



ScienceDirect

Contents lists available at [sciencedirect.com](http://sciencedirect.com)  
Journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

Methodology

## Cross-Attribute Level Effects Models for Modeling Modified 5-Level Version of EQ-5D Health State Values: Is Less Still More?

Zhihao Yang, PhD, Kim Rand, PhD, Jan Busschbach, PhD, Nan Luo, PhD

### ABSTRACT

**Objectives:** Cross-attribute level effects (CALE) model has demonstrated better predictive accuracy for out-of-sample health states than the conventional additive main-effects model in cross-validation analysis of the 5-level version of EQ-5D (EQ-5D-5L) composite time trade-off (cTTO) datasets. In this study, we aimed to further test the performance of CALE model using a different design and modified EQ-5D-5L states.

**Methods:** A total of 29 EQ-5D-5L self-care bolt-off states, 30 EQ-5D-5L states, and 31 EQ-5D-5L vision bolt-on states were selected from the same orthogonal array. A total of 600 university students were interviewed face-to-face to value a subset of these health states using the cTTO method. For each type of health state, we fitted both the conventional main-effects model and the CALE model. Predictive accuracy was assessed in a series of cross-validation analysis using the leave-one-state-out method.

**Results:** Overall, the CALE model outperformed the conventional model for each of the 3 types of health states in predicting the cTTO values of out-of-sample health states. The prediction accuracy of using the CALE model improved with the number of dimensions in health states, for example, the MAE decreased about 24%, 67%, and 77% for the EQ-5D-5L self-care bolt-off, EQ-5D-5L, and EQ-5D-5L vision bolt-on states, respectively, when using CALE models.

**Conclusion:** Our study supported the strengths of the CALE model for modelling the utility values of both original and modified EQ-5D-5L health states. Investigators with limited resources may consider using the CALE model to lower the costs for their valuation studies for EQ-5D-5L or similar health state descriptive systems.

**Keywords:** additive model, cross-attribute level effects model, EQ-5D, value set.

VALUE HEALTH. 2023; ■(■):■-■

### Introduction

The use of EQ-5D in economic evaluation requires a value set, namely, utility values for all health states defined by the instrument. Such a value set is derived using a 2-step approach: first, a subset of the health states is directly valued by members of the general public, and second, the observed values are modeled to predict values for all health states.<sup>1,2</sup> Typically, the model used is a variant of the main-effects model used in the seminal Measurement and Valuation of Health study.<sup>3</sup> In the case of 5-level version of EQ-5D (EQ-5D-5L), the default main-effects model consists of 20 parameters, 4 dummy variables per dimension representing the 4 different levels of problems beyond level 1 (reference level).<sup>1</sup> Using this model as the core, many countries have developed their own EQ-5D-5L value sets.<sup>4,5</sup>

Recently, a constrained main-effects model called cross-attribute level effects (CALE) model was tested<sup>6</sup> and was used to estimate both EQ-5D-5L and 3-level version of EQ-5D value sets.<sup>7,8</sup> The CALE model contains the same main effects in the

conventional model but those are specified with fewer parameters. For example, 8 parameters can be used to specify the 20 main effects for modeling EQ-5D-5L values, including 5 dimension parameters, each for the level-5 problems of a different dimension, and 3 level parameters for the 3 intermediate levels: “slight” (level-2), “moderate” (level-3), and “severe” (level-4) problems. With the assumption that the ratios of the effects of the levels within a dimension are constant across dimensions, the CALE model specifies the 20 main effects in the conventional model with either a dimension parameter or the multiplication of a dimension parameter and a level parameter. In an analysis of the EQ-5D-5L valuation data collected from China, The Netherlands, Spain, and Singapore,<sup>6</sup> the 8-parameter CALE model outperformed the conventional 20-parameter main-effects model in predicting the values of out-of-sample EQ-5D-5L health states.

Although the CALE model is novel and promising, the favorable results observed in the previous study could be specific to the valuation study design, particularly, the choice of the health states for valuation. All the 4 EQ-5D-5L valuation studies used the same

protocol to value the same set of 86 health states using the composite time trade-off (cTTO) method. It is possible that the CALE model would not outperform the conventional model in out-of-sample predictions, had a different set of EQ-5D-5L health states been valued and used for cross-validation analysis. Moreover, the superior out-of-sample prediction could be specific to the (number of) dimensions of the EQ-5D health states. Last but not least, the CALE model may allow estimating an EQ-5D value set with smaller amount of data; nevertheless, this potential was not evaluated in the previous study. The currently recommended sample size of at least 1000 individuals for an EQ-5D-5L valuation study is estimated for the case of using the conventional 20-parameter model.<sup>9</sup> If the 8-parameter CALE model is used, a significantly smaller sample would be required.<sup>10</sup> Accordingly, a smaller number of interviewers and interviewing hours would be needed.<sup>11,12</sup> EQ-5D-5L valuation studies are costly also because of intensive interviewer training and monitoring necessary to ensure data quality.<sup>4,11</sup> Extant EQ-5D-5L valuation studies typically used 5 to 10 interviewers. The large general population sample and the manpower needed for collecting EQ-5D valuation data could deter countries with limited resources to develop their own value sets. Therefore, a smaller valuation study design that allows a value set to be estimated with a lower budget would be desirable.

