC	0.9860	0.9877	0.9883	0.9886
Pearson R	0.9860	0.9877	0.9883	0.9886

ADD20_HYBRID additive 20-parameter hybrid model, *ADD20_TTO* additive 20-parameter C-TTO only model, *CCC* concordance correlation coefficient, *C-TTO* composite time trade-off, *ICC* intra-class correlation coefficient, *MULT8_HYBRID* multiplicative eight-parameter hybrid model, *MULT8_TTO* multiplicative eight-parameter C-TTO only model
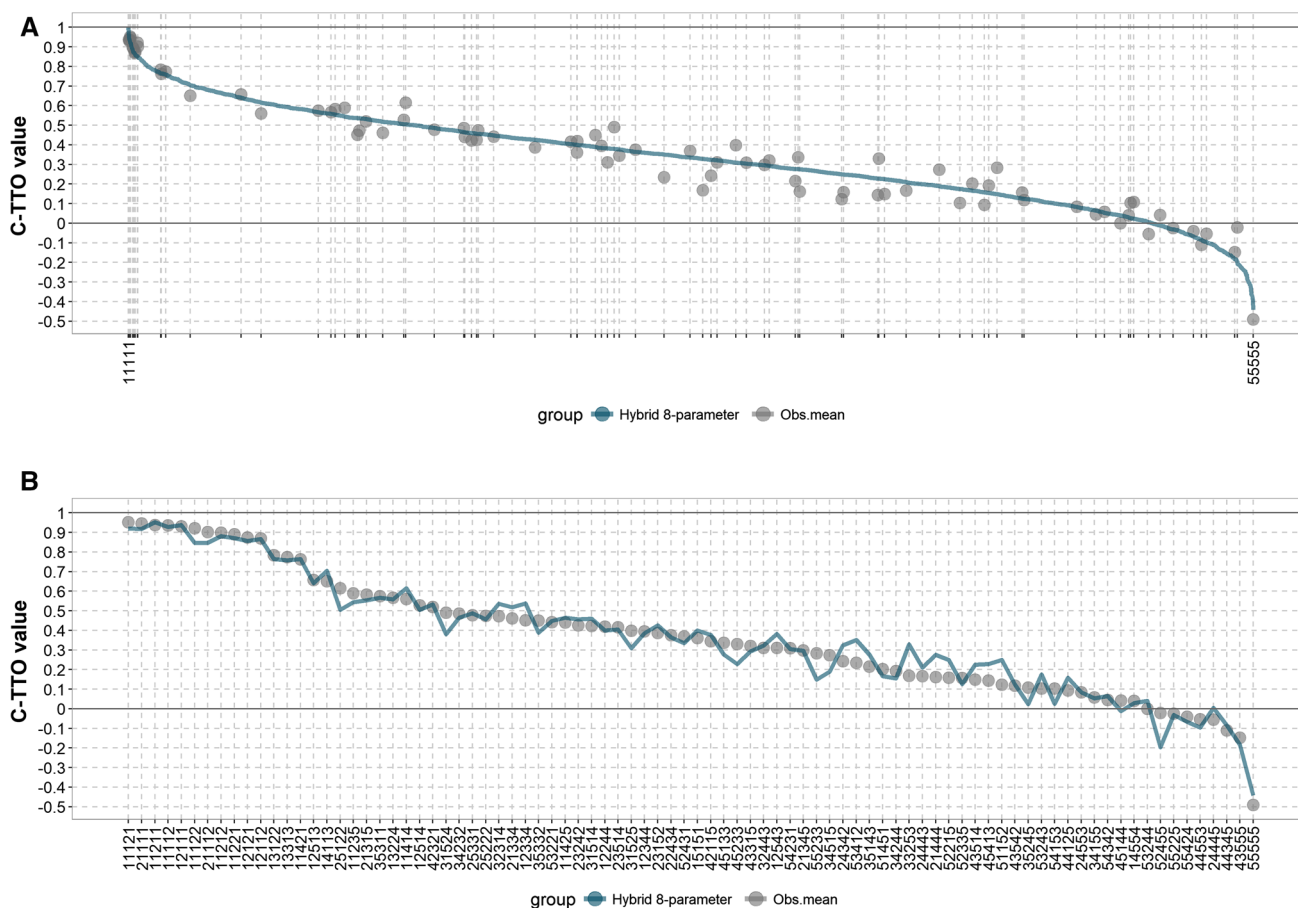
Fig. 1 Observed and MULT8_HYBRID predicted values for the EQ-5D-5L health states. **a** Values for the whole 3125 health states by descending order of MULT8_HYBRID predicted values. **b** Values for 86 health states by descending order of observed mean values. *C-TTO* composite time trade-off, *EQ-5D-5L* five-level EuroQol five-dimensional questionnaire, *MULT8_HYBRID* multiplicative eight-parameter hybrid model

