Applied Health Economics and Health Policy (2023) 21:405–418
https://doi.org/10.1007/s40258-023-00794-9

SYSTEMATIC REVIEW

Discrete Choice Experiments in Health State Valuation: A Systematic Review of Progress and New Trends

Haode Wang[1] · **Donna L. Rowen**[1] · **John E. Brazier**[1] · **Litian Jiang**[2]

Accepted: 12 February 2023 / Published online: 30 March 2023

Abstract

Background Discrete choice experiments (DCEs) are increasingly used in health state valuation studies.

Objective This systematic review updates the progress and new findings of DCE studies in the health state valuation, covering the period since the review of June 2018 to November 2022. The review reports the methods that are currently being used in DCE studies to value health and study design characteristics, and, for the first time, reviews DCE health state valuation studies published in the Chinese language.

Methods English language databases PubMed and Cochrane, and Chinese language databases Wanfang and CNKI were searched using the self-developed search terms. Health state valuation or methodology study papers were included if the study used DCE data to generate a value set for a preference-based measure. Key information extracted included DCE study design strategies applied, methods for anchoring the latent coefficient on to a 0–1 QALY scale and data analysis methods.

Results Sixty-five studies were included; one Chinese language publication and 64 English language publications. The number of health state valuation studies using DCE has rapidly increased in recent years and these have been conducted in more countries than prior to 2018. Wide usage of DCE with duration attributes, D-efficient design and models accounting for heterogeneity has continued in recent years. Although more methodological consensus has been found than in studies conducted prior to 2018, this consensus may be driven by valuation studies for common measures with an international protocol (the 'model' valuation research). Valuing long measures with well-being attributes attracted attention and more realistic design strategies (e.g., inconstant time preference, efficient design and implausible states design) were identified. However, more qualitative and quantitative methodology study is still necessary to evaluate the effect of those new methods.

Conclusions The use of DCEs in health state valuation continues to grow dramatically and the methodology progress makes the method more reliable and pragmatic. However, study design is driven by international protocols and method selection is not always justified. There is no gold standard for DCE design, presentation format or anchoring method. More qualitative and quantitative methodology study is recommended to evaluate the effect of new methods before researchers make methodology decisions.

1 Background

Discrete choice experiment (DCE) has been increasingly employed by health economists in recent years [1]. Carson and Louviere [2] defined DCE as a 'general preference elicitation' survey approach, where respondents are asked to choose between two or more alternatives, where at least one attribute is 'systematically varied'. Individual-level or societal-level preference would be calculated indirectly with the choice data. In the field of health economics, one common use of DCE is to generate health state utility values, which are also referred to as health state preference scores or preference weights [3], where the health states are often defined by preference-based measures (PBMs) [4]. The health state

✉ Haode Wang
hwang165@sheffield.ac.uk

[1] School of Health and Related Research (ScHARR), University of Sheffield, 30 Regent St, Sheffield City Centre, Sheffield S1 4DA, UK

[2] Health Policy Research Unit, Shenzhen Health Development Research and Data Management Center, Shenzhen, Guangdong Province, China

Key Points for Decision Makers

In recent years, the use of discrete choice experiments has been increasingly for health state valuation, especially in developing countries.

Nested methodology, including valuing EQ measures with EQ-VT, EORTC QLU-C10D and SF-6D, are identified and have influenced the design strategies of many studies.

There are methodology uncertainties around time preference, duration level selection and efficient design method. Researchers should fully understand the advantages, disadvantages, and participant characteristics before making a methodology decision.

utility values, multiplied by duration, are quality-adjusted life-years (QALYs) in health technology assessment (HTA).

Analysing discrete choice data follows random utility theory (RUT) [5]. Random utility theory posits that selection indicates an individual preference, and random factors explain factors not accounted for. The characteristics theory of demand, which states that consumer utility is derived from good characteristics instead of real content, served as a theoretical rationale for explaining revealed preference with a pre-determined descriptive system [5, 6]. The PBM descriptive system is naturally suitable for designing discrete choice scenarios, where the health states are described by a limited number of attributes. The latent regression coefficients reflect the ordinal utility effect of each attribute levels. To generate health state values on the QALY scale, the latent preference weights should be anchored onto a full health to dead (1–0) scale [7]. The anchored preference weights are often referred to as 'value sets', as they are used to generate utility values for all health states defined by the PBM.

Since Hakim and Pathak [8] reported the first DCE study measuring health state preferences, DCE has been broadly employed by many health state valuation studies [9, 10] and is also recommended in the European Quality of Life 5 Dimensions 5 Level Version (EQ-5D-5L) valuation protocol EQ-VT [11]. Published comprehensive reviews of the health state valuation DCE literature cover the periods 2007–2018 and 1999–August 2018 [9, 10]. However, the DCE health state valuation literature is rapidly expanding, and the published reviews may not fully summarise the methods being used at this time. This paper updates these published systematic reviews to cover the period June 2018 to November 2022 and reports on the new progress and trends in the field of DCEs in the health state valuation literature since the published reviews.

2 Method

2.1 Literature Search

This literature review was the first to cover both English and Chinese databases. In recent years, DCE has been applied in health state utility generation in many developing countries including China (e.g., generating Chinese value set for the measure Short-Form Six-Dimension version 2 [SF-6D v2] using DCE [12]), but there were no systematically searched studies in non-English literature. We identified terminology for the search terms by considering earlier systematic reviews [9, 10] and translated the English keywords into Chinese. We searched English databases PubMed and Cochrane, and Chinese databases Wanfang and CNKI.

The search terms included descriptive keywords of discrete choice survey (e.g., discrete choice experiment, choice experiments, choice modelling and DCE, etc.), health state valuation (e.g., value set generation, etc.) and Multi-Attribute Utility Instrument (MAUI) (e.g., preference-based measure, PBMs, MAUIs, European Quality of Life 5 Dimensions [EQ-5D], etc.). The research group conducted a scoping review to identify the various names and Chinese translations of the DCE method (i.e., paired comparison, case 3 Best-Worst Scaling [BWS] etc.). The full English and Chinese search terms are included in Appendix 1 in the Supplementary Material. The first-round literature search was completed in March 2021 and was updated in November 2022 to ensure all latest papers were included.

2.2 Inclusion and Exclusion Criteria

Health state valuation or methodology study papers were included if the study used DCE design or paired Case 3 (multi-profile case) BWS data to generate a value set for a PBMs. Case 3 BWS asked respondents to select the best and worst scenarios with more than one multiple-attribute profile, where the design was in line with DCE [13, 14]. Papers were excluded if:

1. Only non-DCE methods were used [15].
2. DCE studies targeting a monetary parameter ratio or willingness-to-pay (WTP) for a certain intervention.
3. Studies valuing partial health states where not all attributes were considered, or health states that were not derived from Preference-Based Measure (PBM), where a value set cannot be developed.
4. Quantitative studies using DCE but not reporting the statistical analysis results, reviews and qualitative studies.
5. Papers published before June 2018.

6. Data generated from software simulation instead of real-world survey, or where the study design is not reported, conference abstracts where full text was not available, and replicated articles in various languages were excluded.

2.3 Data Extraction

The data extraction used a pre-designed data extraction sheet. The data extraction framework included: (1) study general information (sample information, measure and data descriptive characteristics); (2) study design (attributes and levels used, attribute categories, scenario and choice set numbers, anchoring method, question asked and statistical analysis strategy); (3) analysis (latent or anchored result and logical consistency); and (4) reported research limitations and recommendations regarding DCE methodological choices. The information in the data extraction sheet was identified as important in previous reviews [9, 10].