In this study, we aimed to further test the performance of the CALE model relative to the conventional main-effects model. In particular, we compared the 2 models in modeling the cTTO values of both original and modified EQ-5D-5L health states. We tested 2 types of modified EQ-5D health states, one comprising 6 dimensions, that is, the EQ-5D-5L plus a vision dimension (hereafter referred to as “EQ + VI”), and the other comprising 4 dimensions, that is, the EQ-5D-5L without the self-care dimension (hereafter referred to as “EQ – SC”). Vision has been confirmed to be a useful bolt-on dimension,<sup>13</sup> whereas self-care has been found to be the least important EQ-5D dimension in several populations.<sup>5,14</sup> The choice of the dimensions for bolt-on and bolt-off was primarily for assessing the potential of the CALE model for health state descriptive systems that are simpler or more complex than the original EQ-5D system.

## Methods

### Experimental Design

We designed a 2-arm valuation study. The first arm was for participants to value 10 EQ – SC health states followed by 10 or 11 EQ-5D-5L states; the second arm was for participants to value 10 or 11 EQ-5D-5L states followed by 11 EQ + VI health states. EQ – SC health states were developed by simply removing the self-care descriptor and the descriptor of vision bolt-on in the EQ + VI followed the descriptors of the mobility, self-care, and usual activities, that is “no/slight/moderate/severe problems” seeing for the first 4 levels and “unable to” for the fifth level.

We used an existing orthogonal array of  $6 \times 5$  to select 29 EQ – SC, 30 EQ-5D, and 31 EQ + VI states for valuation in this study.<sup>15</sup> This orthogonal array has 6 columns, each representing a different dimension; therefore, the array could be used to select EQ + VI states. To derive the EQ – SC states, we took out the second and last columns of the orthogonal array. Similarly, to derive the EQ-5D states, we took out the last column of the orthogonal array. The orthogonal array remains orthogonal when one or more columns are taken out. Given that the columns in an orthogonal design are open to code, we chose one variant of the design which contained the state 55555 and the most plausible states.<sup>16</sup> Based on previous studies, this design would allow

estimation of values for all EQ – SC, EQ-5D, and EQ + VI health states.<sup>10</sup>

We blocked the EQ – SC, EQ-5D, and EQ – SC states each into 3 blocks using Alg-algorithm in R studio software, with the numbers of states per block being 10 for EQ – SC, 10 or 11 for EQ-5D, and 11 for EQ + VI. We then combined the 3 EQ – SC blocks with the 3 EQ-5D blocks to create blocks 1, 2, and 3 for valuation in the first arm. Similarly, we created block 4, 5, and 6 by combining the EQ-5D and EQ + VI blocks for valuation in the second arm (see all health state blocks in Appendix 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2022.12.012>). Each block has 20 to 22 EQ-5D health states. The first arm of 300 respondents valued the 10 EQ – SC + 10/11 EQ-5D, and the second arm of another 300 respondents valued the 10/11 EQ-5D + 11 EQ + VI. The respondents were reminded of the added dimension by interviewers.

### Participants

We used a student sample drawn from Guizhou Medical University, China. First- and second-year students from the Schools of Public Health, Health and Medicine Management, Stomatology, and Medical Humanities were recruited using advertisement. The inclusion criteria were those who (1) gave informed consent and (2) did not participate in a cTTO valuation interview before. Following sample size considerations put forward by Oppe et al.,<sup>9</sup> we targeted a sample size of 600 participants with 300 for each study arm.

### The Interviews

All consenting participants were interviewed in a designated interview room in the campus to value the health states. The interviews were conducted using the EQ-PVT tool, a Microsoft PowerPoint-based version of EQ-VT v1.1. Each participant was paid 100 RMB (equivalent to 14 euros) upon completion. The Institutional Ethic Review Board of the Guizhou Medical University did not consider an ethic application necessary given that this study was a non-interventional, interview-based study.

All the interviews consisted of 3 sections. First, the study background was introduced. Second, the participants provided personal background information and reported their health states using the EQ-5D or EQ – SC descriptive system. Third, participants were randomized to value 1 block of the EQ – SC and EQ-5D states (study arm 1) or 1 block of the EQ-5D and EQ + VI states (study arm 2). Before start, each participant was familiarized with the cTTO valuation tasks by valuing 3 hypothetical health states (ie, being in wheelchair, a health state worse than being in wheelchair, and a health state better than being in wheelchair) and 3 EQ – SC or EQ-5D practice states. The cTTO valuation task was described in detail elsewhere.<sup>17</sup> Briefly, health states perceived as better than dead by respondents is valued with a 10-year time frame, whereas health states considered worse than dead is values with a 20-year time frame (ie, 10 years in full health followed by 10 years in the health state for valuation). For each health state, an iterative elicitation procedure is used to identify the indifferent point between life A (ie, x years of full health followed by death) and life B (ie, 10 years in the health state for valuation followed by death or followed by 10 years in full health and then death). The value of the health state is given by  $x/10$  for better-than-dead states or  $(x - 10)/10$  for worse-than-dead states.

All interviews were conducted by 8 trained interviewers including 3 graduate students and 5 third-year undergraduate students. All the interviewers had conducted cTTO valuation tasks using EQ-PVT in a previous study and were trained again before data collection in this study. Interviews were conducted in

accordance with current EuroQol guidelines for quality control (QC) in EQ-VT studies. For every 10 consecutive interviews completed by an interviewer, a QC report was generated by Z.Y. to indicate possible protocol violation and provide individualized suggestions for improvement.