C-TTO health state means. This has not been demonstrated in any published hybrid model to date. This supports the notion that C-TTO and DCE tap into the same underlying preference structure, at least in this specific valuation study. We know from secondary analyses of other existing EQ-5D-5L valuation study data that this is not universally true for already published EQ-5D-5L valuation studies. When we perform the same analyses on data from already published EQ-5D-5L valuation studies from other countries, we find that inclusion of DCE data tends to reduce out-of-sample predictive accuracy over left-out TTO data (results not presented here).

Interestingly, while pain/discomfort was the dimension with greatest impact on assigned values in the C-TTO-only models, mobility had relatively greater importance in the hybrids. This subtle but important difference was captured by the combination of response patterns from two conceptually different preference elicitation techniques trying to measure the same thing: matching tasks of the TTO and choice-based tasks of the DCE [15, 30, 35].

Model fit between individual observations and predicted values were not used as a performance indicator in this study considering that results would generally favor increased model complexity. As the validity of predicted values outside the scope of observed data cannot be truly assessed, cross-validation is an attractive proxy [16, 29]. The issue of hybrid models not having apparent counterfactual observed values may be a concern, as the cross-validation procedure was based on C-TTO blocks and values [15]. In this study, predictive accuracy over C-TTO means was chosen as the most important metric, and consequently, cross-validation was performed at the level of C-TTO blocks. If greater emphasis was placed on DCE, some metric of accuracy over DCE data could be chosen as a criterion, and cross-validation using DCE blocks could be employed. A combination approach would also be possible but would potentially indicate different models as superior for different criteria.

The best-performing model in this study was the constrained, non-linear eight-parameter hybrid, which displayed the greatest out-of-sample predictive accuracy over C-TTO

Fig. 2 Comparison of observed rescaled discrete choice experiment (DCE) and MULT8_HYBRID predicted values for the EQ-5D-5L health states included in the EQ-VT design. *EQ-5D-5L* five-level EuroQol five-dimensional questionnaire, *EQ-VT* EuroQol Valuation Technology, *MULT8_HYBRID* multiplicative eight-parameter hybrid model
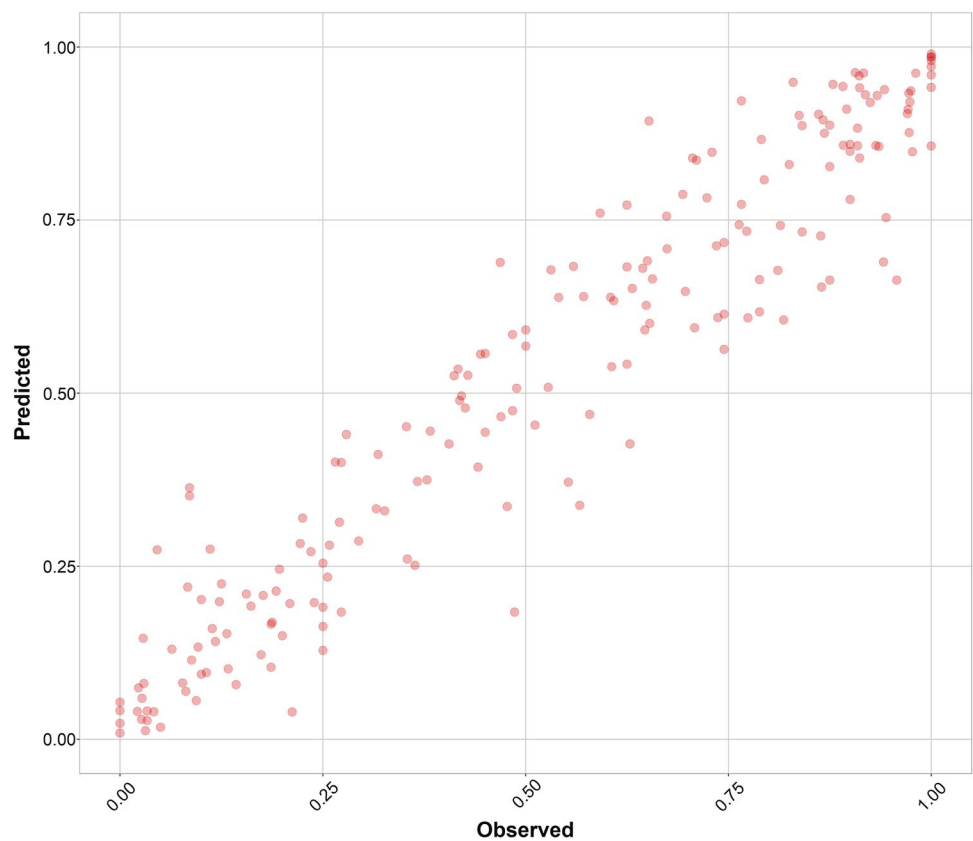
health state mean, produced the widest range of health state values, and the smallest SEs among all models. The highest disutility weight was placed on the mobility dimension and the lowest on usual activities, which is similar to China [16], Hong Kong [17], and Thailand [20]. Many Asian countries [13, 14, 16–18, 20] having published EQ-5D-5L value sets to date place the highest emphasis on the mobility dimension. The similarities in ranking of dimensions usually end there as cultural and socioeconomic factors come into play to shape the preferences of the people. Indonesia and Malaysia are neighbors, and the citizens of the two countries share similar backgrounds but display quite distinct value set characteristics. 'Pain and discomfort' was ranked a close second in terms of preference in the Malaysian value set but was surprisingly the least important dimension from the Indonesian perspective. Indonesia also has a health state 55555 value of −0.865, which was significantly lower than Malaysia's. These differences further support the need for country-specific value sets to better capture the health preferences of the population.