3 Results

3.1 Identified Studies

The search identified 1133 English language records and 46 Chinese language studies using DCE and preference elicitation search terms, where 1172 articles were included after duplicate checks. A total of 1106 records were from PubMed, 20 articles from the Cochrane database, 16 articles from Wanfang, and 30 from CNKI. All studies reported DCE study design and no case 3 BWS articles were identified. After screening titles and abstracts, 1063 articles were excluded, leaving 109 articles. The assessment of full articles excluded a further 44 articles. One Chinese language article, and 64 English articles satisfied the inclusion criteria (Table 1).

3.2 Trend of Publication

An increasing number of works were identified in the reviewed years in compared with the published reviews up to 2018 [9, 10]. More papers were published in 2021 ($n = 19$) than 2020 ($n = 17$) and 2019 ($n = 9$) but cannot easily be compared to 2022 ($n = 15$), since this is not a full calendar year. The majority of studies were conducted in Organisation for Economic Co-operation and Development (OECD) countries, examples including the UK ($n = 13$), the USA ($n = 6$), Australia ($n = 6$) and Netherlands ($n = 7$). Other countries with more than one research identified were Germany, China ($n = 4$ for each country), Italy ($n = 3$), Canada, France, Poland, Spain, Hungary and Slovenia ($n = 2$ for each country). Denmark, Egypt, Ethiopia, Malaysia, Mexico, New Zealand, Japan, Peru, Portugal, Russia, Slovenia, Tunisia, Philippines and Thailand all provided one. Compared with published reviews in 1999–2018, there was an increase in the proportion of studies coming from 'developing' countries (15% in 1999–2018 compared with 25% in 2019–2022), but the three leading publication sources were the UK, the USA, and the Netherlands (Fig. 1).

3.3 Sample Size and PBM Measures

Most studies ($n = 60$) sampled the general population, and stratified the respondents by gender, age, educational level and region. Other studies collected data from adolescents [52], parents [44], diabetic macular oedema patients [28], elderly group [35, 65] or people with haemophilia [57]. The sample size varied among the studies. Forty-nine studies (46 valued by general public and 3 valued by a specific group) interviewed over 1000 respondents, with a sample size ranging from 220 to 13,623 (Table 1). The average sample size was 1704.

The proportion of studies administrated online was similar to the previous review. Thirty-seven studies (60%) collected data with an online panel. Mulhern et al. [10] identified 37 (59%) of all the papers employed online administration mode. Of the 25 off-line studies, 21 employed software-assistant data collection. Two studies used mixed data collection strategy and one study did not mention their data gathering method (Table 2).

EuroQol health-related quality of life (HRQoL) (5D-3L, 5D-5L and 5D-Y) measures were the most valued PBMs. The EuroQol international protocols for valuing EQ-5D-5L [11] recommend using the time trade-off discrete choice experiment (TTO-DCE) method; and 21 papers generated EQ-5D-5L value set under the recommended framework. Nine studies generated EQ-5D-Y value sets or assessed preference heterogeneity. Other studies measured generic PBMs including EQ-5D-3L ($n = 4$), EQ-5D bolt-on/bolt-off measures ($n = 2$), SF-6D v2 ($n = 4$), EQ-5D-5L plus Adult Social Care Outcomes Toolkit (ASCOT) ($n = 1$), the informal caregivers' life quality measure CarerQol-7D ($n = 8$) and infant health-related quality of life instrument measure (IQI) ($n = 1$). Various condition-specific PBMs were also valued, such as the European Organisation for Research and Treatment of Cancer (EORTC) utility measure instrument QLU-C10D [82] ($n = 6$), the impact of self-management on quality of life in diabetes measure HASMID, a tool for palliative and supportive care ICECAP-SCM [25], diabetic retinopathy measure DRU-I [20], traumatic brain injury outcome measure QOLIBRI-OS [77], Alzheimer measure AD-5D [24] and cerebral palsy measure CP-6D [19, 24] (Fig. 2). Some measures

Table 1 Study categorisation

Study	Year	Categorisation		Characteristics		Measure
		Data source	Research objective	Country[a]	Sample	
Al Shabasy et al. [16]	2022	Primary	Value set development	Egypt	General public	EQ-5D-5L
Andrade et al. [17]	2020	Primary	Value set development	French	General public	EQ-5D-5L
Augustovski et al. [18]	2020	Primary	Methodology research	Peru	General public	EQ-5D-5L
Bahrampour et al. [19]	2021	Primary	Value set development	Australia	General public	CP-6D
Baji et al. [20]	2020	Primary	Methodology research	Hungary, Poland, Slovenia	General public	CarerQol-7D
Bouckaert et al. [21]	2021	Primary	Value set development	Belgium	General public	EQ-5D-5L
Chemli et al. [22]	2021	Primary	Value set development	Tunisia	General public	EQ-5D-3L
Chen et al. [23]	2021	Primary	Value set development	Australia	General public	QCE
Comans et al. [24]	2020	Primary	Preference comparison	Australia	General public	AD-5D
Dams et al. [25]	2021	Primary	Value set development	Germany	General public	ICECAP-SCM
Doherty et al. [26]	2021	Secondary	Methodology research	Ireland	General public	EQ-5D-5L
Dufresne et al. [27]	2021	Primary	Value set development	Canada	0–17 children and patients	SF-6Dv2
Fenwick et al. [28]	2020	Primary	Value set development	Australia	Patients	DRU-I
Ferreira et al. [29]	2019	Primary	Value set development	Portugal	General public	EQ-5D-5L
Finch et al. [30]	2021	Primary	Value set development	Italy	General public	EQ-5D-5L
Finch et al. [31]	2021	Primary	Value set development	Spain	General public	QLU-C10D
Gamper et al. [32]	2020	Primary	Value set development	Austria, Italy, Poland	General public	QLU-C10D
Gutierrez-Delgado et al. [33]	2021	Primary	Value set development	Mexico	General public	EQ-5D-5L
Hansen et al. [34]	2022	Secondary	Methodology research	Norway, Netherlands and Austria	General public	EQ-5D-5L
Himmler et al. [35]	2022	Primary	Value set development	Netherlands	Elderly people	WOOP
Hoogendoorn et al. [36]	2019	Primary	Preference comparison	Netherlands	General public	EQ-5D-5L with bolt-on
Jansen et al. [37]	2021	Primary	Value set development	Netherlands	General public	QLU-C10D
Jensen et al. [38]	2021	Primary	Value set development	Denmark	General public	EQ-5D-5L
Jiang et al. [39]	2022	Primary	Value set development	USA	General public	Neck Disability Index
Jonker et al. [40]	2019	Primary	Methodology research	Netherlands	General public	EQ-5D-5L
Jyani et al. [41]	2022	Primary	Value set development	India	General public	EQ-5D-5L
Kemmler et al. [42]	2019	Primary	Value set development	Germany	General public	QLU-C10D (Germany ½ versions)
King et al. [43]	2021	Primary	Value set development	Australia	General public	FACT-8D
Krabbe et al. [44]	2020	Primary	Value set development	Hong Kong, UK, USA	General public and primary caregivers	IQI
Kreimeier et al. [45]	2022	Primary	Value set development	Germany	General public	EQ-5D-Y
Lim et al. [46]	2018	Primary	Methodology research	Netherlands	General public	EQ-5D-5L
Ludwig et al. [47]	2018	Primary	Value set development	Germany	General public	EQ-5D-5L
Malik et al. [48]	2022	Primary	Value set development	Pakistan	General public	EQ-5D-3L
Marten et al. [49]	2020	Secondary	Methodology research	UK	General public	EQ-5D-5L
McTaggart-Cowan et al. [50]	2019	Primary	Value set development	Canada	General public	QLU-C10D
Miguel et al. [51]	2022	Primary	Value set development	Philippines	General public	EQ-5D-5L
Mott et al. [52]	2021	Primary	Value set development	UK	General public and adolescence (age 11 to 17)	EQ-5D-Y-3L
Mulhern et al. [53]	2019	Primary	Value set development	Australia	General public	EQ-5D-5L and ASCOT
Mulhern et al. [54]	2020	Primary	Methodology research and value set development	UK	General public	SF-6Dv2

