

RESEARCH ARTICLE

Open Access



# Similar responses to EQ-5D-3L by two elicitation methods: visual analogue scale and time trade-off

Xiuying Wang<sup>1†</sup>, Lin Zhuo<sup>2†</sup>, Yifei Ma<sup>3</sup>, Ting Cai<sup>3</sup>, Aviva Must<sup>4</sup>, Ling Xu<sup>5</sup> and Lang Zhuo<sup>3\*</sup> 

## Abstract

**Background:** Health-related quality of life (HRQoL) is often measured using EQ-5D-3L by the elicitation methods of visual analogue scale (VAS) and time trade-off (TTO). Although many countries have constructed both national VAS and TTO value sets, the fact that VAS and TTO value sets produces different values bewilders researchers and policymakers. The aim of this study is to explore certain conditions which could yield similar value sets using VAS and TTO.

**Methods:** A homogeneous sample of medical school students was selected to value 18 hypothetical health states using VAS and TTO methods. The 18 hypothetical health states were produced by orthogonal design (L18, 2\*3<sup>7</sup>). The range of rescaled values was transformed into  $-1 \sim 0 \sim 1$ . The investigations via different methods were carried out by computer-assisted personal interviewing with a wash-time interval of 72 h. Value sets for VAS and TTO were constructed using general least square regression models. Independent variables were composed of 10 dummy variables from 5 dimensions and including or omitting both constant and N3 terms.

**Results:** Three hundred thirteen medical students participated. The mean age was  $21.03 \pm 0.44$  years and 56.2% were female. The four regression models (for each method with and without constant and N3 terms) were all statistically significant ( $P < 0.05$ ) with high goodness-of-fit (Adj.  $R^2 > 0.94$  and MAE  $< 0.033$ ). Differences between the coefficients of the 10 dummy variables corresponding to each model were all less than 0.059. Pearson correlation coefficients between observed means and predicted values exceeded 0.981. Fitted curves of VAS and TTO largely coincided.

**Conclusions:** VAS and TTO can generate similar responses under certain conditions, suggesting that the two valuation methods could be equivalent intrinsically. The VAS method appears a more valid approach for valuation in the general population due to its greater simplicity and feasibility.

**Keywords:** Quality of life measurement, EQ-5D-3L, Visual analogue scale (VAS), Time trade-off (TTO), Valuation method

## Background

Worldwide, people are getting healthier, living longer and spending more time with often debilitating chronic diseases [1]. Patients with chronic diseases are confronted with reduced quality of life (QoL) while the management of chronic disease significantly drains the

human and financial resources in the health system [2]. The past decades have seen a growing body of research into quality-adjusted life-years (QALYs) which have emerged as one of the key outcome measures in health resources allocation [3, 4]. QALYs are calculated using the time in a specific health state multiplied by a score representing the value of that specific health state [5]. Health-related quality of life (HRQoL) has been widely used to indicate the utility of any specific health state [6]; the score of utility is anchored at 0 (death) and full

\* Correspondence: [zhuolang@xzhmu.edu.cn](mailto:zhuolang@xzhmu.edu.cn)

<sup>†</sup>Xiuying Wang and Lin Zhuo contributed equally to this work.

<sup>3</sup>School of Public Health, Xuzhou Medical University, Xuzhou 221004, Jiangsu, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

health (1). To date, several multi-attribute utility-based instruments (MAUI) have been proposed for measuring HRQoL, e.g. EQ-5D [7], Short Form 36 (SF-36) [8], WHOQOL-BREF [9], and Health Utility Index (HUI) [10, 11]. Among these, EQ-5D is the most concise [12, 13].

The EQ-5D, an acronym for “European Quality of Life with 5 Dimensions”, is a generic instrument that was published by the EuroQol Group in 1990 [14], consisting of a EQ-5D descriptive system and a Visual Analogue Scale [3]. The EQ-5D descriptive system comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) with 3 levels each (no problem, moderate problem, and severe problem), thus defining 243 ( $3^5$ ) distinct health states to characterize HRQoL. EQ-5D has been used in health-value research [15], cost-utility analysis [16], and population health services surveys [17, 18] in many countries.

Visual analogue scale (VAS) or time trade-off (TTO) [19] are commonly used for EQ-5D value set elicitation. The first pair of VAS and TTO value sets for the EQ-5D was derived from the general population of the United Kingdom in 1990s [20]. Subsequently, Germany, Spain, Denmark, Argentina, and Sweden etc. have generated both VAS and TTO value sets [20, 21]. However, all value sets identified in pairs offered inconsistent values for HRQoL, that is, the VAS and TTO value sets in the same country generated different values for same health states [20, 21]. For example, the VAS value set predicts lower scores than TTO value set for mild health states and higher scores for severe health states in the United Kingdom [21]; in Sweden, the predicted TTO values are uniformly higher than VAS values [22]. Dominant explanations for the discrepancies include: that different instruments measured different aspects of health-related quality of life and thus yielded different results [23]; VAS values did not relate to years of life, VAS values were not useful for economic analysis [24]. However, some health economic critics contended that the discrepancies in the value sets preclude reaching a valid conclusion [25]. Furthermore, some policymakers argued that these inconsistencies undermine the fundamental strength and validity of HRQoL measurement [4, 26]. These criticisms demonstrate the need for better understandings of these discrepancies.

