ORIGINAL PAPER

Discrete choice experiments to generate utility values for multi-attribute utility instruments: a systematic review of methods

Mina Bahrampour[1] · Joshua Byrnes[1] · Richard Norman[2] · Paul A. Scuffham[1] · Martin Downes[1]

Abstract

Objectives In recent years, discrete choice experiments (DCEs) have become frequently used to generate utility values, but there are a diverse range of approaches to do this. The primary focus of this systematic review is to summarise the methods used for the design and analysis of DCEs when estimating utility values in both generic and condition-specific preference-based measures.

Methods Published literature using DCEs to estimate utility values from preference-based instruments were identified from MEDLINE, Embase, Cochrane Library and CINAHL using PRISMA guidelines. To assess the different DCE methods, standardised information was extracted from the articles including the DCE design method, the number of choice sets, the number of DCE pairs per person, randomisation of questions, analysis method, logical consistency tests and techniques for anchoring utilities. The CREATE checklist was used to assess the quality of the studies.

Results A total of 38 studies with samples from the general population, students and patients were included. Values for health states described using generic multi attribute instruments (MAUIs) (especially the EQ-5D) were the most commonly explored using DCEs. The studies showed considerable methodology and design diversity (number of alternatives, attributes, sample size, choice task presentation and analysis). Despite these differences, the quality of articles reporting the methods used for the DCE was generally high.

Conclusion DCEs are an important approach to measure utility values for both generic and condition-specific instruments. However, a gold standard method cannot yet be recommended.

Keywords Discrete choice experiment · Conjoint analysis · Health state valuation · Preference-based measures · Utility · Systematic review

JEL Classification I10

✉ Mina Bahrampour
mina.bahrampour@griffithuni.edu.au

1 Centre for Applied Health Economics, School of Medicine and Menzies Health Institute Queensland, Griffith University, Nathan, QLD, Australia

2 School of Public Health, Curtin University, Perth, WA, Australia

Introduction

Resources in health-care are scarce [1]. Finding an efficient way to assign these limited resources has been one of the objectives of economic evaluation. Economic evaluations, which are designed to determine the most efficient and cost-effective interventions [2], generate evidence for prioritising the best set of interventions to maximise total health. To do this, a number of different methods are available. These methods compare the interventions and programs through two dimensions, cost and consequences. The consequences can be either monetary or non-monetary values [3].

One popular method used in economic evaluation is cost–utility analysis (CUA) [4] in which the main outcome is defined with respect to quality adjusted life years (QALYs)

[5]. The estimation of QALYs requires utility values for defined health states [6]. Usually, these values, which are also known as preference scores, utilities or weights [7], are anchored on a scale where one is representative of full health and zero represents death [8]. To ensure comparability between different areas of health, health states are often defined using a multi-attribute utility instrument (MAUI). MAUIs can be either generic or disease-specific [9] and the decision on what kind of MAUIs to use to value any particular health state is still debatable.

MAUIs describe health states based on dimensions of quality of life and are associated with a scoring algorithm to estimate the relative value of that health state. These values are then multiplied by the duration in each health state to calculate QALYs [10]. Frequently used preference-based methods can be classified into either cardinal methods or more recently, ordinal methods. Cardinal methods for utility estimation assume preference can be expressed quantitatively; however, ordinal methods, such as ranking models [11] and Discrete Choice Experiments (DCEs) [12–14], do not rely on an individual providing a score for a particular health state; rather they are based on rankings instead [15]. In SG, the participants are asked the risk they are going to take for achieving better health and in TTO participants are asked about the amount of time they are willing to trade off for better health. There are some concerns with using SG and TTO preference based methods [16, 17]. The tasks in SGs and TTOs might be more difficult to administer in comparison with DCE [18]. Valuing health states worse than dead can be problematic with these techniques [19]. These problems have led to the use of ordinal methods such as DCEs [20]; however, these methods have their own limitations [21–23]. A DCE is a preference-based technique that asks individuals to choose between health states over a set of different hypothetical scenarios in order to elicit their preferred health state and the relative weights for various attributes embodied within health states. DCEs for MAUIs are similar to TTO and SG, in the case that all of the models are presented by their attributes to describe health states. However, DCEs are generally easier to administer and they have greater reliability due to reduced measurement error [16, 18, 24]. The derived values from DCEs are based on the preferred health state or health profile elicited by the choices made by respondents [25].