Table 1 (continued)

Study	Year	Categorisation		Characteristics		Measure
		Data source	Research objective	Country[a]	Sample	
Nerich et al. [55]	2021	Primary	Value set development	France	General public	QLU-C10D
Norman et al. [56]	2019	Primary	Value set development	UK	General public	QLU-C10D
O'Hara et al. [57]	2021	Primary	Methodology research	USA	General public and people with haemophilia	EQ-5D-5L
Omelyanovskiy et al. [58]	2021	Primary	Value set development	Russia	General public	EQ-5D-3L
Pattanaphesaj et al. [59]	2018	Primary	Value set development	Thailand	General public	EQ-5D-5L
Pahuta et al. [60]	2021	Primary	Value set development	USA	General public	SOSGOQ-8D
Pickard et al. [61]	2019	Primary	Value set development	USA	General public	EQ-5D-5L
Prevolnik and Ogorevc [62]	2021	Primary	Value set development	Slovenia	General public	EQ-5D-Y
Ramos-Goñi et al. [63]	2022	Primary	Methodology research and value set development	Spain	General public	EQ-5D-Y
Ramos-Goñi et al. [64]	2022	Primary	Methodology research	USA and UK	General public	EQ-5D-Y
Ratclife et al. [65]	2022	Primary	Value set development	Australia	Home care and residential care aged people	QOL-ACC
Rencz et al. [66]	2022	Primary	Value set development	Hungary	General public	EQ-5D-Y
Revicki et al. [67]	2021	Primary	Value set development	USA	General public	QLU-C10D
Rogers et al. [68]	2022	Primary	Value set development	UK	Adolescents and General public	CARIES-QC-U
Rowen et al. [69]	2018	Primary	Value set development	UK	General public	HASMID
Rowen et al. [70]	2021	Primary	Methodology research	UK	General public	DMD-QoL
Roudijk et al. [71]	2022	Primary	Value set development	Netherlands	General public	EQ-5D-Y
Shafie et al. [72]	2019	Primary	Value set development	Malaysia	General public	EQ-5D-5L
Shah et al. [73]	2020	Primary	Methodology research	UK	General public	EQ-5D-Y-3L and EQ-5D-3L
Shiroiwa et al. [74]	2021	Primary	Value set development	Japan	General public	EQ-5D-Y
Sullivan et al. [75]	2020	Primary	Value set development and Methodology research	New Zealand	General public	EQ-5D-5L
Tsuchiya et al. [76]	2019	Primary	Methodology research	UK	General public	EQ-5D-3L EQ-4D-3L
Voormolen et al. [77]	2020	Primary	Value set development	Italy, Netherlands, UK	General public	QOLIBRI-OS
Webb et al. [78]	2020	Primary	Methodology research	UK	General public	EQ-5D-3L
Welie et al. [79]	2020	Primary	Value set development	Ethiopia	General public	EQ-5D-5L
Wu et al. [12]	2021	Primary	Value set development	China	General public	SF-6Dv2
Wu et al. [12, 80]	2020	Primary	Methodology research	China	General public	SF-6Dv2
Zhu et al. [81]	2022	Primary	Value set development	China	General public	CQ-11D

Shah et al. [73] used EQ-5D-Y and EQ-5D-3L

AD-5D Alzheimer's Disease Five Dimensions, *ASCOT* Adult Social Care Outcomes Toolkit, *CARIES-QC-U* Caries Impacts and Experiences Questionnaire for Children, *CarerQol-7D* Care related quality of life-7 dimensions, *CP-6D* cerebral palsy quality of life-6 dimensions, *CQ-11D* Chinese medicine quality of life-11 dimensions, *DMD-QoL* Duchenne muscular dystrophy quality of life, *DRU-I* Diabetic Retinopathy Utility Index, *EQ-5D-3L/5L* European Quality of Life 5 Dimensions 3 Level/5 Level Version, *EQ-5D-Y* European Quality of Life 5 Dimensions Youth Bolt on/off EQ-5D with bolt on/off dimensions, *FACT-8D* Functional Assessment of Cancer Therapy Eight Dimension, *HASMID* Health and Self-Management in Diabetes, *ICECAP-SCM ICECAP* Supportive Care Measure, *IQI* infant health-related quality of life instrument, *QLU-C10D* European Organisation for Research and Treatment of Cancer (EORTC) cancer utility measure instrument, *QOL-ACC* quality-of-life aged care consumers, *QOLIBRI-OS* Quality of Life after Brain Injury overall scale, *SF-6D v2* Short-Form Six-Dimension Version 2, *SOSGOQ-8D* Spine Oncology Study Group Outcomes Questionnaire-8 dimensions, *WOOP* well-being of older people