### Data Analysis

By arm and type of health states, we fitted the cTTO data into both the conventional and CALE models, including the 16-, 20-, and 24-parameter models for EQ – SC, EQ-5D, and EQ + VI and the 7-, 8-, and 9-parameter CALE models for EQ – SC, EQ-5D, and EQ + VI, respectively. Below we showed the formula of the random effects conventional additive model (formula 1) and CALE model (formula 2) for EQ + VI health states as a reference. To account for the left-censored at  $-1$  nature of cTTO values, all models were censored at  $-1$  by using a tobit estimator. For each model specification, we applied a fixed intercept model for all participants and a random intercept model with random intercepts at the level of individual study participants to account for the panel nature of the data.

$$\begin{aligned} \text{EQ+VI disutility}_{ij} = & \alpha_i + \beta_1 \text{MO}_{2j} + \beta_2 \text{MO}_{3j} + \beta_3 \text{MO}_{4j} + \beta_4 \text{MO}_{5j} + \beta_5 \text{SC}_{2j} + \beta_6 \text{SC}_{3j} + \beta_7 \text{SC}_{4j} + \\ & \beta_8 \text{SC}_{5j} + \beta_9 \text{UA}_{2j} + \beta_{10} \text{UA}_{3j} + \beta_{11} \text{UA}_{4j} + \beta_{12} \text{UA}_{5j} + \beta_{13} \text{PD}_{2j} + \beta_{14} \text{PD}_{3j} + \beta_{15} \text{PD}_{4j} + \\ & \beta_{16} \text{PD}_{5j} + \beta_{17} \text{AD}_{2j} + \beta_{18} \text{AD}_{3j} + \beta_{19} \text{AD}_{4j} + \beta_{20} \text{AD}_{5j} + \beta_{21} \text{VI}_{2j} + \beta_{22} \text{VI}_{3j} + \beta_{23} \text{VI}_{4j} + \beta_{24} \text{VI}_{5j} + \varepsilon_{ij} \end{aligned} \quad (1)$$

$$\begin{aligned} \text{EQ+VI disutility}_{ij} = & \alpha_i + (\beta_{\text{MO} \times \text{MO}_{2j}} + \beta_{\text{SC} \times \text{SC}_{2j}} + \beta_{\text{UA} \times \text{UA}_{2j}} + \beta_{\text{PD} \times \text{PD}_{2j}} + \beta_{\text{AD} \times \text{AD}_{2j}} + \beta_{\text{VI} \times \text{VI}_{2j}}) \\ & \text{L}_{2j} + (\beta_{\text{MO} \times \text{MO}_{3j}} + \beta_{\text{SC} \times \text{SC}_{3j}} + \beta_{\text{UA} \times \text{UA}_{3j}} + \beta_{\text{PD} \times \text{PD}_{3j}} + \beta_{\text{AD} \times \text{AD}_{3j}} + \beta_{\text{VI} \times \text{VI}_{3j}}) \\ & \text{L}_{3j} + (\beta_{\text{MO} \times \text{MO}_{4j}} + \beta_{\text{SC} \times \text{SC}_{4j}} + \beta_{\text{UA} \times \text{UA}_{4j}} + \beta_{\text{PD} \times \text{PD}_{4j}} + \beta_{\text{AD} \times \text{AD}_{4j}} + \beta_{\text{VI} \times \text{VI}_{4j}}) \\ & \text{L}_{4j} + (\beta_{\text{MO} \times \text{MO}_{5j}} + \beta_{\text{SC} \times \text{SC}_{5j}} + \beta_{\text{UA} \times \text{UA}_{5j}} + \beta_{\text{PD} \times \text{PD}_{5j}} + \beta_{\text{AD} \times \text{AD}_{5j}} + \beta_{\text{VI} \times \text{VI}_{5j}}) + \varepsilon_{ij} \end{aligned} \quad (2)$$

$u_i \sim \text{iid } N(0, \sigma_u^2)$  is a individual level random intercept;  $\varepsilon_{ij} \sim \text{iid } N(0, \sigma_\varepsilon^2)$  is an error term.

The interpretation of the additive model is described in detail in published EQ-5D valuation studies. In the CALE model, each dimension has 1 dimension parameter. For example, 5 dimension parameters were used to represent the disutility of having problems at level 5 on each dimension ( $\beta_{\text{MO}}$ ,  $\beta_{\text{SC}}$ ,  $\beta_{\text{UA}}$ ,  $\beta_{\text{PD}}$ , and  $\beta_{\text{AD}}$ ) for EQ-5D health state; a set of scalars L2-L4 (ie, level parameters) was estimated to identify where the cutoffs of the intermediate levels were located for all dimensions. iid is independent and identically distributed,  $\varepsilon$  is the error term,  $i$  is the respondent, and  $j$  accounts for the panel structure of the data set. To calculate the decrement of a dimension at certain level, one needs to multiply the level parameter with the dimension parameter. For example, the disutility of level 2 on mobility was  $\beta_{\text{MO}} \times \text{L}_2$  and the disutility of level 5 on mobility was  $\beta_{\text{MO}}$ .