There have been several investigations in recent decades into the causes of the observed disagreement [27, 28]. In 2009, Craig et al. demonstrated the extent of agreement between VAS and ranking, another elicitation method of health states [29, 30]. To date, there has been relatively little research focusing on the agreement between VAS and TTO. This study attempts to redress this deficit. We assert that the difference between the two elicitation methods stem from several biases. First, the contrasting levels of complexity between the two methods may result in

differences in comprehension and adherence of the respondents to the protocol [31]. Second, the lack of a lower boundary in the transformed HRQoL values gives respondents too much “free space” in which to make a choice [32]. Furthermore, traditional interviews, with pencil and paper, cannot provide immediate feedback on inconsistencies, which inevitably results in errors and decreases validity. Finally, pre-selected health states in MVH protocol comprised an empirical sample, which does not represent the full underlying population of all health states. By eliminating these potential sources of biases, we hypothesized that we would achieve relatively similar responses using VAS and TTO methods. If achieved, it would serve to strengthen the foundation of HRQoL measurement.

## Methods

The current experimental study was designed to obtain value sets by VAS and TTO elicitation methods using an adapted measurement and valuation of health (MVH) protocol. We improved our study design in several aspects as outlined below.

### Homogeneous sampling

Mortimer et al. have argued that variation between individuals is more important in explaining variation in predicted quality-of-life weights than the choice of elicitation technique [33]. In comparison to VAS, TTO is cognitively burdensome and challenging to administer, as pointed out by Craig [24]. These observations suggest that sample homogeneity offers the opportunity to improve both validity and feasibility. Therefore, we chose a homogeneous sample of third-year undergraduates in Xuzhou Medical University.

### Sample size

According to Chevalier et al. [34], the sample size needed for a general value set was calculated following the formula:  $n = \frac{Z_{1-\frac{\alpha}{2}}^2 \sigma^2}{\delta^2}$ , where  $Z_{1-\frac{\alpha}{2}}$  represents the percentile of the normal distribution used as the critical value in a two-tailed test of size  $\alpha$  ( $Z_{1-\frac{\alpha}{2}} = 1.96$  for a 0.05 level test).  $\delta$  is the tolerated margin of error ( $\delta = 0.05$ ), and  $\sigma$  is an estimate of the standard deviation from a pilot survey ( $\sigma = 0.4$ ). Application of this formula results in 246 samples needed to obtain an estimation of the mean with a 95% ( $1-\alpha$ ) probability that the true mean falls in the interval of observed mean  $\pm \delta$ . Taking into account the expected compliance rate, we increased the sample size to 350, which is accordant to the one recommended by Lamers et al. [35].

### Selection of health states

Although 43 health states in MVH protocol [36] and 97 health states in Paris protocol [37] were recommended

to derive the EQ-5D value set, these states were chosen arbitrarily. A fully balanced set of health states is needed to represent all health states. Put another way, a representative sample of health states is critical to draw a valid inference. In this study, a total of 18 hypothetical health states was created by orthogonal design (L18,  $2^3 \times 3^7$ ), which is an approach increasingly adopted [38, 39]. Table 1 presents the 18 hypothetical health states used in this study, ordered as in the actual protocol.

### The valuation tasks

Because of its greater simplicity, VAS valuation was conducted prior to TTO valuation. According to Ebbinghaus' theory of forgetting curve [40], a 72-h between-task interval was employed to reduce the negative influence of the retention. Before each experimental trial, the participants attended a classroom instruction 1) explaining the purpose of the study; 2) introducing EQ-5D and VAS or TTO valuations; 3) demonstrating the interface of computer-assisted personal-interviewing (CAPI) software and its functions; and 4) reinforcing with opportunities for practice with the CAPI software. In the software, the state of full health (11111) was assigned a value of 10 as an anchor point. The participants were required to value the other 17 health states using VAS or TTO methods.

In the process of VAS valuation, a tailored scale, similar to a thermometer, was shown to the participants on the CAPI interface. “-10 (worst imaginable state)” was labeled at the lower end of the scale; “0 (dead)” was

labeled at the midpoint of the scale; “10 (full health)” was labeled at the top of the scale. Participants were then asked to rank one health state each time on the scale at the point  $x$ , to indicate how good or bad they deemed the state. A utility weight for each state was calculated as  $x/10$ . Accordingly, the range of the transformed values was  $-1 \sim 0 \sim 1$ . In this scheme,  $-1$ ,  $0$ , and  $1$  indicate the imaginable worst state, being dead, and full health, respectively [32, 41].

In the process of TTO valuation, the participants were assumed to be in the impaired health state for 10 years followed by immediate death. If the health state was considered better than death (BTD), then the participants were asked to trade for  $t$  years in full health, where  $t$  decreased from 10 to 0 with a decrement of 1 year followed by immediate death. The elicitation process ended when the participant was indifferent to either in full health for  $t$  years or in the impaired state for 10 years. One decimal was permitted if the participant believed that 1 year was not adequately precise. Utility weight for BTD was calculated as  $t/10$ . In the case of a health state being regarded as worse than death (WTD), the participants would prefer to live  $t$  years in full health to compensate for tolerating the impaired state for  $(10-t)$  years. The years for compensation decreased from 10 to 0 with a decrement of 1 year followed by immediate death. The elicitation process ended when the participant was indifferent to either  $(10-t)$  years of tolerating the health state plus  $t$  years of full health or immediate

**Table 1** 18 health states created by orthogonal design (L18,  $2^3 \times 3^7$ )

State number	Mobility	Self-Care	Usual Activities	Pain/Discomfort	Anxiety/Depression
1	1	1	1	1	1
2	3	1	1	2	2
3	2	2	1	1	3
4	1	3	2	1	2
5	1	2	3	2	1
6	2	1	2	3	1
7	2	2	2	2	2
8	2	3	3	1	1
9	1	1	3	3	2
10	3	2	1	3	1
11	3	1	2	1	3
12	1	3	1	2	3
13	1	2	2	3	3
14	3	3	2	2	1
15	2	1	3	2	3
16	2	3	1	3	2
17	3	2	3	1	2
18	3	3	3	3	3