[a]Studies included population in more than one country

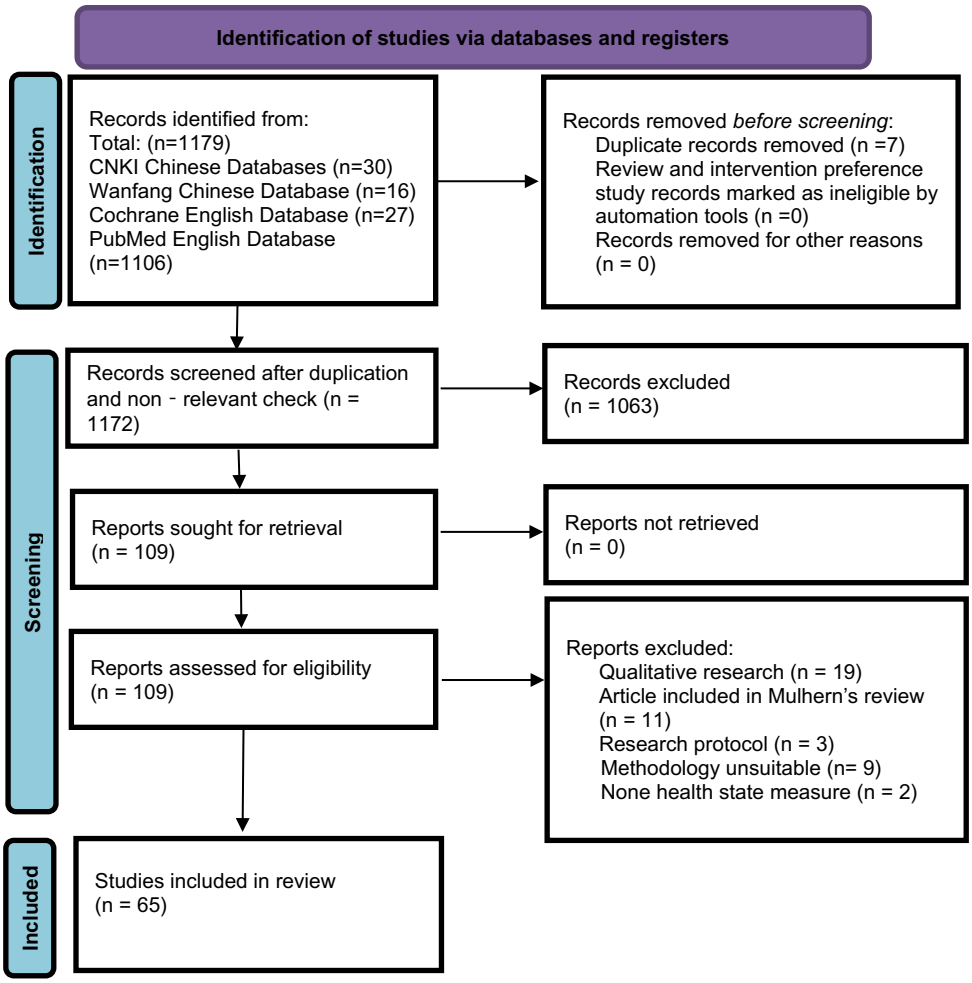

Fig. 1 Study selection (PRISMA STANDARD). Our review was conducted in line with Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA [100])

covered a broader well=being change instead of limiting the measurement domains in the regime of HRQoL [84] and this review found that 14 out of 65 articles valued measures covering well-being.

3.4 Attributes and Choice Sets

Included measures were described by different numbers of attributes, ranging from 5 to 13. Thirty-two studies included a duration attribute ($n = 29$) or included a 'death' scenario ($n = 3$) to collect relative preference for anchoring. The range of duration levels for all the included studies was 2 months to 15 years. The most common duration levels as recommended by the QLU-C10D valuation studies were 1, 2, 5, and 10 years [56, 83]. However, other condition-specific PBM and SF-6D valuation studies had 1-, 4-, 7- and 10-year duration levels. We did not find a study that combined the two duration-level designs together. Three studies reported that they conducted a qualitative interview for deciding duration levels, but evidence from published studies was the most common source of duration levels. Most studies ($n = 28$) selected the n-1 number of duration levels compared with other attributes. However, the EQ-5D severity-stratified study from Lim et al. [46] used 12 duration levels, although there was no explanation for the duration selection process.

Valuation studies can present choice tasks with either paired scenarios [12, 73] or triplet scenarios [85]. Pairs ($n = 65$) were used more commonly than triplets ($n = 3$). The additional scenario was either described as dead or as the worst health state of valued measure. The total number of choice sets ranged from 28 to 960, with an average number of 269 (Table 2). Over 60% ($n = 34$) of all studies presented fewer than 12 DCE tasks per respondent, and the number of tasks varied from 7 to 28 (including dominant task or consistency test task).

3.5 Study Design and Presentation

The choice of design and presentation method is influenced by protocols or early valuation works. Apart from EQ measures, the 'model' valuation studies led the design strategy

Table 2 Experiment design characteristics

Characteristics	Level	Identified studies
Attributes number (range: 5–13)	Range: 5–13	
	5	38
	6–7	10
	8–10	6
	11–12	10
	13	1
Number of levels	3	15
	4	11
	5	34
	6	5
Anchoring	Anchoring with cTTO data	24
	Anchoring with duration	29
	Anchoring with VAS data	3
	Others[a] (rescaling)	5
	No anchoring	4
Choice set	Pairs	62
	Triples[b]	3
Choice tasks per participant	Range: 7–28	
	≤12	38
	13–28	25
	Not mention	2
Survey mode	Interview offline	25
	Interview online	37
	Mixed	2
	Not mention	1
Question asked	Prefer	32
	Better	10
	"Which one do you pick"	4
	Following EQ-VT (v1/v2 question format mentioned)	14
	Not mention	5
Total number of choice sets	Range: 28–960	
	≤120	20
	121–196	26
	197–960	16
	Not mentioned	3
	In total	65

cTTO composite time trade-off, *EQ-VT* EuroQol Valuation Technology, *VAS* visual analogue scale

[a]Re-scaling anchoring option includes re-scale with existing tariff or re-scale with the minimum/maximum utility value

[b]All of the studies included used death as the third state

choice. For example, all QLU-C10D valuation studies employed Australia and UK valuation protocols [56, 83], where the DCE was recommended as the single method for value set development. Typically, it is not feasible to value all pairwise comparisons with full factorial design in valuation studies. Almost all studies applied mathematical algorithms to eliminate the DCE scenario number and generate an efficient design. The extracted efficient design approaches are optimising D-efficiency/C-efficiency with non-zero ($n = 26$) and zero priors ($n = 1$). Informative prior values could be applied in minimising the D-error in efficient design, where the priors may be from a pilot study or extracted data from published articles (e.g., taking Netherlands EQ-5D-5L values as the fixed prior values for 15 EQ-5D valuation studies). Studies with prior distribution information and value uncertainty iteratively extracted priors with a Bayesian method ($n = 25$). As recommended by the EuroQol EQ-5D-5L international valuation protocol, Bayesian-efficient design has been applied by a larger proportion of DCE studies (38% vs 30% from 1999–2018 review [10]). Apart from informative prior, researchers may use non-informative prior values, which were applied to design a small-scale pilot study, followed by design update with the pilot data values and distributions. Other design strategies included fractional factorial design ($n = 8$), C-efficient design ($n = 1$), full/fractional factorial design ($n = 1$), and others (hand selection and self-adaptive) ($n = 2$). Four studies applied mixed design strategy, including both D-efficient and suppressing unrealistic/severe health states using hand selection (Table 3).

A study design with all attributes varied in each choice set provided higher statistical efficiency with given number of respondents, yet it simultaneously increased respondents' confusion, misunderstanding, and dropout rate [86]. Twenty-five studies presented the choice set with strategies to reduce the cognitive burden and increase respondent participation. These strategies included overlap and visually attractive choice set presentation. Twenty-two studies introduced within-dimension overlap, and 18 studies highlighted the dimensions that differ within a choice set using different colours (yellow or light grey). Three studies presented the attributes that differed within a choice set before fixed overlap levels, where the other studies ($n = 62$) showed mixed fixed and differed attributes in each health state. All valuation studies using a long measure (attribute number larger than 9) applied some degree of attribute overlap.