Next, we used a cross-validation procedure similar to the one used in a previous study of CALE models<sup>6</sup> to compare the performance of the CALE and conventional models in prediction of out-of-sample health state values. This procedure sequentially left

out part of the data for out-of-sample validation and used the remainder for model estimation. We performed 2 cross-validation analyses by leaving out (1) 1 state and (2) 1 block of health states sequentially for all states and blocks. To compare prediction accuracy, we calculated the mean absolute errors (MAE), root mean squared errors (RMSE), Pearson product-moment correlation (Pearson R), Intraclass correlation coefficient and Lin's concordance correlation coefficient using observed and model predicted values for the left-out states part. The observed mean was also calculated using tobit estimator. All analyses were conducted in R studio and Stata. As in the cross-validation process, there are multiple models; we chose to report the coefficients of full models (all data were used) for reference.

### Results

In total, 601 students participated in the study and finished the valuation tasks. On average, the participants spent 36.8 minutes (SD: 10.8) on all the cTTO tasks including 6 practice tasks and considered 6.8 life durations (SD: 2.0) before reaching the

indifferent point in the tasks. Figure 1 shows the histograms of the cTTO values by arm and type of health states. In general, the value distributions are similar between arms and types of health states. In all histograms, there is a “spike” at  $-1$ , suggesting that some values may be lower than  $-1$  if the cTTO procedure had been designed to allow such values.

All coefficients were significant at 0.05 level for CALE models (Table 1). For conventional additive models (Table 2), there were some insignificant coefficients, which all occurred on second level, except for the coefficient of the third level vision, which is also the only nonmonotonic coefficient. From Table 1, it can be seen these 2 arms of students had slightly different preference; for example, the rank order of dimension weights was different across 2 arms. Nevertheless, between 2 sets of coefficients within each arm, there was an agreement in terms of rank order of dimension weight and level weight similarity.

Table 3 shows the cross-validation results of the leave-one-state-out analysis. In overall, the CALE model outperformed the conventional model in 3 types of health states across 2 arms. The relative decrease of MAE/RMSE ranged between 24% and 115%

**Table 1.** Model coefficients of CALE models.

Parameters	First arm		Second arm					
	EQ – SC, n = 2999		EQ-5D_1, n = 3200		EQ-5D_2, n = 3209		EQ + VI, n = 3298	
	Coef, SE	Coef, SE	Coef, SE	Coef, SE	Coef, SE	Coef, SE		
Intercept	0.053, 0.027	0.008, 0.023	0.052, 0.023	0.018, 0.022				
VI				0.388, 0.018				
MO	0.341, 0.017	0.299, 0.016	0.293, 0.016	0.266, 0.017				
SC		0.299, 0.016	0.256, 0.016	0.214, 0.017				
UA	0.386, 0.017	0.358, 0.017	0.316, 0.016	0.253, 0.017				
PD	0.404, 0.017	0.358, 0.017	0.397, 0.016	0.330, 0.017				
AD	0.476, 0.017	0.444, 0.017	0.416, 0.016	0.379, 0.017				
L2	0.113, 0.025	0.135, 0.024	0.079, 0.025	0.146, 0.022				
L3	0.263, 0.021	0.242, 0.020	0.265, 0.020	0.260, 0.020				
L4	0.757, 0.021	0.748, 0.021	0.736, 0.021	0.705, 0.021				
Log likelihood	–1431.92	–1463.1	–1421.7	–1489.9				

Note. All coefficients were significant at 0.05 level.

AD indicates anxiety/depression; CALE, cross-attribute level effects; Coef, coefficient; EQ – SC, EQ-5D-5L without the self-care dimension; EQ + VI, EQ-5D-5L plus a vision dimension; EQ-5D-5L, 5-level version of EQ-5D; L2, level 2; L3, level 3; L4, level 4; Mo, mobility; PD, pain/discomfort; SC, self-care; UA, usual activities.

when using the CALE model (with a random intercept) compared with the conventional model (with a random intercept). Compared with a fix intercept, using a random intercept marginally improved the performance for the CALE model for all 4 arms but led to mixed results for the conventional model (Table 3). In either model, as the number of dimensions increased, the MAE/RMSE increased and the Intraclass correlation coefficient/concordance correlation coefficient/Pearson C decreased. Similar results were observed in the leave-one-block-out cross-validation analysis (Table 4).

## Discussion

This study provided new evidence for the superior out-of-sample predictions of the CALE model over the conventional main-effects model for modeling health state values including EQ-5D-5L and 2 types of modified EQ-5D-5L states. All previous comparisons of the 2 models were secondary analyses of EQ-5D-5L valuation data collected using the same EQ-VT protocol. Therefore, it was possible that the results of those comparisons were specific to the study design. For example, the 86 health states included in the EQ-VT protocol may coincidentally favor the assumption of the CALE model. Nevertheless, this study is a dedicated investigation into the relative merits of the CALE model over the conventional model. We collected and modeled TTO data from a different set of EQ-5D-5L health states that is sufficient for estimating a value set. We also collected and modeled TTO data for 2 types of modified EQ-5D health states. Hence, our study suggests that the advantage of the CALE model in out-of-sample predictions is generalizable to other valuation study designs and even modified EQ-5D health states. This, together with findings from previous studies,<sup>6–8</sup> clearly indicated that the CALE model should be considered together with the conventional main-effects model in future EQ-5D-5L valuation studies. The conventional model is a useful tool for examining data quality and identifying possible interactions between dimensions and levels. The CALE model, if found to have better out-of-sample predictions, could be used to

produce the value set given that its predictions should be closer to the true values than the currently used model.