The numbers 1, 2, and 3 in five dimensions represent level 1 (no problem), level 2 (moderate problem) and level 3 (severe problem), respectively

death. A decimal was permitted if the participant believed 1 year was not adequately precise. Utility weights for WTD states were calculated as  $-t/10$ . The range of transformed TTO values was also  $-1 \sim 0 \sim 1$ . Thus, the two evaluation methods are on the same scale, similar to the EuroQol Group Valuation Technology (EQ-VT) protocol described by Oppe M. et al. [42].

### Quality control

The valuation processes were carried out in a computer lab. After instruction, the participants carried out the task individually. There were three types of approaches to optimize the quality of valuation. These are: an acceptable predicted value for the distinct participant; the number of inconsistencies is three or less; the absolute value of the difference between the standard deviation of the 18 health states and 2.5 is less than 0.5. The three types of approaches are explained in detail as below.

After valuing 18 hypothetical states, participants were asked to describe their own health state using the EQ-5D-3L descriptive system in the CAPI software. Then a predicted value of the participant was derived from a multiple linear regression model, which was generated from the previously valued 18 health states. Additionally, the number of logical inconsistencies was also fed back to the participant based on the multiple linear regression models. For example, when disutility is adopted as independent, the level 3 (severe problems) of each dimension should have a higher value (in absolute term) than the level 2 (moderate problems), and the level 2 should be higher than the level 1. Accordingly, there are 15 comparable pairs in total. The standard deviation of the 18 values of the hypothetical states was also presented. Participants were asked to optimize their previous values to improve the results, but they retained the ability to keep the initial results if they wanted.

Because individuals differ in their ability to grasp the abstract health states, we used three distinct approaches to accommodate individual differences: including numbers, words, and pictures. For example, the abbreviation 31122 represents a health state with extreme problems in walking, no problems with self-care, no problems with performing usual activities, moderate pain, and moderate anxiety. The numbers and words were both presented. Additionally, a cobweb diagram [43] illustrating the hypothetical state on the five dimensions and the three severity levels was also shown to the respondents on the interface of the CAPI. (The interfaces of the CAPI software are presented in the “[Supplementary Materials](#)”.)

### Data cleaning

Standard protocols for data cleaning do not yet exist. We excluded the participants who met at least one of the following criteria: 1) could not complete the value

task; 2) had more than four logical inconsistencies; 3) produced more than two outliers. Outliers were defined as those meeting all the following 1) data-points distinctly isolated from the whiskers in a box-plot; 2) distance between outliers and the nearest quartile was more than 3 times the interquartile range; and 3) values less than the 5th percentile or more than 95th percentile [43].

### Modeling

All 18 health states, including the anchor point 11111, were all used for building models for maximum use of the data. Studies have shown that the anchor point exerts a great impact on the value set [44, 45].

We employed disutility (1-utility) as the dependent variable. Independent variables included 10 dummy variables and N3. The dummy variables consisted of MO2, MO3, SC2, SC3, UA2, UA3, PD2, PD3, AD2, and AD3, which represent the main effect of any movement from no problem to moderate or severe problem for each dimension. N3 equals 1 if any dimension is level 3, 0 otherwise [43]. Table 2 shows the dummy variables used in the modeling.

We adopted a general least squares (GLS) regression method to produce multilevel models, since each participant valued 17 health states. Specifications of the models defined in this study are also presented in Table 2. All statistical analysis and modeling were conducted using STATA/SE 12.0 (StataCorp, College Station, TX) with  $\alpha$  set at 0.05 to declare statistical significance.

## Results

### Sample characteristics

We enrolled 350 medical students in this survey and excluded 37 participants including 2 for incompleteness, 26 who had more than 4 inconsistencies, and 9 who had more than 2 outliers. The final sample consisted of 313 participants with an average age of  $21.03 \pm 0.44$  years; 56.2% were female.

### Descriptive statistics of 18 health states

Overall, 5634 VAS values (96.1% BTD) were collected with an average of  $4.788 \pm 2.703$ ; the state of 33,333 received 197 out of 218 WTD values. All 5634 TTO values (94.2% BTD) were collected with an average of  $4.310 \pm 2.457$ ; the state of 33,333 received 260 out of 326 WTD values. Other health states had only a small number of WTD values (VAS < 3, TTO < 6).

Raw data were then transformed into utility values by dividing by 10. Table 3 shows the mean, standard deviation, median, and quartiles of transformed values of VAS and TTO for each health state. Differences between the means for TTO and VAS are generally less than 0.1 (Table 3).