Included studies involved a randomisation process of choice sets or sample randomisation to increase face validity. Thirty-three studies applied a process of blocking choice sets, including Balanced Incomplete Block Design (BIBD) or partial block design, to guarantee a balanced severity level distribution. Twenty-nine studies randomised the choice sets into blocks without stratification. Respondents in 6 studies answered a fixed number of unduplicated DCE questions. Some studies developed in-block randomisation: 16 studies randomised the choice set order in a fixed DCE block; 12 studies randomised the scenario sequence (left-right

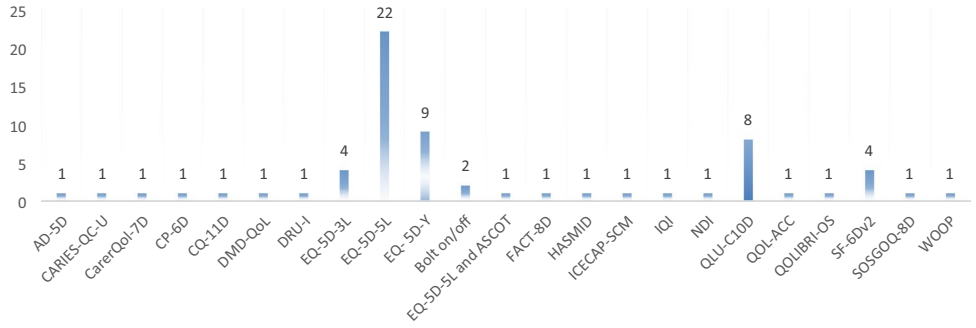

Fig. 2 Measures used in identified articles. *AD-5D* Alzheimer's Disease Five Dimensions, *ASCOT* Adult Social Care Outcomes Toolkit, *CARIES-QC-U* Caries Impacts and Experiences Questionnaire for Children, *CarerQol-7D* Care related quality of life-7 dimensions, *CP-6D* cerebral palsy quality of life-6 dimensions, *CQ-11D* Chinese medicine quality of life-11 dimensions, *DMD-QoL* Duchenne muscular dystrophy quality of life, *DRU-I* Diabetic Retinopathy Utility Index, *EQ-5D-3L/5L* European Quality of Life 5 Dimensions 3 Level/5 Level Version, *EQ-5D-Y* European Quality of Life 5 Dimensions Youth Bolt on/off EQ-5D with bolt on/off dimensions, *FACT-8D* Functional Assessment of Cancer Therapy Eight Dimension, *HASMID* Health and Self-Management in Diabetes, *ICECAP-SCM ICECAP* Supportive Care Measure, *IQI* Infant health-related quality of life instrument, *QLU-C10D* European Organisation for Research and Treatment of Cancer (EORTC) cancer utility measure instrument, *QOL-ACC* quality-of-life aged care consumers, *QOLIBRI-OS* Quality of Life after Brain Injury overall scale, *NDI* neck disability index, *SF-6D v2* Short-Form Six-Dimension Version 2, *SOSGOQ-8D* Spine Oncology Study Group Outcomes Questionnaire-8 dimensions, *WOOP* well-being of older people. Note Shah et al. [73] used EQ-5D-Y and EQ-5D-3L

Table 3 Design method used

Design type	Approach	Identified study
Efficient design	Bayesian efficient design	25
	D-efficient (with fixed/zero prior)	26
	C-efficient (with fixed/zero prior)	2
Fractional factorial	Randomised design and orthogonal method	8
Factorial design	Full factorial/fractional factorial design (including adaptive DCE)	2
Other	Others (hand selection and self-adaptive)	2
	In total	65

The study that used efficient design and excluded the combinations of dimension levels that were considered highly implausible in practice, has been classified as efficient design

DCE discrete choice experiment

randomisation); and 20 studies arranged the measure attributes in a random order (randomised dimensions).

3.6 Statistical Analysis

Table 4 displays the number of articles using different statistical analysis models. The main effect linear utility function ($n = 19$) and main effect interacted with duration ($n = 26$) were the most frequently used model functions. The main effect model captured only single-parameter main effects without interactions or extra dummies, while the main effect interacted with duration model estimated attribute coefficients with duration interactions. Both model specification forms assume that there was no dimensional interaction between PBM attributes [5]. To consider the non-duration interaction, two studies considered the interaction between measure attributes with linear model [25, 69] or included an extra dummy to capture the impact of extreme health states ($n = 3$). Shafie et al. [72] and two other studies used an eight-parameter non-linear constrained model, where the parameter representing level 5 and one parameter for levels 2, 3 and 4 (L2, L3, L4) were included. A hybrid model function ($n = 13$) used both DCE and composite time trade-off (cTTO) data, where the majority of EuroQol measure valuation studies were included this model.

For the regression model, the conditional logit model ($n = 48$) was the starting point of most studies.

A conditional logit model follows that is consistent with the random utility theory and assumes no scale or preference heterogeneity [87]. On the other hand, 33 studies did not consider homogeneous model and applied the mixed logit model ($n = 24$) or latent-class model ($n = 9$) to control for individual heterogeneity. Twenty-four studies used a hybrid model, which jointly modelled both DCE and TTO preference data. Seven studies considered the possible heteroscedasticity issue with conditional logit model, and estimated the scale effect with scale-assessment models [88]. The Zermelo Bradley Terry model (ZBT model) with a unilinear time preference [18] assumption appeared twice, and the mean individual preference model showed once (Table 2). Studies reported

Table 4 Modelling and regression characteristics

Characteristics	Approach	Identified study
Model function	Main effect linear utility function[a]	19
	Main effect interacted with duration (with and without constant time assumption[b])	26
	Main effect with extra term (dead dummy or worst/N3 state)	3
	Hybrid model[c]	13
	Main effect with constrained model (eight-parameter)	3
	Personal value function[d]	1
Regression function[e]	Conditional logit model (random/fix effect, scale-adjusted	48
	Mixed logit/latent-class logit model (heterogeneity model)	33
	Likelihood function (TT0/BWS with DCE data)	24
	Scale-assessment models/poolability	7
	ZBT model with power function	2
	Mean individual preference	1
	Total identified article	65

BWS Best–Worst Scaling, *cTTO* composite time trade-off, *DCE* discrete choice experiment, *ZBT* Zermelo Bradley Terry model

[a]Main effect linear utility function is the model function with only DCE data and considers no dimensional interaction or co-effect with duration attribute

[b]Research considered both interaction with duration and extra term is classified as main effect with extra term

[c]Hybrid model is the main effect function with cTTO or BWS data

[d]The personal value function is a self-adjusted health state valuation function, where the social preference is the average of personal preference

[e]Research can use more than one regression function

model performance with logical judgements: if the 'worse level' has higher latent or anchored disutility value, then the item coefficient would be regarded as inconsistent. Our updated review found that over 60% of all studies reported some degree of inconsistency with the conditional logit model. The significant inconsistency rate is acceptable compared with other valuation methods [9, 10].

3.7 Anchoring

To generate utility values on a QALY scale, latent coefficients should be anchored on the 0 (full health) to 1 (dead) QALY scale, which can be done using a variety of different methods [7], with a specific anchoring method. Sixty-one studies anchored the latent coefficient by using: extra TTO data ($n = 24$); visual analogue scale (VAS) data ($n = 3$); duration attribute for estimating relative preference with time ($n = 29$); re-scaling method with or without additional data ($n = 5$).

4 Discussion

4.1 Summary of Trends

This review generated a richer picture of health state valuation with the discrete choice method. Compared with the published reviews covering the time period 1999–2018 [10] and 2007–2018 [9], this review identified a larger average number of published studies in the reviewed years (2019–2022), and, for the first time, included studies published in the Chinese language. The research concluded that not only innovative DCE study design methods used in recent years, but also a widening range of countries launched large-scale experiments to test the feasibility of this method and reached positive outcomes. This trend indicates that DCE is a valuable and feasible methodology for valuing health states and has attracted attention in many countries [48].