Even though DCEs are predominantly applied to elicit the relative strength of preferences of different attributes within programs or between programs, the use of DCE to value health states for estimating utility weights is increasing [26]. There are key stages to perform a DCE for a valuation study. Before starting to design the DCE, the first step is identifying relevant dimensions of the health outcome measure and assigning levels to these dimensions. The dimensions and their levels will be described when using a classification (descriptive) system. To design the DCE, the next step is applying statistical design theory to draw an independent sample of health states from the full set of health states, for which preferences will be elicited. Then the choice sets will be presented to respondents, and they will be asked to express their preferences by choosing within these choice sets.

There are many different methodological and design considerations when using a DCE to value utilities. The validity of these methods is uncertain. Indeed, the methods to derive DCEs are still evolving, and hence it is important to understand how researchers are developing and using DCEs to help define and inform best practice for future studies. This paper reports a systematic review based on methods to determine how DCEs are conducted to estimate utility values for MAUIs. The main aim of the review was to summarise the methods used to design and analyse DCEs when generating utility weights in both generic and condition-specific preference-based measures.

Methods

This systematic review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) guidelines [27]. The protocol for this review was registered in the international database of prospectively registered systematic reviews in health and social care (PROSPERO), Centre for Reviews and Dissemination, University of York (CRD42017075603).

Search strategy

In order to identify the relevant published literature, multiple electronic databases were included in the search strategy: Medline, EMBASE, Cochrane Library and CINAHL. The selected studies were restricted to English language articles published in peer-reviewed journals. References of the included articles were cross checked to identify other relevant publications. There were no limits on time of publication.

A preliminary scoping search was conducted to identify terminology for the search terms and the type of studies that are likely to be available and the research team discussed and approved the list of key search terms. The main search terms were related to commonly used DCE terminology, quality of life terminology and different multi-attribute instruments; a full description of the search strategy can be found in the supplementary file. Inclusion and exclusion criteria were as follows:

Studies had to meet the following criteria to be included:

- Use of a DCE for preference elicitation; and

- Use of an existing MAUI

Studies were excluded if:

- HRQoL was valued using only non-DCE techniques such as TTO or SG;
- A DCE was used, but did not report the results of the DCE design in their study; or
- Results were based on simulated data

More generally, methodological articles with no data, systematic reviews, working papers, protocols, editorials or letters and abstracts were excluded from the identified articles.

Study selection and quality assessment

MB reviewed article titles and excluded those that clearly did not match the inclusion criteria. MB reviewed the abstracts that were remaining to assess if they met the inclusion/exclusion criteria. After exclusion by title and abstract, MB reviewed the full text of the remaining studies and decided on the final included article. SK assessed all the articles at the abstract stage to ensure consistent selection/rejection. If there were any differences in agreement, adjudication was undertaken by MD. Assessment of the quality of the articles in this review has been done using the CREATE checklist [28]. This checklist was designed to assess the quality of valuation studies of multi-attribute utility-based instruments. The CREATE checklist includes 21 items in seven categories: descriptive system; health states valued; sampling; preference data collection; study sample; modelling; and scoring algorithm. For this study, each item in the checklist was scored either yes or no; however, some items were not applicable in some studies, such as studies that compared models but did not report the algorithms. For each study, the percentage score of reported items was calculated (yes was scored as 1 and no as 0).

Data extraction

Data were extracted from eligible articles using a predefined data extraction sheet (supplementary file). This file included the general information of the studies such as study country, study year; DCE design methods such as the number of choice sets, number of DCE pairs, randomisation of questions, logical consistency tests and techniques for anchoring utilities; it also consisted of questions regarding other preference eliciting methods also conducted, analysis software and statistical models. These categories were chosen as they have been used in previous reviews and have been included in checklists to develop DCEs [29, 30].

Results

Selection process

After initial searches were completed, a total of 3162 studies were identified. After reading of titles and abstracts, 141 articles were identified as potentially meeting the inclusion criteria. In the full article review stage, 38 articles were included in the final evidence synthesis (Fig. 1).

Study descriptive data

From the literature search, a total of 38 articles were used in the systematic review (Fig. 1).