**Table 2** Definition of dummy variables and model specification

Dummy variables	Definition	
MO2	1 if mobility is level 2; 0 otherwise	
SC2	1 if self-care is level 2; 0 otherwise	
UA2	1 if usual activities is level 2; 0 otherwise	
PD2	1 if pain/discomfort is level 2; 0 otherwise	
AD2	1 if anxiety/depression is level 2; 0 otherwise	
MO3	1 if mobility is level 3; 0 otherwise	
SC3	1 if self-care is level 3; 0 otherwise	
UA3	1 if usual activities is level 3; 0 otherwise	
PD3	1 if pain/discomfort is level 3; 0 otherwise	
AD3	1 if anxiety/depression is level 3; 0 otherwise	
N3	1 if at least one level 3; 0 otherwise	
constant	loss of utility of any health state away from full health	
<b>Model Specification</b>	<b>f (x)</b>	<b>Methods</b>
Model 1	$f(\text{MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3 N3})$	VAS model with constant and N3
Model 2	$f(\text{MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3 N3})$	TTO model with constant and N3
Model 3	$f(\text{MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3})$	VAS model without constant and N3
Model 4	$f(\text{MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3})$	TTO model without constant and N3

**Table 3** Mean, standard deviation, median and quartiles for rescaled VAS and TTO values ( $n = 313$ )

States	VAS					TTO				
	Mean	Std. Deviation	P25	P50	P75	Mean	Std. Deviation	P25	P50	P75
11111	1.0000	0.0000	1	1	1	1.0000	0.0000	1	1	1
31122	0.6874	0.1628	0.6	0.7	0.8	0.6125	0.1685	0.5	0.6	0.7
22113	0.5679	0.1858	0.4	0.6	0.7	0.5193	0.1613	0.4	0.5	0.6
13212	0.6178	0.1797	0.5	0.6	0.7	0.4867	0.1632	0.4	0.5	0.6
12321	0.5882	0.1577	0.5	0.6	0.7	0.5025	0.1465	0.4	0.5	0.6
21231	0.5171	0.1377	0.4	0.5	0.6	0.4816	0.1383	0.4	0.5	0.5
22222	0.4613	0.1468	0.4	0.5	0.5	0.4155	0.1556	0.3	0.4	0.5
23311	0.4451	0.1387	0.4	0.4	0.5	0.4300	0.1483	0.4	0.4	0.5
11332	0.4300	0.1454	0.3	0.4	0.5	0.4503	0.1446	0.4	0.5	0.5
32131	0.3962	0.1422	0.3	0.4	0.5	0.3963	0.1486	0.3	0.4	0.5
31213	0.4962	0.1857	0.4	0.5	0.6	0.3980	0.1450	0.3	0.4	0.5
13123	0.4553	0.1307	0.4	0.5	0.6	0.4426	0.1431	0.4	0.5	0.5
12233	0.4335	0.1921	0.3	0.4	0.6	0.3842	0.1941	0.3	0.4	0.5
33221	0.4764	0.2063	0.3	0.4	0.6	0.3835	0.1412	0.3	0.4	0.4
21323	0.4006	0.1817	0.3	0.4	0.6	0.3647	0.1381	0.3	0.4	0.4
23132	0.3946	0.2201	0.2	0.4	0.5	0.2794	0.1550	0.2	0.2	0.3
32312	0.3884	0.2301	0.2	0.4	0.5	0.2904	0.1428	0.2	0.3	0.3
33333	-0.1378	0.2465	-0.2	-0.1	0	-0.0801	0.1509	-0.125	-0.1	-0.028
Total	0.4788	0.2703	0.3	0.5	0.6	0.4310	0.2457	0.3	0.4	0.5

### Regression analyses

Four models and the goodness-of-fit indices for each are shown in Table 4. All models and all coefficients were statistically significant ( $P < 0.05$ ). All models passed the Breusch-Pagan/Cook-Weisberg test, which indicated the presence of homoscedasticity. Four regression coefficients for each dummy variable are very close to one another (Table 4). The greatest difference is only 0.059 produced by PD3 between Model 1 and Model 2; PD2 produces the least difference of 0.01 between Model 1 and Model 2.

High levels of the goodness-of-fit statistics are seen for all models. Adjusted-R squares all exceed 0.94. Pearson correlation coefficients ( $r$ ) between observed means and predicted values are higher than 0.98. Mean absolute error (MAE) between observed means and predicted values is less than 0.04. No errors in logic were observed among the predicted values of 243 health states.

Figure 1a delineates the predictions derived from Model 1 (VAS) and Model 2 (TTO). Figure 1b delineates the predictions derived from Model 3 (VAS) and Model 4 (TTO). The points of each health state in Fig. 1a and b are almost overlapping except the states of 21232 and

32211, which differ slightly. This indicates that similar results obtained under certain conditions supports different model specifications.