4.2 Study Design

Some of the methodology consensus reported by the Mulhern et al. [10] review has been reinforced in the last three years. Online DCE with the general population was popular during the COVID-19 pandemic. Online DCE is a less costly and more flexible option for a large-scale survey. However, it is worth noting that in order to undertake surveys online, participants require internet connection, an appropriate device and some level of computer literacy. The equipment requirement may affect the representativeness of an online survey, and data quality can be lower [89]. Researchers should consider the representativeness of a mode when conducting the study and use data quality-control strategies (for example, a time check) to minimise and assess the mode of administration influence.

The experimental design selection is an important step and is largely influenced by "model" designs. The valuation of EuroQol measures boosted the population of EQ-VT (version 1 and 2). The Australian EORTC QLU-C10D valuation method [90] and the SF-6D valuation in the UK, where a DCE with duration method was used, have been referred to in valuation studies in many other countries. One advantage of using the DCE with duration design is that the DCE data can be anchored without extra cardinal data [86]. Common DCE with duration levels for the duration attribute of 1, 4, 7, or 10 years/1, 2, 5, or 10 years are from the SF-6D v2 and QLU-C10D valuation studies [54, 90]. The increasing trend of using D-efficient DCE design with priors considering no dimensional interaction is influenced by "model" or protocol designs. Although the main effect and the interaction with duration designs remained dominant, we should consider study designs catered for health-attribute interactions. An ISPOR report [91] indicated that estimating interaction effects among measured attributes should rely on quantitative analysis instead of assuming that the interactions are not statistically significant. In conclusion, we found that the revealed "methodological consensus" might be influenced by the measure that is valued, rather than academic agreement [10]. Whilst this reflects a policy-making demand for generic PBM tariffs and cancer-specific tariffs, it is recommended that, before making a decision, the attribute interaction significance, design strategy and selection of the duration levels be explored further using qualitative methods.

The reporting of models accounting for heterogeneity and heteroskedasticity has become more common in recent DCE valuation studies. Although 62% of all the included studies reported inconsistencies (inconsistency for the coefficients and inconsistency for the health state rating) with the conditional logit model, it is still worthwhile to test conditional logit model as it takes advantage of the largest amount of information [92]. For individual-level inconsistencies, Doherty et al. (2021) and Wang et al. (2021) evaluated attribute non-attendance [26, 93] and concluded that some respondents are less likely to consider the physical dimensions. Non-attendance of attributes happens when respondents use simplified strategies or heuristics to make decisions [94], which creates a systematic bias violating the discrete choice assumption that the individual considers all the information and may not be identified under homogeneity assumption [95]. The heterogeneity models tend to be more promising practices with prior research group knowledge and large samples. With the support of stratified-group evidence or identified preference variation, heterogeneity models can be selected without testing conditional logit model.

4.3 Measure and Prior Selection

The preference for valuing EQ-5D-5L and EORTC QLQ-C30 with informative priors increased the proportion of efficient design with fixed and Bayesian priors. For studies without prior information, using non-informative priors for the pilot design and updated with Bayesian method is commonly applied. It is theoretically expected that using Bayesian design may maximise logarithm of the determinant estimator information matrix [96], with a price of extra effort on prior data collection. However, using non-informative priors and informative priors may not cause systematic differences. Kesselsa et al. [97] presented a case study where non-informative prior efficient design did not cause variation with a sample size greater than 1000. It is reasonable for valuation studies with a large sample size to use non-informative priors, without pilot information update in the design stage.

The last trend is for methods applied for valuing measures with a large number of health and well-being attributes. It is recognised that outcome attributes beyond health, such as carer well-being, are not fully covered by physical and mental health measures [4, 98]. More attributes will be considered if we generate QALY beyond health. Sullivan et al. [75] suggest a replicate adaptive design strategy called PAPRIKA to value the long measures. The experiment consisted of an average of 5 binary search questions (to find the dividing line of better or worse than death) and 20 adaptive DCE pairs (to generate value sets). However, there are still many uncertainties around long measure valuation studies and statistical analysis. For example, the constant time preference assumption may not hold and the general relationship between overall health and well-being may significantly interact.

4.4 Remaining Questions

Although there was a wide range of health state valuation applications using DCE, some questions from the previous review remain [10]. First, the popularity of anchoring with duration, question wording and data analysis is driven by

"model" international protocols. It must be noted that standardisation is a double-edged sword instead of a promised solution. Researchers should consider pros and cons of all feasible options before valuing new measures, instead of picking the standard method of EQ measures, EORTC QLQ-C30 and SF-6D. Second, DCE study design details depend on social-demographic factors and participant background. There is no single best answer. Third, there is still no gold standard for study design, especially for long measures.

This review identified several remaining research gaps. The first was around the modelling function used. The majority of DCE data were modelled using the main effect approach or the main effect interacted with duration approach, where the non-duration interaction term was not considered [53]. Future methodology studies are required to instruct the interaction modelling and interpretation of qualitative evidence [99]. Second, there is no comprehensive feasibility and efficiency comparisons of various DCE study design strategies. Although there is some empirical evidence to suggest [94] that overlapping and colour-coding design strategies reduced the dropout rate of EQ-5D-5L valuation studies by 4%, and dimension-level and question-level randomisation have little influence on the results [100], a more comprehensive comparison of design methods, especially with a broader health dimensions, will be valuable to instruct long-measure valuation. Further evidence is required to generate a more integrated criteria, considering statistical, respondent and cost efficiency for study design selection.

Regarding the duration level selection and time preference assumption, there is still no consensus. Future qualitative and quantitative studies are required to compare various duration level selection patterns and time preference functions, to understand how time influences result consistency.

5 Limitations

One limitation of this review is that the author found some conference reports in either Chinese or English language using the discrete choice method. However, there was no peer reviewed publication records for those works, and there was only one Chinese language paper that satisfied the inclusion criteria. A quality evaluation of the studies was not undertaken in this review, as the methodology differences limited the cross-comparison validity.

6 Conclusion

This review provides up-to-date information of health state valuation studies using the DCE method. The number of published studies continues to grow dramatically and there is more homogeneity in the methods used in the published articles, but this is likely impacted using international protocols for some measures. Like previous reviews, this study did not find a 'gold standard' or consensus in the DCE health state valuation study design strategy or universally accepted criteria to evaluate the validity of included design strategies. Further research, especially qualitative research to assess the impact of different methodologies, is recommended to inform practice in health state valuation using DCE.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40258-023-00794-9.

Author Contributions HW carried out the literature search in English literature, extracted the data, led the data synthesis and interpretation and developed the first draft of the manuscript. LJ searched the Chinese literature database and checked the result for qualified Chinese literatures. DR and JB supported the design of data extraction sheet, data extraction process and were involved in the development of the manuscript. Author would like to express special thanks for the funder EuroQol group.

Funding This review was partly funded by the EuroQol Group. Mr. Wang was funded by EuroQol Group Scholarship.

Data Availability Study characteristics extracted from the literature to generate the review dataset. Due to the large amount of data extracted and large number of literature in the database, please contact Haode Wang (hwang165@sheffield.ac.uk) for accessing the full data.

Code Availability Not applicable

Declarations

Conflicts of Interest HW, DR and JB are funded by the EuroQol Group. There are no potential conflicts of interest of LJ.