There are a number of study limitations to be considered. Malaysia as a country consists of Peninsular Malaysia (79.6% of the population, 11 states, two Federal Territories) in the West and Borneo island (20.4% population, two states, one Federal Territory) in the East. Our study only involved sampling clusters in Peninsular Malaysia due to resource constraints. Generally, the focus of economic development in Peninsular Malaysia has resulted in better job and study opportunities, leading to the mobility of East Malaysians to the West. Conducting our interviews in gathering spots such as hospitals, colleges, and market areas allowed us to maximize the respondent characteristic sampling. Accordingly, subsequent studies focusing on valuation data of East Malaysia residents should be conducted in the future to assess if there are any significant differences in the preference patterns compared with people residing in West Malaysia. Also, due to the nature of the sampling strategy, we do not have a response rate to report. There were potential respondents who approached the interviewer booths expressing interest in the study while there were those who declined when first approached but came back at a later time. This was common in rural areas where they had to finish chores before returning when they had free time. Based on the similarities of the sample characteristics with the general population, we believe the representativeness of Malaysians in terms of age, gender, ethnicity, and urbanicity is accounted for. However, those employed full-time or self-employed may be slightly under-represented and those outside the labor force over-represented in our study. Further analysis comparing the mean difference in observed C-TTO values between these two groups was not statistically significant, indicating the slight skew in employment status representativeness

was not a major factor for concern. Finally, in reflecting the number of respondents who were excluded during the data analysis stage, there was a slight learning effect observed. Most of the excluded respondents were from the first cluster. However, the quality control monitoring process [25], together with experience, aided interviewers in improving the interview process, leading to only 1.1% (12 respondents) being excluded in total.

5 Conclusion

This study demonstrated that the constrained eight-parameter hybrid model has the best potential in representing the Malaysian value set. The establishment of the Malaysian EQ-5D-5L value set will facilitate its application in health-related research, especially in the Malaysian health technology assessment scene.

Compliance with Ethical Standards

Conflict of interest Nan Luo and Kim Rand-Hendriksen are members of the EuroQol Research Foundation. Asrul Akmal Shafie, Nan Luo, and Kim Rand-Hendriksen have received grants from EuroQol Research Foundation. There is no other conflict of interest.