### Discussion

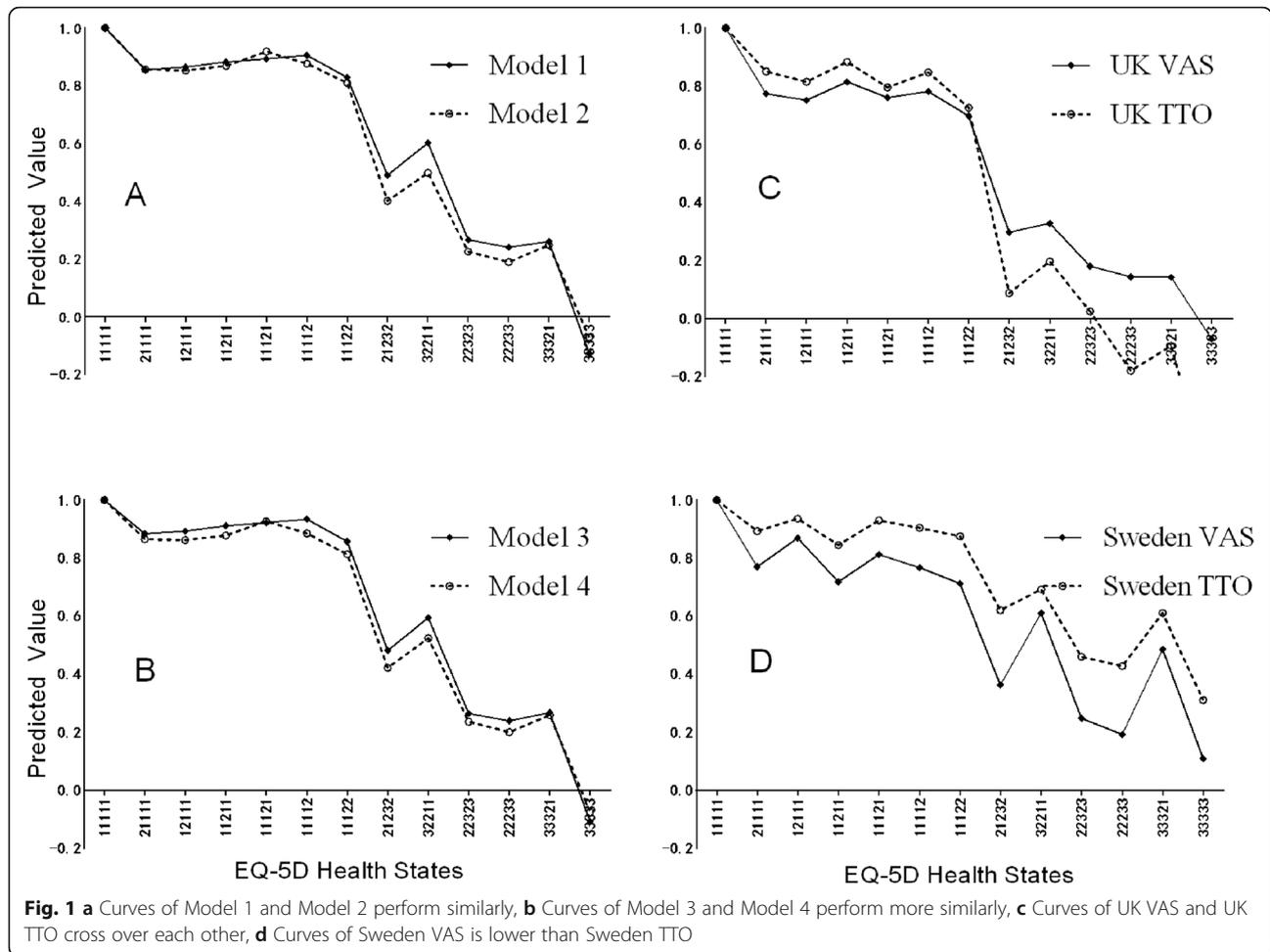
Previous studies comparing VAS and TTO have documented many defects regarding TTO, including that TTO generates more inconsistencies than VAS [46]; TTO is burdensome [31]; TTO values are negatively related to the duration of optimal health states and positively related to the poorer health states [24]; TTO is prone to interviewer effects [47]. After controlling some conditions through the design of the present study, we found that VAS and TTO generate similar responses. Comparing the predicted values in our study (Fig. 1a and b) to prior reports in the literature (Fig. 1c and d) [21, 22], results derived from the present are more similar. Four specific features of our study design provide a basic framework for explaining the discrepancies between VAS and TTO.

The key explanation for the ability to generate similar responses was the homogeneous sample chosen from a medical university. Medical students are superior to the

**Table 4** Coefficients and indices of the goodness-of-fit of VAS and TTO models

	Model 1 (VAS)	Model 2 (TTO)	Model 3 (VAS)	Model 4 (TTO)
11111	1	1	1	1
constant	-0.0299	-0.0141	-	-
MO2	-0.1154	-0.1292	-0.1158	-0.1340
MO3	-0.2180	-0.2058	-0.2111	-0.2159
SC2	-0.1060	-0.1329	-0.1064	-0.1377
SC3	-0.2172	-0.2220	-0.2103	-0.2321
UA2	-0.0879	-0.1170	-0.0884	-0.1218
UA3	-0.2413	-0.2103	-0.2345	-0.2205
PD2	-0.0767	-0.0673	-0.0772	-0.0721
PD3	-0.2553	-0.1967	-0.2484	-0.2068
AD2	-0.0646	-0.1096	-0.0650	-0.1145
AD3	-0.2087	-0.1893	-0.2018	-0.1994
N3	0.0439	-0.0318	-	-
Adj. R-square	0.9409	0.9499	0.9550	0.9664
AIC	-5285.48	-6918.29	-5252.47	-6876.14
BIC	-5192.57	-6825.38	-5172.83	-6796.50
MAE	0.0304	0.0269	0.0327	0.0310
# > 0.05	4	1	4	4
# > 0.1	0	0	0	0
$r$	0.9847	0.9879	0.9813	0.9867
Logical error num.	0	0	0	0

All models and regression coefficients were significant ( $P < 0.05$ ); Adj. R-square adjusted R-square, AIC Akaike information criterion, BIC Bayesian information criterion, MAE mean absolute error between observed mean and predicted value # > 0.05, number of MAE > 0.05 out of 18 states; # > 0.1, number of MAE > 0.1 out of 18 states,  $r$ , correlation coefficient between observed means and predicted values, Logical error num., number of inconsistencies among all predicted health state values



general population as study subjects in two respects. First, they have less difficulty in understanding the TTO method than the general population. It was widely reported that TTO is more difficult to understand than VAS for the general population [31, 36, 48]. Medical students are young, enthusiastic, and committed to health and healthcare. Therefore, they are able to reduce error in the face of the complex TTO valuation task. Second, medical students are highly adherent. The TTO method involved a term of “immediate death”, which can elicit antipathy and non-adherence in the general population [49, 50]. Medical students are more comfortable with these concepts, more devoted to improving the quality of life of their patients, and thereby more likely to complete the trade-off task.