Ethics Approval Not applicable

Consent to Participate Not applicable

Consent for Publication (from Patients/Participants) Not applicable

References

1. Soekhai V, et al. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019;37(2):201–26.

2. Carson RT, Louviere JJ. A common nomenclature for stated preference elicitation approaches. Environ Resource Econ. 2011;49(4):539–59.
3. Lancsar E, et al. The relative value of different QALY types. J Health Econ. 2020;70: 102303.
4. Drummond MF, et al. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2015.
5. Clark MD, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883–902.
6. Lancaster KJ. A new approach to consumer theory. J Polit Econ. 1966;74(2):132–57.
7. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Making. 2015;35(3):328–40.
8. Hakim Z, Pathak DS. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Econ. 1999;8(2):103–16.
9. Bahrampour M, et al. Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods. Eur J Health Econ. 2020;21(7):983–92.
10. Mulhern B, et al. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. Pharmacoeconomics. 2019;37(1):29–43.
11. Oppe M, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17(4):445–53.
12. Wu J, et al. Valuation of SF-6Dv2 health states in China using time trade-off and discrete-choice experiment with a duration dimension. Pharmacoeconomics. 2021;39(5):521–35.
13. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res. 2010;10(3):259–67.
14. Marley AAJ, Louviere JJ, Flynn TN. The BWS multi-profile case. In: Marley AJ, Louviere JJ, Flynn TN, editors. Best–worst scaling: theory, methods and applications. Cambridge: Cambridge University Press; 2015. p. 89–113.
15. Wittenberg E, et al. Using best–worst scaling to understand patient priorities: a case example of Papanicolaou tests for homeless women. Ann Fam Med. 2016;14(4):359–64.
16. Al Shabasy S, et al. The EQ-5D-5L valuation study in Egypt. Pharmacoeconomics. 2022;40(4):433–47.
17. Andrade LF, et al. A French value set for the EQ-5D-5L. Pharmacoeconomics. 2020;38(4):413–25.
18. Augustovski F, et al. Peruvian valuation of the EQ-5D-5L: a direct comparison of time trade-off and discrete choice experiments. Value Health. 2020;23(7):880–8.
19. Bahrampour M, et al. Utility values for the CP-6D, a cerebral palsy-specific multi-attribute utility instrument, using a discrete choice experiment. Patient. 2021;14(1):129–38.
20. Baji P, et al. Development of population tariffs for the CarerQol instrument for Hungary, Poland and Slovenia: a discrete choice experiment study to measure the burden of informal caregiving. Pharmacoeconomics. 2020;38(6):633–43.
21. Bouckaert N, et al. An EQ-5D-5L value set for Belgium. Pharmacoecon Open. 2022;6:823–36.
22. Chemli J, et al. Valuing health-related quality of life using a hybrid approach: Tunisian value set for the EQ-5D-3L. Qual Life Res. 2021;30:1445–55.
23. Chen G, et al. Quality of care experience in aged care: an Australia-Wide discrete choice experiment to elicit preference weights. Soc Sci Med. 2021;289: 114440.
24. Comans TA, et al. Valuing the AD-5D dementia utility instrument: an estimation of a general population tariff. Pharmacoeconomics. 2020;38(8):871–81.
25. Dams J, et al. German tariffs for the ICECAP-supportive care measure (ICECAP-SCM) for use in economic evaluations at the end of life. Eur J Health Econ. 2021;22(3):365–80.
26. Doherty E, et al. An exploration on attribute non-attendance using discrete choice experiment data from the Irish EQ-5D-5L national valuation study. Pharmacoecon Open. 2021;5(2):237–44.
27. Dufresne É, et al. SF-6Dv2 preference value set for health utility in food allergy. Allergy. 2021;76(1):326–38.
28. Fenwick EK, et al. Validation of a novel diabetic retinopathy utility index using discrete choice experiments. Br J Ophthalmol. 2020;104(2):188–93.
29. Ferreira PL, et al. A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. Qual Life Res. 2019;28(12):3163–75.
30. Finch AP, et al. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. Soc Sci Med. 2022;292: 114519.
31. Finch AP, et al. Estimation of an EORTC QLU-C10 value set for Spain using a discrete choice experiment. Pharmacoeconomics. 2021;39(9):1085–98.
32. Gamper EM, et al. EORTC QLU-C10D value sets for Austria, Italy, and Poland. Qual Life Res. 2020;29(9):2485–95.
33. Gutierrez-Delgado C, et al. EQ-5D-5L health-state values for the Mexican population. Appl Health Econ Health Policy. 2021;19(6):905–14.
34. Hansen TM, Stavem K, Rand K. Sample size and model prediction accuracy in EQ-5D-5L valuations studies: expected out-of-sample accuracy based on resampling with different sample sizes and alternative model specifications. MDM Policy Pract. 2022;7(1):23814683221083840.
35. Himmler S, et al. Estimating an anchored utility tariff for the well-being of older people measure (WOOP) for the Netherlands. Soc Sci Med. 2022;301: 114901.
36. Hoogendoorn M, et al. Exploring the impact of adding a respiratory dimension to the EQ-5D-5L. Med Decis Making. 2019;39(4):393–404.
37. Jansen F, et al. Netherlands utility weights for the EORTC cancer-specific utility instrument: the Netherlands EORTC QLU-C10D. Qual Life Res. 2021;30:2009–19.
38. Jensen CE, et al. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. Appl Health Econ Health Policy. 2021;19(4):579–91.
39. Jiang EX, et al. Calculating ex-ante utilities from the modified Japanese Orthopedic Association score: a prerequisite for quantifying the value of care for cervical myelopathy. Spine (Phila Pa 1976). 2022;47(7):523–30.
40. Jonker MF, et al. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Econ. 2019;28(3):350–63.
41. Jyani G, et al. Development of an EQ-5D value set for India using an extended design (DEVINE) study: the Indian 5-level version EQ-5D value set. Value Health. 2022;25(7):1218–26.
42. Kemmler G, et al. German value sets for the EORTC QLU-C10D, a cancer-specific utility instrument based on the EORTC QLQ-C30. Qual Life Res. 2019;28(12):3197–211.
43. King MT, et al. The functional assessment of cancer therapy eight dimension (FACT-8D), a multi-attribute utility instrument derived from the cancer-specific FACT-general (FACT-G) quality of life questionnaire: development and australian value set. Value Health. 2021;24(6):862–73.
44. Krabbe P, et al. A two-step procedure to generate utilities for the Infant health-related Quality of life Instrument (IQI). PLoS ONE. 2020;15(4): e0230852.