This study highlighted the importance of cross-validation for model evaluation in EQ-5D value set studies. In most EQ-5D value set studies, model selection is based on in-sample model fit, that is, agreement between observed and predicted values for the health states whose values are used to estimate the model. Such an approach is at the risk of selecting an overfitted model that may provide inferior predictions for health states that are not included in the value set studies. Predictions for health states not included in the valuation studies are important because only a very small portion of the EQ-5D health states are valued in value set studies. In the case of EQ-5D-5L, the recommended EQ-VT protocol included only 86 of 3125 health states. Cross-validation favors the model that provides the best predictions for out-of-sample health states. It reduces the risk of selecting an overfitted model. Indeed, the conventional model outperformed the CALE model in in-sample model fit (data not shown) but underperformed in out-of-sample model fit in our study, suggesting overfitting. Therefore, we advocate the use of cross-validation to guide the model selection process in future EQ-5D value set studies, irrespective of the use of the CALE model.

In addition to providing a new model specification for the standard EQ-5D-5L valuation studies, the use of the CALE model could significantly lower the costs of value set studies. Taking EQ-5D-5L as an example, the CALE model needs less than half of data for using the conventional 20-parameter model to estimate a value set with comparable prediction accuracy. This is important for countries where data collection is difficult or resources are limited. The current EQ-VT protocol requires face-to-face interviews of 1000 individuals who constitute a nationally representative general population sample. This sample size is based on the statistical power needed for running the 20-parameter model.<sup>9</sup> The data collection is typically completed by 5 to 10 interviewers who are intensively trained before and closely monitored by a specialized data QC team during the data collection period that may last 3 to 6 months.<sup>4</sup> Given that most countries worldwide have not established their EQ-5D value sets, a lighter,

**Table 2.** Model coefficients of conventional additive models.

Parameters	First arm		Second arm					
	EQ – SC, n = 2999		EQ-5D_1, n = 3200		EQ-5D_2, n = 3209		EQ + VI, n = 3298	
	Coef, SE	Coef, SE	Coef, SE	Coef, SE	Coef, SE	Coef, SE		
Intercept	0.036, 0.028	0.014, 0.024	0.054, 0.024	0.022, 0.023				
mo2	0.074, 0.020	0.036,* 0.019	0.015,* 0.019	0.049, 0.018				
mo3	0.131, 0.022	0.102, 0.020	0.094, 0.020	0.070, 0.021				
mo4	0.282, 0.021	0.238, 0.020	0.223, 0.019	0.199, 0.020				
mo5	0.356, 0.020	0.303, 0.019	0.289, 0.019	0.270, 0.019				
sc2		0.018,* 0.019	0.030,* 0.019	0.056, 0.019				
sc3		0.062, 0.020	0.052, 0.020	0.075, 0.020				
sc4		0.251, 0.021	0.193, 0.020	0.188, 0.021				
sc5		0.274, 0.020	0.256, 0.019	0.213, 0.020				
ua2	0.039, 0.020	0.052, 0.019	0.008,* 0.018	0.005,* 0.018				
ua3	0.099, 0.020	0.078, 0.020	0.050, 0.019	0.045, 0.019				
ua4	0.294, 0.020	0.251, 0.020	0.197, 0.019	0.173, 0.019				
ua5	0.383, 0.020	0.367, 0.019	0.318, 0.019	0.235, 0.019				
pd2	0.055, 0.022	0.046, 0.020	0.043, 0.020	0.063, 0.021				
pd3	0.098, 0.021	0.085, 0.020	0.101, 0.020	0.103, 0.021				
pd4	0.301, 0.020	0.286, 0.020	0.314, 0.020	0.280, 0.020				
pd5	0.410, 0.020	0.348, 0.020	0.391, 0.019	0.319, 0.020				
ad2	0.038,* 0.021	0.060, 0.019	0.031,* 0.019	0.057, 0.020				
ad3	0.117, 0.021	0.083, 0.020	0.140, 0.020	0.142, 0.020				
ad4	0.351, 0.020	0.308, 0.020	0.306, 0.019	0.310, 0.019				
ad5	0.472, 0.019	0.452, 0.019	0.423, 0.019	0.374, 0.019				
vi2				0.027,* 0.025				
vi3				0.017,* 0.022				
vi4				0.159, 0.023				
vi5				0.411, 0.019				
Log likelihood	-1428.63	-1455.57	-1414.90	-1450.90				

Each dimension has four dummy variables, e.g., mo3 is the dummy variable for having moderate problems in mobility dimension.

EQ – SC indicates EQ-5D-5L without the self-care dimension; EQ + VI, EQ-5D-5L plus a vision dimension; EQ-5D-5L, 5-level version of EQ-5D.

\*Not significant at 0.05 level.

lower-cost EQ-5D-5L valuation protocol would be very attractive. This cost-saving feature of the CALE model could also benefit the development of new preference-based instruments.