The EQ-5D, a generic MAUI, was the most commonly used instrument ($n = 24$) for valuation of health states through a DCE, of which 17 studies used the more recent EQ-5D-5L, seven studies used the EQ-5D-3L and one study used both [31]. Other generic MAUIs were the SF-6D ($n = 2$) and the CHU-9D ($n = 1$). Condition-specific MAUIs included cancer-specific instruments (the QLU-C10D [23, 32–34], derived from the EORTC QLQ-C30, ($n = 4$), Asthma Quality of life (AQL-5D) [13] ($n = 1$), an influenza-related health-related quality of life measure ($n = 1$), the over active bladder questionnaire (OAB-5D) [13, 17] ($n = 1$), sexual quality of life (SQOL-3D) ($n = 1$), a Glaucoma utility index ($n = 1$), a Diabetes Health Profile-Five Dimension ($n = 1$) and the ICE-CAP Supportive Care Measure [35] ($n = 1$) (Table 1).

The majority of studies recruited participants from a single country: the UK ($n = 12$), Australia ($n = 7$), the Netherlands ($n = 4$), Germany ($n = 2$), Spain ($n = 2$), the USA ($n = 2$), Canada ($n = 1$), Indonesia ($n = 1$) and Japan ($n = 1$). Six studies were carried out in more than one country: Xie et al. [36] included study participants from UK, Canada, Spain, the Netherlands, China, Japan, Korea and Thailand; two studies [33, 34] recruited participants in France and Germany; Krabbe et al. studies [7, 37] recruited participants in Canada, England, The Netherlands and the USA, and Pullenayegum et al. [38] recruited participants in Canada and the UK. Thirty-three of the studies sampled from the general population, stratified by age and gender (supplementary file Table A1.1). Scalone et al. [39] sampled university students in Netherlands. Stolk et al. [16] elicited values from both a general population sample and students. Ratcliffe et al. [40] used adolescents as the study sample, as their questionnaire was designed for young adults. Burr et al. [25], which used a disease specific MAUI, used patients to elicit values. Sample sizes ranged from 102 to 8222 respondents (Supplementary file table A1.2).
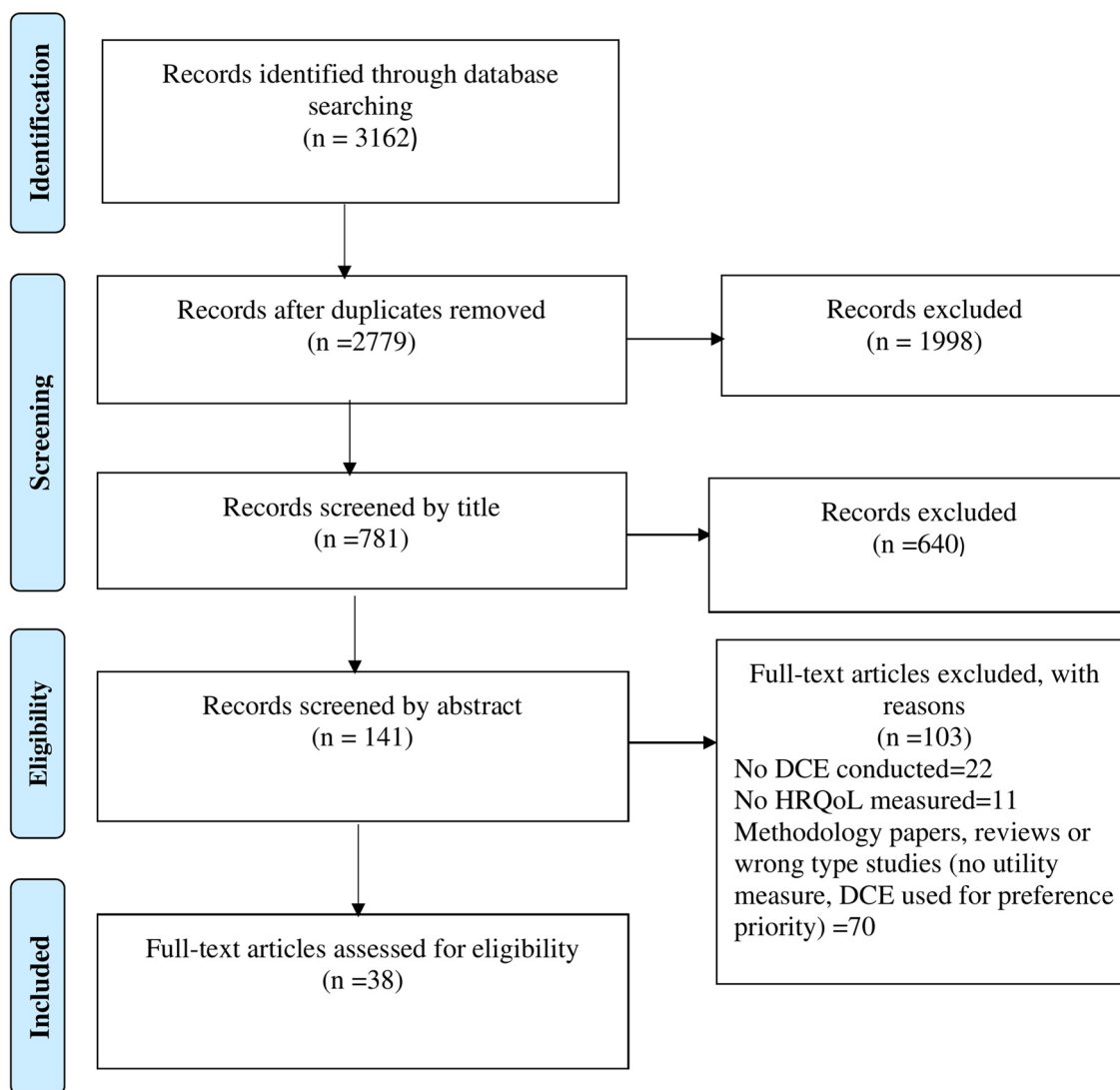
Fig. 1 PRISMA flow diagram summarizing the study selection process