45. Kreimeier S, et al. EQ-5D-Y value set for Germany. Pharmacoeconomics. 2022;40:217–29.
46. Lim S, et al. Severity-stratified discrete choice experiment designs for health state evaluations. Pharmacoeconomics. 2018;36(11):1377–89.
47. Ludwig K, Graf VDSJ, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics. 2018;36(6):663–74.
48. Malik M, et al. A pilot study of valuation methods of the EQ-5D and the impact of literacy, cultural and religious factors on preferences. Value Health Reg Issues. 2022;30:48–58.
49. Marten O, et al. Implausible states: prevalence of EQ-5D-5L states in the general population and its effect on health state valuation. Med Decis Making. 2020;40(6):735–45.
50. McTaggart-Cowan H, et al. The EORTC QLU-C10D: the Canadian valuation study and algorithm to derive cancer-specific utilities from the EORTC QLQ-C30. MDM Policy Pract. 2019;4(1):2381468319842532.
51. Miguel RTD, et al. Estimating the EQ-5D-5L value set for the Philippines. Qual Life Res. 2022;31(9):2763–74.
52. Mott DJ, et al. Valuing EQ-5D-Y-3L health states using a discrete choice experiment: do adult and adolescent preferences differ? Med Decis Making. 2021;41:584–96.
53. Mulhern B, et al. Investigating the relative value of health and social care related quality of life using a discrete choice experiment. Soc Sci Med. 2019;233:28–37.
54. Mulhern BJ, et al. Valuing the SF-6Dv2 classification system in the United Kingdom using a discrete-choice experiment with duration. Med Care. 2020;58(6):566–73.
55. Nerich V, et al. French value-set of the QLU-C10D, a cancer-specific utility measure derived from the QLQ-C30. Appl Health Econ Health Policy. 2021;19(2):191–202.
56. Norman R, et al. U.K. utility weights for the EORTC QLU-C10D. Health Econ. 2019;28(12):1385–401.
57. O'Hara J, et al. Evidence of a disability paradox in patient-reported outcomes in haemophilia. Haemophilia. 2021;27(2):245–52.
58. Omelyanovskiy V, et al. Valuation of the EQ-5D-3L in Russia. Qual Life Res. 2021;30:1997–2007.
59. Pattanaphesaj J, et al. The EQ-5D-5L valuation study in Thailand. Expert Rev Pharmacoecon Outcomes Res. 2018;18(5):551–8.
60. Pahuta MA, et al. Calculating utilities from the Spine Oncology Study Group Outcomes Questionnaire: a necessity for economic and decision analysis. Spine (Phila Pa 1976). 2021;46(17):1165–71.
61. Pickard AS, et al. United States valuation of EQ-5D-5L health states using an international protocol. Value Health. 2019;22(8):931–41.
62. Prevolnik RV, Ogorevc M. EQ-5D-Y value set for Slovenia. Pharmacoeconomics. 2021;39(4):463–71.
63. Ramos-Goñi JM, et al. Accounting for unobservable preference heterogeneity and evaluating alternative anchoring approaches to estimate country-specific EQ-5D-Y value sets: a case study using Spanish preference data. Value Health. 2022;25(5):835–43.
64. Ramos-Goñi JM, et al. Does changing the age of a child to be considered in 3-level version of EQ-5D-Y discrete choice experiment-based valuation studies affect health preferences? Value Health. 2022;25(7):1196–204.
65. Ratcliffe J, et al. Valuing the quality-of-life aged care consumers (QOL-ACC) instrument for quality assessment and economic evaluation. Pharmacoeconomics. 2022;40(11):1069–79.
66. Rencz F, et al. Value set for the EQ-5D-Y-3L in Hungary. Pharmacoeconomics. 2022;40:205–15.
67. Revicki DA, et al. United States utility algorithm for the EORTC QLU-C10D, a multiattribute utility instrument based on a cancer-specific quality-of-life instrument. Med Decis Making. 2021;41(4):485–501.
68. Rogers HJ, et al. Adolescent valuation of CARIES-QC-U: a child-centred preference-based measure of dental caries. Health Qual Life Outcomes. 2022;20(1):18.
69. Rowen D, et al. Estimating a preference-based single index measuring the quality-of-life impact of self-management for diabetes. Med Decis Making. 2018;38(6):699–707.
70. Rowen D, et al. Deriving a preference-based measure for people with Duchenne muscular dystrophy from the DMD-QoL. Value Health. 2021;24(10):1499–510.
71. Roudijk B, et al. A value set for the EQ-5D-Y-3L in the Netherlands. Pharmacoeconomics. 2022;40:193–203.
72. Shafie AA, et al. EQ-5D-5L valuation for the Malaysian population. Pharmacoeconomics. 2019;37(5):715–25.
73. Shah KK, et al. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. Eur J Health Econ. 2020;21(7):1091–103.
74. Shiroiwa T, et al. Valuation survey of EQ-5D-Y based on the international common protocol: development of a value set in Japan. Med Decis Making. 2021;41(5):597–606.
75. Sullivan T, et al. A new tool for creating personal and social EQ-5D-5L value sets, including valuing "dead." Soc Sci Med. 2020;246: 112707.
76. Tsuchiya A, et al. Manipulating the 5 dimensions of the EuroQol instrument: the effects on self-reporting actual health and valuing hypothetical health states. Med Decis Making. 2019;39(4):379–92.
77. Voormolen DC, et al. Health-related quality of life after traumatic brain injury: deriving value sets for the QOLIBRI-OS for Italy, The Netherlands and The United Kingdom. Qual Life Res. 2020;29(11):3095–107.
78. Webb E, et al. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. Eur J Health Econ. 2020;21(5):787–800.
79. Welie AG, et al. Valuing health state: an EQ-5D-5L value set for Ethiopians. Value Health Reg Issues. 2020;22:7–14.
80. Xie S, et al. Do discrete choice experiments approaches perform better than time trade-off in eliciting health state utilities? evidence from SF-6Dv2 in China. Value Health. 2020;23(10):1391–9.
81. Zhu W, et al. The Chinese Medicine Life Quality Evaluation Scale (CQ-11D) report (Chinese). China J Pharm Econ. 2022;17(5):16–20.
82. King MT, et al. QLU-C10D: a health state classification system for a multi-attribute utility measure based on the EORTC QLQ-C30. Qual Life Res. 2016;25(3):625–36.
83. King MT, et al. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. Pharmacoeconomics. 2018;36(2):225–38.
84. Brazier J, et al. A review of generic preference-based measures for use in cost-effectiveness models. Pharmacoeconomics. 2017;35(Suppl 1):21–31.
85. Mulhern B, et al. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences. Med Decis Making. 2017;37(3):285–97.
86. Bansback N, et al. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31(1):306–18.
87. Hauber AB, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value in Health. 2016;19(4):300–15.
88. Manahan R, et al. PSAT097 patient preference research: preferred adjunctive medication attributes of adult patients with classic congenital adrenal hyperplasia. J Endocr Soc. 2022;6(Suppl 1):A118.

89. Rowen D, et al. Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. Soc Sci Med. 2022;309: 115227.
90. Norman R, et al. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual Life Res. 2016;25(3):637–49.
91. Reed Johnson F, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value Health. 2013;16(1):3–13.
92. de Bekker-Grob EW, et al. Are healthcare choices predictable? The impact of discrete choice experiment designs and models. Value in Health. 2019;22(9):1050–62.
93. Wang K, et al. Using eye-tracking technology with older people in memory clinics to investigate the impact of mild cognitive impairment on choices for EQ-5D-5L health states preferences. Appl Health Econ Health Policy. 2021;19(1):111–21.
94. Jonker MF, et al. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value Health. 2018;21(8):993–1001.
95. Stolk EA, et al. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value Health. 2010;13(8):1005–13.
96. Gotwalt CM, Jones BA, Steinberg DM. Fast computation of designs robust to parameter uncertainty for nonlinear settings. Technometrics. 2009;51(1):88–95.
97. Kessels R, et al. Rejoinder: the usefulness of Bayesian optimal designs for discrete choice experiments. Appl Stoch Model Bus Ind. 2011;27(3):197–203.
98. Brazier J, Tsuchiya A. Improving cross-sector comparisons: going beyond the health-related QALY. Appl Health Econ Health Policy. 2015;13(6):557–65.
99. Karimi M, Brazier J, Paisley S. How do individuals value health states? A qualitative investigation. Soc Sci Med. 2017;172:80–8.
100. Moher D, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev. 2015;4(1):1.