Our study demonstrated that it works well for expanded EQ-5D health states to cover vision function, suggesting possible application for estimating EQ-5D bolt-on value sets. If country-specific bolt-on value sets are considered appropriate, using the current EQ-VT protocol may not be economically viable or sustainable even for resource-rich countries, given the increasingly number of bolt-ons that may be potentially useful.<sup>18</sup> Investigators aiming to develop value sets for instruments dissimilar to the EQ-5D, such as EQ-Health and Wellbeing<sup>19</sup> and SF-6D, may also take advantage of the CALE model to lower their research costs. Theoretically, the more dimensions an instrument has and the more similar the level descriptors are, the more efficient the CALE model can be. Future research is warranted to assess this potential of the CALE model.

This study has several limitations. First, a student sample was used in this experimental study due to budget constraint. It has been shown that the preference of students is more homogeneous

than the general public.<sup>10</sup> In addition, it is clear that the student sample has a very different preference over EQ-5D health states compared with the general public; for example, mobility was valued the lowest and second lowest in the first and second arms respectively in this study, but was the most important dimension in Chinese value set.<sup>8</sup> In addition to the preference difference, the time spent on the interview (36.8 minutes) was shorter in this study than national valuation studies, for example, 58.3 minutes for the United States<sup>20</sup> and 42.6 minutes for the Italy,<sup>21</sup> which may be due to university students who tend to be more intellectual and engaged in the cTTO task and our interviewers who were experienced with conducting cTTO interviews. Second, the current study assumes that a main-effects model is sufficient for modeling health state values. We did not test interaction effects mainly because our study design (ie, selection of health states) was optimized for main-effects models. In the recent Slovenia 3-level version of EQ-5D valuation study, a variant of the CALE model with an exponential parameter was used to represent the marginal utility decrease theory.<sup>7</sup> In the marginal utility decrease theory, respondents may display diminishing sensitivity to health

**Table 3.** Results of the leave-one-state-out cross-validation analysis.

Model prediction measure	Arm and type of health state	CALE with a fix intercept	CALE with a random intercept	Additive with a fix intercept	Additive with a random intercept	Relative reduction,* %
MAE	EQ – SC	0.028 <sup>†</sup>	0.029	0.033	0.036	24
	EQ-5D_1	0.059	0.058 <sup>†</sup>	0.109	0.125	115
	EQ-5D_2	0.055	0.052 <sup>†</sup>	0.089	0.087	67
	EQ + VI	0.077	0.074 <sup>†</sup>	0.124	0.131	77
RMSE	EQ – SC	0.036 <sup>†</sup>	0.036 <sup>†</sup>	0.038	0.042	16
	EQ-5D_1	0.078	0.077 <sup>†</sup>	0.139	0.157	103
	EQ-5D_2	0.070	0.065 <sup>†</sup>	0.133	0.113	75
	EQ + VI	0.100	0.095 <sup>†</sup>	0.159	0.168	77
CCC	EQ – SC	0.996 <sup>†</sup>	0.996 <sup>†</sup>	0.996 <sup>†</sup>	0.995	0
	EQ-5D_1	0.986 <sup>†</sup>	0.986 <sup>†</sup>	0.956	0.945	4
	EQ-5D_2	0.988	0.990 <sup>†</sup>	0.958	0.969	2
	EQ + VI	0.978	0.980 <sup>†</sup>	0.949	0.943	4
Pearson R	EQ – SC	0.997 <sup>†</sup>	0.997 <sup>†</sup>	0.996	0.995	0
	EQ-5D_1	0.987 <sup>†</sup>	0.987 <sup>†</sup>	0.956	0.945	4
	EQ-5D_2	0.988	0.990 <sup>†</sup>	0.958	0.969	2
	EQ + VI	0.978	0.980 <sup>†</sup>	0.950	0.945	4
Log likelihood	EQ – SC	–1922.638 <sup>†</sup>	–1924.083	–1924.382	–1927.686	0
	EQ-5D_1	–1944.192	–1944.021 <sup>†</sup>	–2063.179	–2109.647	8
	EQ-5D_2	–1942.907	–1926.877 <sup>†</sup>	–2059.587	–2004.059	4
	EQ + VI	–2031.667	–2023.504 <sup>†</sup>	–2177.614	–2205.389	8

CALE indicates cross-attribute level effects; CCC, concordance correlation coefficient; EQ – SC, EQ-5D-5L without the self-care dimension; EQ + VI, EQ-5D-5L plus a vision dimension; EQ-5D-5L, 5-level version of EQ-5D; MAE, mean absolute errors; RMSE, root mean squared errors.

\*Relative reduction was based on the results of random intercept models.

<sup>†</sup>The best performance.

problems when combined, so that the perceived disutility of problems on 2 separate dimensions at the same time may be smaller than the sum of the disutility of each problem in isolation. Third, only vision was used to expand the EQ-5D health states in

this study and the vision dimension was described differently with other vision bolt-on studies.<sup>22,23</sup> We encourage future studies to use well-established bolt-on descriptors. Our finding about the CALE model may not be generalizable to valuation of other health