Adopting an adjusted scale is the second important contributor to the similar responses we demonstrated. The scale used in our study is not the same as the 20 cm vertical scale calibrated from 0 to 100 that is standard in EQ-VAS. We used a scale of  $-10 \sim 0 \sim 10$  for both methods for several reasons. First, we sought to apply the same scale in the elicitation, to minimize systematic

measurement error. Second, by using revised scale, VAS possesses explicit anchor points as TTO does, and an anchored scale has the advantage of simplifying the process of the VAS valuation. Third, the range of  $-10 \sim 0 \sim 10$  has fixed boundaries for the transformed values of HRQoL. Fixed boundaries are critical in the process of elicitation and modeling [45]. However, in the MVH protocol, both ranges of VAS and TTO methods are transformed into  $-\infty \sim 0 \sim 1$  [36], so the lower boundary is unbounded. As Seymour et al. demonstrated, the “ceiling” effect can be controlled by adjusting VAS and TTO ranges between  $-1$  and  $1$  [51]. EQ-VT protocol, the latest version of TTO elicitation, has also adopted  $-1 \sim 0 \sim 1$  [37]. Dolan has also introduced a fix-boundary rescaling method for WTD states. Although there was a typo in the alternative formula, the idea of “compensation” reflects the essence of trade-offs [36].

The third point worthy of mention is the computer-assisted personal-interviewing process. As reported by Ramos-Goni et al., interviewer effects were identified in many valuation studies [52]. In the CAPI process, a standardized procedure of valuation was embedded to

facilitate self-administration, and should eliminate the potential interviewer bias [34, 35, 53, 54]. Furthermore, the CAPI process has been successful in other research ways: assisting and simplifying. Assisting means that inconsistent and irrational values would be fed back to the participants in real time to improve the reliability of the valuation. In addition, the use of computer-assisted personal-interviewing process reduces the need for the “warm-up” steps of ranking and pairwise comparison, which should be advantageous given the finding that ranking leads to a higher rate of inconsistency than VAS [55].

Selecting fully balanced health states represents the final key factor contributing to similar responses. According to the definition of the QALYs, the utility values of health states must lie on an interval scale anchored at 0 (death) and 1 (full health). Therefore, fully balanced states are expected to receive well-distributed utility values. A total of 18 health states selected via an orthogonal design have the considerable advantage of balanced distribution, which should represent all possible levels for each dimension. There is no contradiction between the balanced distribution of health states and a stabilized standard deviation of utility values. The stabilized standard deviation also plays an important role in the modeling algorithms as well as facilitating the comparability of the resultant value sets. Additionally, Sun et al. suggested that reasonable parity of health states should produce better results [56].

Although most predicted values were close to each other based on methods of VAS and TTO, the two states of 21,232 and 32,211 were slightly separated. This suggests that there may be other factors which were overlooked. For example, Augestad et al. pointed out that the attitudes toward death may influence the value sets [44]. The use of “death” is inevitable in the TTO method. This essential difference between the VAS and TTO is difficult to eliminate. Additionally, the process of comparing the current health state to “death” might cause “noise” since it is metaphysically unknown [44]. Badia et al. found that the VAS is more feasible and reliable than TTO in the Spanish population [57]. Taking our results into consideration, one could conclude that VAS should occupy a position of relative advantage over TTO in the general population, especially to older adults living in rural areas.

In summary, this study surfaced some previously neglected biases and provided experimental evidence that VAS and TTO can generate similar results under specific conditions. The similarity might shed light on the intrinsic equality of both methods. VAS would therefore seem to serve as a substitute for TTO, especially in a general population survey due to its relative ease and convenience. The major strength of this study was its experimental study design. The main weakness is the

presence of separation over a small number of health states, suggesting unmeasured characteristics. Despite these encouraging results, many unanswered questions remain, such as the extent to which the new scale influenced the VAS valuation, and identification of the effect of each restrictive condition. Future work would tackle these issues.

## Conclusions

The data reported here have further strengthened our speculation that VAS and TTO methods in valuation of EQ-5D health states could be intrinsically equivalent. If confirmed, the VAS method is more applicable than TTO for health valuation in the general population due to its simplicity and superior feasibility.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12874-020-01008-9>.

**Additional file 1.** The interfaces of VAS and TTO in the CAPI software.

## Abbreviations

BTD: Better than death; CAPI: Computer-Assisted Personal Interviewing; EQ-5D: EuroQol 5 dimensions; EQ-VT: EuroQol Group Valuation Technology; HRQoL: Health-related Quality of Life; HUI: The Health Utility Index; MAUI: Multi-attribute utility-based instruments; MAE: The mean absolute error; GLS: General least squares; MVH: Measurement and Valuation of Health; SF-36: The Short Form of 36 questions; SG: The Standard gamble; TTO: The time trade-off; VAS: The visual analogue scale; WTD: Worse than death

## Acknowledgments

We thank the National Nature Science Foundations of China and Jiangsu Education Department for funding this study. A special word of thanks is given to Professor Paul Kind, University of Leeds, United Kingdom, for his suggestions for improving the data reporting and analysis of the manuscript.

## Authors' contributions

XW and LZ1 wrote the paper. YM and TC contributed to the work of data collection and analysis. LX and AM participated in interpreting the findings. LZ2 participated in designing the study. All authors read and approved the final manuscript.