Administration

Commercial providers administered the majority of the questionnaires (Table A1.1). 30 studies used an online panel to gather data, four studies used interviews and follow-up postal questionnaires and two studies used computer-assisted personal interviews (CAPI) and two studies did not mention their data gathering method.

Modelling DCE data

A number of the included studies developed a task beyond a simple comparison of two or more health states. For example, nine studies added the "duration" of being in a given health state to their DCE questionnaire (DCEtto); 12 studies

included a separate TTO experiment of which one also included a willingness to pay (WTP) task [41]; two articles included a ranking task [17, 26], and two studies [16, 42] included a visual analogue scale (VAS) task in addition to a TTO and DCE. Two studies [37, 43] included a VAS and seven studies also included a best–worst-scaling (BWS) task in addition to a DCE, the BWS type has not been reported (Supplementary file table A1.2). Seven studies used dead in their survey as a third option in the choice set.

Attributes and levels

The number of possible health states covered by the instruments ranged from 64 to 2 million. The number of choice sets included in the studies ranged from 24 to 1620. The

Table 1 Background information of the studies

	Complete name	N=38
MAUIs:	Complete name	
EQ-5D-5L		
EQ-5D-3L	European quality of life 5 dimensions 5 level	18
QLQ-C30 (QLU-C10D)	European quality of life 5 dimensions 3 level	8
SF-6D	Quality of life questionnaire	4
CHU-9D	Short-form questionnaire-6 dimensions	2
AQL-5D	Child health utility-9 Dimensions	1
OAB-5D + AQL-5D	Asthma quality of life-5 Dimensions	1
SQOL-3D	over active bladder quality of life-5 dimensions	1
ICECAP SCM	Sexual quality of life 3 dimension	1
Diabetes health profile-Five dimension	ICEpop CAPability measure Supportive Care Measure	1
Influenza-related health-related quality	Diabetes health profile	1
	Influenza-related health-related quality	1
Number of choices	< 8	8
	08-Dec	15
	13–16	12
	16 <	3

One extra MAUI as one study used both EQ-5D-3L and EQ-5D-5L

DCE choice tasks per respondent varied between 6 and 32 choices (Supplementary file table A1.1).

Design

The way that DCE designs were reported in the studies varied; the reported methods were full factorial ($n=1$), fractional factorial ($n=3$), efficient design (d-efficient and two-way interaction; $n=11$), Bayesian methods ($n=12$), d-optimal ($n=5$) and orthogonal methods ($n=3$). In these studies, 14 studies reported using non-zero priors: 12 studies got the priors from previous studies that used the same questionnaire, one study mentioned the use of a pilot study to get priors and one study did not mention the method. Five studies used zero priors. Some studies reported using design generator software such as Ngene ($n=6$) or SAS® ($n=7$) to design their DCE (Supplementary file table A1.1).