**Table 4.** Results of the leave-one-block-out cross-validation analysis.

Model prediction measure	Descriptive system	CALE with a fix intercept	CALE with a random intercept	Additive with a fix intercept	Additive with a random intercept	Relative reduction,* %
MAE	EQ – SC	0.029 <sup>†</sup>	0.029 <sup>†</sup>	0.048	0.053	87
	EQ-5D_1	0.090	0.086 <sup>†</sup>	0.124	0.118	38
	EQ-5D_2	0.068	0.065 <sup>†</sup>	0.098	0.094	43
	EQ + VI	0.094	0.091 <sup>†</sup>	0.116	NA <sup>‡</sup>	23
RMSE	EQ – SC	0.036 <sup>†</sup>	0.036 <sup>†</sup>	0.059	0.066	84
	EQ-5D_1	0.120	0.114 <sup>†</sup>	0.165	0.161	42
	EQ-5D_2	0.087	0.085 <sup>†</sup>	0.127	0.125	48
	EQ + VI	0.119	0.113 <sup>†</sup>	0.139	NA <sup>‡</sup>	17
CCC	EQ – SC	0.996 <sup>†</sup>	0.996 <sup>†</sup>	0.991	0.988	1
	EQ-5D_1	0.967	0.970 <sup>†</sup>	0.939	0.942	3
	EQ-5D_2	0.981	0.982 <sup>†</sup>	0.962	0.963	2
	EQ + VI	0.970	0.972 <sup>†</sup>	0.961	NA	1
Pearson R	EQ – SC	0.997 <sup>†</sup>	0.997 <sup>†</sup>	0.991	0.989	1
	EQ-5D_1	0.967	0.971 <sup>†</sup>	0.940	0.943	3
	EQ-5D_2	0.981	0.983 <sup>†</sup>	0.962	0.963	2
	EQ + VI	0.970	0.973 <sup>†</sup>	0.962	NA <sup>‡</sup>	1
Log likelihood	EQ – SC	–1908.041	–1444.340 <sup>†</sup>	–1926.013	–1482.074	3
	EQ-5D_1	–2005.303	–1607.555 <sup>†</sup>	–2116.835	–1757.763	9
	EQ-5D_2	–2038.608	–1510.797 <sup>†</sup>	–2107.717	–1639.605	9
	EQ + VI	–2079.145	–1603.095 <sup>†</sup>	–2129.141	NA <sup>‡</sup>	2

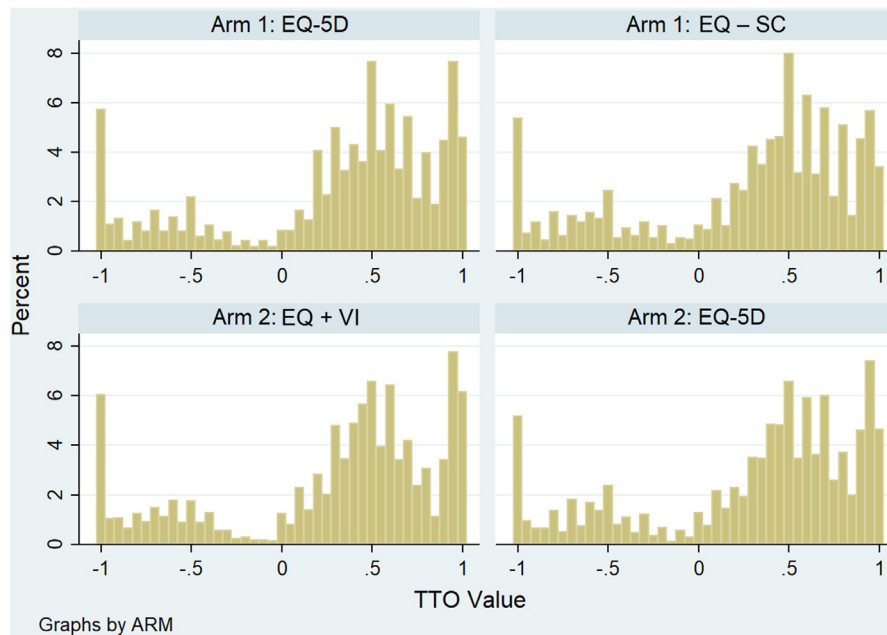
CALE indicates cross-attribute level effects; CCC, concordance correlation coefficient; EQ – SC, EQ-5D-5L without the self-care dimension; EQ + VI, EQ-5D-5L plus a vision dimension; EQ-5D-5L, 5-level version of EQ-5D; MAE, mean absolute errors; NA, not applicable; RMSE, root mean squared errors.

\*Relative reduction was based on the results of random intercept models for EQ – SC and EQ-5D models, but was based on fixed intercept models for EQ + VI.

<sup>†</sup>Indicates the best performance.

<sup>‡</sup>The model could not converge for the leave-one-out by block for the EQ + VI states.

**Figure 1.** cTTO value distributions by arms and types of health states.



cTTO indicates composite time trade-off.

dimensions such as those studied as bolt-ons to the EQ-5D.<sup>22,24,25</sup> To address these limitations, future studies should consider using a general public sample and a different bolt-on item. Finally, our study used a smaller design than the 86-state design of the EQ-VT studies. Given that the conventional 20-parameter model is proven to experience overfitting, a smaller design favors the CALE model. Hence, it would be useful to compare the performance 2 models using a larger design in the future.

## Conclusion

The CALE model is proved to be an attractive model specification for estimating EQ-5D-5L value sets. We strongly recommend investigators of future EQ-5D-5L value set studies to consider it together with the conventional main-effects model. Investigators with limited resources may consider using the CALE model to lower the costs for their valuation studies for EQ-5D-5L or similar health state descriptive systems.

## Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2022.12.012>.