Ethical approval The study received ethical approval from the Malaysia Medical Research & Ethics Committee (ID NMRR-13-1377-18574) and was conducted in accordance with the Declaration of Helsinki.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. Appl Health Econ Health Policy. 2017;15(2):127–37. https://doi.org/10.1007/s40258-017-0310-5.
2. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics. 2014;32(4):367–75.
3. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53–72.
4. Herdman M, Gudex C, Lloyd A, Janssen M, 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(10):1727–36. https://doi.org/10.1007/s11136-011-9903-x.
5. Brooks R, Rabin R, De Charro F. The measurement and valuation of health status using EQ-5D: a European perspective: evidence from the EuroQol BIO MED research programme. Berlin: Springer Science & Business Media; 2003.
6. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. Value Health. 2009;12(8):1194–200. https://doi.org/10.1111/j.1524-4733.2009.00581.x.
7. Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. Med Decis Mak. 2014;34(1):8–20. https://doi.org/10.1177/0272989x13480852.
8. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health J Int Soc Pharmacoecon Outcomes Res. 2014;17(4):445–53.
9. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. Med Care. 2016;54(1):98–105. https://doi.org/10.1097/mlr.0000000000000447.
10. Augustovski F, Rey-Ares L, Irazola V, Garay O, Gianneo O, Fernández G, et al. An EQ-5D-5L value set based on Uruguayan population preferences. Qual Life Res. 2015;18(7):A810–1. https://doi.org/10.1016/j.jval.2015.09.104.
11. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Econ. 2018;27(1):7–22. https://doi.org/10.1002/hec.3564.
12. Versteegh MM, Vermeulen KM, Evers SM, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. Value Health. 2016;19(4):343–52. https://doi.org/10.1016/j.jval.2016.01.003.
13. Kim S-H, Ahn J, Ock M, Shin S, Park J, Luo N, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res. 2016;25(7):1845–52. https://doi.org/10.1007/s11136-015-1205-2.
14. Ikeda S, Shiroiwa T, Igarashi A, Noto S, Fukuda T, Saito S. Developing a Japanese version of the EQ-5D-5L value set. J Natl Inst Public Health. 2015;64(1):47–55.
15. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med Care. 2014;55(7):e51–8. https://doi.org/10.1097/mlr.0000000000000283.
16. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. Value Health. 2017;20(4):662–9. https://doi.org/10.1016/j.jval.2016.11.016.

17. Wong ELY, Ramos-Goñi JM, Cheung AWL, Wong AYK, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. Patient Patient Centered Outcomes Res. 2018;11(2):235–47. https://doi.org/10.1007/s40271-017-0278-0.

18. Purba FD, Hunfeld JAM, Iskandarsyah A, Fitriana TS, Sadarjoen SS, Ramos-Goñi JM, et al. The Indonesian EQ-5D-5L value set. Pharmacoeconomics. 2017;35(11):1153–65. https://doi.org/10.1007/s40273-017-0538-9.

19. Ludwig K, Graf von der Schulenburg J-M, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics. 2018;36(6):663–74. https://doi.org/10.1007/s40273-018-0615-8.

20. Pattanaphesaj J, Thavorncharoensap M, Ramos-Goni JM, Tongsiri S, Ingsrisawang L, Teerawattananon Y. The EQ-5D-5L valuation study in Thailand. Expert Rev Pharmacoecon Outcomes Res. 2018;18(5):551–8. https://doi.org/10.1080/14737167.2018.1494574.

21. Hobbins A, Barry L, Kelleher D, Shah K, Devlin N, Goni JMR, et al. Utility values for health states in ireland: a value set for the EQ-5D-5L. Pharmacoeconomics. 2018;36(11):1345–53. https://doi.org/10.1007/s40273-018-0690-x.

22. Shafie AA. EuroQol 5-dimension measures in Malaysia. In: Michalos AC, editor. Encyclopedia of quality of life and well-being research. Springer: Berlin; 2014. p. 2041–4.

23. Oppe M, Van Hout B. The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. EuroQol working paper series. 2017. p. 17003.

24. Department of Statistics Malaysia. Population distribution and basic demographic characteristic report 2010. https://www.statistics.gov.my/index.php?r=column/ctheme&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09&bul_id=MDMxdHZjWTk1SjFzTzNkRXYzcVZjdz09. Accessed 7 May 2015.

25. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. Value Health. 2017;20(3):466–73. https://doi.org/10.1016/j.jval.2016.10.012.

26. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics. 2016;34(10):993–1004. https://doi.org/10.1007/s40273-016-0404-1.

27. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31(1):306–18. https://doi.org/10.1016/j.jhealeco.2011.11.004.

28. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value Health. 2010;13(8):1005–13.

29. Rand-Hendriksen K, Ramos-Goñi JM, Augestad LA, Luo N. 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. https://doi.org/10.1016/j.jval.2017.03.013.

30. Oppe M, van Hout B, editors. The optimal hybrid: experimental design and modeling of a combination of TTO and DCE. In: 27th scientific plenary meeting of the EuroQol group-proceedings. Rotterdam: EuroQol Group Executive Office; 2010.

31. Shiroiwa T, Ikeda S, Noto S, Igarashi A, Fukuda T, Saito S, et al. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. Value Health. 2016;19(5):648–54. https://doi.org/10.1016/j.jval.2016.03.1834.

32. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2016.

33. Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc. 1984;79(387):575–83.

34. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Stat Surv. 2010;4:40–79.

35. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Mak. 2015;35(3):328–40. https://doi.org/10.1177/0272989X14559542.