## Funding

The project was funded by Overseas Research Program for Excellent Teachers and Principals of Colleges and Universities in Jiangsu Province (Grant number 20180120); National Natural Science Foundation of China (Grant number 71373183) and Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

## Availability of data and materials

Raw data may be made available upon reasonable request from the corresponding author.

## Ethics approval and consent to participate

This study was approved by the ethics committee of Xuzhou Medical University (EA#20161014002) and written informed consent was obtained from all individual participants in this study.

## Consent for publication

Not applicable.

## Competing interests

No authors had financial conflicts of interest to disclose regarding the contents of this manuscript.

**Author details**

<sup>1</sup>Department of Nephrology, Xuzhou Central Hospital, Xuzhou Clinical School of Xuzhou Medical University, Xuzhou, Jiangsu, China. <sup>2</sup>School of Basic Medical Sciences, Xinxiang Medical University, Xinxiang, Henan, China. <sup>3</sup>School of Public Health, Xuzhou Medical University, Xuzhou 221004, Jiangsu, China. <sup>4</sup>Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, USA. <sup>5</sup>Center for Health Statistics and Information, National Health Commission, Beijing, China.

Received: 13 September 2019 Accepted: 5 May 2020

Published online: 14 May 2020

**References**

- Global Burden of Disease Cancer C, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. *JAMA Oncol.* 2017;3(4):524–48.
- Weernink MGM, Groothuis-Oudshoorn CGM, IJzerman MJ, van Til JA. Valuing treatments for Parkinson disease incorporating process utility: performance of best-worst scaling, time trade-off, and visual analogue scales. *Value Health.* 2016;19(2):226–32.
- Brooks R. EuroQol: the current state of play. *Health Policy.* 1996;37(1):53–72.
- McDonough CM, Tosteson AN. Measuring preferences for cost-utility analysis: how choice of method may influence decision-making. *Pharmacoeconomics.* 2007;25(2):93–106.
- Heintz E, Wirehn AB, Peebo BB, Rosenqvist U, Levin LA. QALY weights for diabetic retinopathy—a comparison of health state valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. *Value Health.* 2012;15(3):475–84.
- Haring R, Feng YS, Moock J, Volzke H, Dorr M, Nauck M, Wallaschofski H, Kohlmann T. Self-perceived quality of life predicts mortality risk better than a multi-biomarker panel, but the combination of both does best. *BMC Med Res Methodol.* 2011;11:103.
- Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Busschbach J, Buxton M, Dolan P, Kind P, Krabbe P, et al. A single European currency for EQ-5D health states. Results from a six-country study. *Eur J Health Econ.* 2003;4(3):222–31.
- Crockett AJ, Cranston JM, Moss JR, Alpers JH. The MOS SF-36 health survey questionnaire in severe chronic airflow limitation: comparison with the Nottingham health profile. *Qual Life Res.* 1996;5(3):330–8.
- Xia P, Li N, Hau KT, Liu C, Lu Y. Quality of life of Chinese urban community residents: a psychometric study of the mainland Chinese version of the WHOQOL-BREF. *BMC Med Res Methodol.* 2012;12:37.
- Mo F, Choi BC, Li FC, Merrick J. Using health utility index (HUI) for measuring the impact on health-related quality of life (HRQL) among individuals with chronic diseases. *ScientificWorldJournal.* 2004;4:746–57.
- Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Oper Res.* 1982;30(6):1043–69.
- Hawthorne G, Richardson J, Day NA. A comparison of the assessment of quality of life (AQoL) with four other generic utility instruments. *Ann Med.* 2001;33(5):358–70.
- Brauer CA, Rosen AB, Greenberg D, Neumann PJ. Trends in the measurement of health utilities in published cost-utility analyses. *Value Health.* 2006;9(4):213–8.
- Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. *Ann Med.* 2001;33(5):337–43.
- Lee YJ, Lee S, Song S, Lee S, Suh D. Health utility in Myelofibrosis patients and comparison between health utility values measured by time trade off and visual analogue scale. *Value Health.* 2017;20(5):A115–6.
- O'Boyle CJ, Royston CM, Sedman PC. Cost-utility analysis of open versus laparoscopic groin hernia repair: results from a multicentre randomized clinical trial (Br J Surg 2001; 88: 653–61). *Br J Surg.* 2001;88(11):1543–5.
- Sun S, Chen J, Johannesson M, Kind P, Xu L, Zhang Y, Burstrom K. Population health status in China: EQ-5D results, by age, sex and socio-economic status, from the National Health Services Survey. *Qual Life Res.* 2008;20(3):309–20.
- Tan Z, Liang Y, Liu S, Cao W, Tu H, Guo L, Xu Y. Health-related quality of life as measured with EQ-5D among populations with and without specific chronic conditions: a population-based survey in Shaanxi Province China. *PLoS one.* 2013;8(7):e65958.
- Bala MV, Wood LL, Zarkin GA, Norton EC, Gafni A, O'Brien BJ. Are health states "timeless"? The case of the standard gamble method. *J Clin Epidemiol.* 1999;52(11):1047–53.
- Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Med Decis Making.* 2014;34(1):8–20.
- Szende AG, Oppe M, Devlin NJ. EQ-5D value sets: inventory, comparative review and user guide; 2007. p. 1–102.
- Burstrom K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin LA, Zethraeus N. Swedish experience-based value sets for EQ-5D health states. *Qual Life Res.* 2014;23(2):431–42.
- Stavem K. Quality of life in epilepsy: comparison of four preference measures. *Epilepsy Res.* 1998;29(3):201–9.
- Craig BM. The duration effect: a link between TTO and VAS values. *Health Econ.* 2009;18(2):217–25.
- Cooper A, Wallman JK, Gulfe A. What PASSes for good? Experience-based Swedish and hypothetical British EuroQol 5-dimensions preference sets yield markedly different point estimates and patient acceptable symptom state cut-off values in chronic arthritis patients on TNF blockade. *Scand J Rheumatol.* 2016;45(6):470–3.
- Zrubka Z, Beretzky Z, Hermann Z, Brodsky V, Gulacsi L, Rencz F, Baji P, Golicki D, Prevolnik-Rupel V, Pentek M. A comparison of European, polish, Slovenian and British EQ-5D-3L value sets using a Hungarian sample of 18 chronic diseases. *Eur J Health Econ.* 2019;20(Suppl 1):119–32.
- Mendeloff J. Measuring elusive benefits: on the value of health. *J Health Polit Policy Law.* 1983;8(3):554–80.
- De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ.* 2000;9(2):109–26.
- Craig BM, Busschbach JJ, Salomon JA. Keep it simple: ranking health states yields values similar to cardinal measurement approaches. *J Clin Epidemiol.* 2009;62(3):296–305.
- Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. *Med Care.* 2009;47(6):634–41.
- Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. *Pharmacoeconomics.* 2000;17(2):151–65.
- Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Med Care.* 2007;45(3):238–44.
- Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. *Med Decis Making.* 2008;28(1):66–89.
- Chevalier J, de Pourouville G. Valuing EQ-5D using time trade-off in France. *Eur J Health Econ.* 2013;14(1):57–66.
- Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ.* 2006;15(10):1121–32.
- Dolan P. Modeling valuations for EuroQol health states. *Med Care.* 1997;35(11):1095–108.
- Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goni JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics.* 2016;34(10):993–1004.
- Bailey H, Kind P, La Foucade A. Results from an exploratory study to test the performance of EQ-5D-3L valuation subsets based on orthogonal designs, and an investigation into some modeling and transformation alternatives for the utility function. *Health Econ Rev.* 2014;4(1):29.
- 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.
- Murre JM, Dros J. Replication and analysis of Ebbinghaus' forgetting curve. *PLoS One.* 2015;10(7):e0120644.
- Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value Health.* 2014;17(5):597–604.
- Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–53.