## Article and Author Information

**Accepted for Publication:** December 12, 2022

**Published Online:** xxxx

doi: <https://doi.org/10.1016/j.jval.2022.12.012>

**Author Affiliations:** Health Services Management Department, Guizhou Medical University, Guiyang, China (Yang); Center of Medicine Economics and Management Research, Guizhou Medical University, Guiyang, China (Yang); Health Services Research Center, Akershus University Hospital,

Lørenskog, Norway (Rand); Math in Health B.V., Rotterdam, The Netherlands (Rand); Medical Psychiatry and Psychotherapy, Erasmus Medical Center, Rotterdam, The Netherlands (Busschbach); Saw Swee Hock School of Public Health, National University of Singapore, Singapore (Luo).

**Correspondence:** Nan Luo, PhD, Saw Swee Hock School of Public Health, National University of Singapore, 16 Medical Dr, Block MD3, 117597, Singapore. Email: [ephln@nus.edu.sg](mailto:ephln@nus.edu.sg)

**Author Contributions:** *Concept and design:* Yang, Rand, Busschbach, Luo  
*Acquisition of data:* Yang  
*Analysis and interpretation of data:* Yang, Rand, Busschbach, Luo  
*Drafting of the manuscript:* Yang, Rand, Busschbach, Luo  
*Critical revision of the paper for important intellectual content:* Yang, Rand, Luo, Busschbach  
*Statistical analysis:* Yang, Rand  
*Obtaining funding:* Yang  
*Administrative, technical, or logistic support:* Yang

**Conflict of Interest Disclosures:** Drs Yang and Luo reported receiving grants from the EuroQol Research Foundation during the conduct of the study and outside the submitted work. Dr Rand is the current Chair of the EuroQol Scientific Executive Committee. Dr Busschbach reports grants from the EuroQol Research Foundation outside the submitted work. Drs Yang, Rand, Busschbach, and Luo are EuroQol members. Dr Luo is an editor for *Value in Health* and had no role in the peer-review process of this article. No other disclosures were reported.

**Funding/Support:** This work was supported by the grant 20170640 from the EuroQol Research Foundation, the Netherlands.

**Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## REFERENCES

1. Yang Z, Luo N, Bonsel G, Busschbach J, Stolk E. Effect of health state sampling methods on model predictions of EQ-5D-5L values: small designs can suffice. *Value Health*. 2019;22(1):38–44.

2. Yang Z, Luo N, Bonsel G, Busschbach J, Stolk E. Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter. *Value Health*. 2018;21(4):456–461.
3. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
4. 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.
5. Wang P, Liu GG, Jo MW, et al. Valuation of EQ-5D-5L health states: a comparison of seven Asian populations. *Expert Rev Pharmacoecon Outcomes Res*. 2019;19(4):445–451.
6. Rand-Hendriksen K, Ramos-Goni 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):945–952.
7. Prevolnik Rupel V, Srakar A, Rand K. Valuation of EQ-5D-3L health states in Slovenia: VAS based and TTO based value sets. *Zdr Varst*. 2020;59(1):8–17.
8. 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):662–669.
9. Oppe M, Hout B. The 'power' of eliciting EQ-5D-5L values: the experimental design of the EQ-VT; 2017, EuroQol Working Paper Series 17003. <https://euroqol.org/publications/working-papers/>. Accessed December 1, 2022.
10. Yang Z, Luo N, Oppe M, Bonsel G, Busschbach J, Stolk E. Toward a smaller design for EQ-5D-5L valuation studies. *Value Health*. 2019;22(11):1295–1302.
11. Ramos-Goni JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20(3):466–473.
12. Purba FD, Hunfeld JA, Iskandarsyah A, et al. Employing quality control and feedback to the EQ-5D-5L valuation protocol to improve the quality of data collection. *Qual Life Res*. 2017;26(5):1197–1208.
13. 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):1037–1042.
14. Yang Z, Purba FD, Shafie AA, et al. Do health preferences differ among Asian populations? A comparison of EQ-5D-5L discrete choice experiments data from 11 Asian studies. *Qual Life Res*. 2022;31(7):2175–2187.
15. Sloane NJA. A library of orthogonal arrays. NeilSloane. <https://neilsloane.com/oadir/>. Accessed December 1, 2018.
16. Yang Z, Feng Z, Busschbach J, Stolk E, Luo N. How prevalent are implausible EQ-5D-5L health states and how do they affect valuation? A study combining quantitative and qualitative evidence. *Value Health*. 2019;22(7):829–836.
17. 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.
18. Finch AP, Brazier J, Mukuria C, Bjorner JB. An exploratory study on using principal-component analysis and confirmatory factor analysis to identify bolt-on dimensions: the EQ-5D case study. *Value Health*. 2017;20(10):1362–1375.
19. Brazier J, Peasgood T, Mukuria C, et al. The EQ health and wellbeing: overview of the development of a measure of health and wellbeing and key results. *Value Health*. 2022;25(4):482–491.
20. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22(8):931–941.
21. Finch AP, Meregaglia M, Ciani O, Roudijk B, Jommi C. 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.
22. Longworth L, Yang Y, Young T, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technol Assess*. 2014;18(9):1–224.
23. Gandhi M, Ang M, Teo K, et al. A vision 'bolt-on' increases the responsiveness of EQ-5D: preliminary evidence from a study of cataract surgery. *Eur J Health Econ*. 2020;21(4):501–511.
24. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value Health*. 2015;18(1):52–60.
25. 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.