43. Zhuo L, Xu L, Ye J, Sun S, Zhang Y, Burstrom K, Chen J. Time trade-off value set for EQ-5D-3L based on a nationally representative Chinese population survey. *Value Health*. 2018;21(11):1330–7.
44. Augestad LA, Rand-Hendriksen K, Stavem K, Kristiansen IS. Time trade-off and attitudes toward euthanasia: implications of using 'death' as an anchor in health state valuation. *Qual Life Res*. 2013;22(4):705–14.
45. Lee BS, Kymes SM, Nease RF Jr, Sumner W, Siegfried CJ, Gordon MO. The impact of anchor point on utilities for 5 common ophthalmic diseases. *Ophthalmology*. 2008;115(5):898–903 e894.
46. Lee JY, Lee E. Assessment of utility for heart failure using visual analogue scale (vas), time-trade off (Tto) and Euroqol-5 dimension (Eq-5d) in the Korean general population. *Value Health*. 2016;19(7):A868–9.
47. Stiggelbout AM, Eijkemans MJC, Kiebert GM, Kievit J, Leer JWH, DeHaes HJCM. The 'utility' of the visual analog scale in medical decision making and technology assessment - is it an alternative to the time trade-off? *Int J Technol Assess Health Care*. 1996;12(2):291–8.
48. Emilsson L, Lindahl B, Koster M, Lambe M, Ludvigsson JF. Review of 103 Swedish healthcare quality registries. *J Intern Med*. 2014.
49. Jin X, Liu GG, Luo N, Li H, Guan H, Xie F. Is bad living better than good death? Impact of demographic and cultural factors on health state preference. *Qual Life Res*. 2016;25(4):979–86.
50. Yin Z, Li J, Ma K, Ning X, Chen H, Fu H, Zhang H, Wang C, Bruera E, Hui D. Development of palliative Care in China: a tale of three cities. *Oncologist*. 2017;22(11):1362–7.
51. Seymour J, McNamee P, Scott A, Tinelli M. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. *Health Econ*. 2010;19(6):683–96.
52. Ramos-Goni JM, Pinto-Prades JL, Oppe M, Cabases JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2014;55:e51–8.
53. Scalone L, Cortesi PA, Ciampichini R, Belisari A, D'Angiolella LS, Cesana G, Mantovani LG. Italian population-based values of EQ-5D health states. *Value Health*. 2013;16(5):814–22.
54. Wittrup-Jensen KU, Lauridsen J, Gudex C, Pedersen KM. Generation of a Danish TTO value set for EQ-5D health states. *Scand J Public Health*. 2009;37(5):459–66.
55. Andrade MV, Noronha K, Kind P, Reis Cde B, de Carvalho LR. Logical inconsistencies in 3 preference elicitation methods for EQ-5D health states: a study in the Brazilian population. *Med Decis Making*. 2016;36(2):242–52.
56. Sun S, Chen J, Kind P, Xu L, Zhang Y, Burstrom K. Experience-based VAS values for EQ-5D-3L health states in a national general population health survey in China. *Qual Life Res*. 2015;24(3):693–703.
57. Badia X, Monserrat S, Roset M, Herdman M. Feasibility, validity and test-retest reliability of scaling methods for health states: the visual analogue scale and the time trade-off. *Qual Life Res*. 1999;8(4):303–10.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