Statistical analyses and software

The majority of the studies ($n=23$) used conditional logistic regression models. Two [16, 17] of these studies used rank-ordered logit in addition to conditional logit. Two studies used multinomial logit [35, 44]. Five articles used random-effects probit models and two studies applied multinomial probit. Multi-level mixed effects logit, latent class method, generalized estimation equation (GEE) and conditional probit were each used in one study. One study mentioned a new model, the Zermelo–Bradley–Terry (ZBT) Model [45], to value health states.

Stata was the most commonly used software to analyse the data ($n=18$). SAS, R, Matlab and WinBUGs were other software used to analyse the DCE data (Supplementary file table A1.1).

Table 2 Consistency methods

Number of studies	Method of consistency	Theory stated behind method
9	Dominant choices	Using dominant choices among choice sets where one health state is better than all other states or is the worst state among all, or using duration as a dominant option, it means checking consistency with respondents who always chose the health profiles with the longest duration
6	Best and worst health state	Coefficients logically consistent if the worse health states has the least value and increases logically as the health state gets better (best health state has greatest value) [44]
7	Question ordering	Checking the answers to see whether the participant always selects the alternative that is on the left-hand side or on the right-hand side

Randomisation and consistency

Randomisation was usually done in a few ways; by randomly allocating respondents to a block of choice tasks ($n = 19$), which was used in the majority of studies, or randomisation was achieved based on the order of choice sets or the question position among participants ($n = 13$). In some studies, respondents were randomly assigned to choice-sets ($n = 5$) (Supplementary file table A1.2).

Methods for assessing consistency included consistency in respondents and consistency in coefficients: one study reported that it had consistency in responses [25] (Table 2) and five studies discussed consistency in coefficients but did not mention which method was used to check for this [26]. In 12 studies, the authors did not report any consistency test.

Anchoring

All four methods for anchoring described by Rowen et al. [13] were observed across the included studies as well as three additional methods not described by Rowen. The first method is to anchor using the dummy coefficient for dead, an approach used in only one study in this review [16]. The second method is to anchor the worst state using TTO, again used in a single study [46]. The third method is to map the DCE on to TTO, which was used in four papers [26, 47–49]. The last approach is a hybrid of methods one and two used in two papers [42, 46]. Two studies [13, 50] reported all anchoring methods in their study. The Bansback et al. study [24] used the value of coefficients in the conditional logit for anchoring when an attribute for duration is included (DCEtto). The same method as Bansback's study was used in 13 of the included studies. Two articles just mentioned anchoring between 0 and 1 for dead and full health, respectively, and five studies anchored between − 1 and 1, best and worst health state. One study [40] rescaled their utilities using an algorithm developed based on SG method. Six studies did not report any anchoring method (Supplementary file table A1.2).

Quality assessment using the CREATE checklist

In general, the majority of the studies scored well against the CREATE checklist [28] with an average score of 86.5 per cent (ranging from 68.4 to 100 per cent). The item "sample size/power calculations are stated and rationalized" was the least observed item in the studies. The average score for subgroup health states values was the highest (average of 99 percent).

Discussion

A systematic review by de-Bekker Grob et al. [30] showed that there were only two published studies that had used a DCE to value utility within the QALY framework between 2000 and 2010, and the review by Clark et al. [20] showed that between 2009 and 2012, four studies had used DCEs to value utility within the QALY framework. In recent years, the use of DCEs to estimate utility values has increased. This review addressed the methods when designing and analysing DCEs to value health states and outlined a variety of methods when developing a DCE for MAUIs. There is no apparent standardised method for design or analysis; however, there were similarities in the studies reviewed. For instance, the majority of studies used an online panel and a sample of general population. There are divergent opinions on whose preference should be considered in valuation studies. On the one hand, some researchers believe that the general population should be the respondents as they are representative of the insured population (either via social or private insurance schemes) and, therefore, collectively they should dictate values. On the other hand, some researchers suggest that a person that has the disease can better value the health state and their opinion is more realistic [51]. Maybe that it is not one or the other; rather, both are relevant but should be used for different purposes. This debate is ongoing for valuation studies of condition specific MAUIs.

Variation in methodology and design might be considered useful by some researchers. It might promote learning about the impact of methodological choices and help guide decisions about future protocols or elements of new methodology that may be applicable to a wide range of DCE studies. However, while the variation in methods remains, it is important for decision makers to understand the features and limitations of utility value sets in order to rely on these for decision making. Researchers need to continue to strive for value sets that are valid and reflect true preferences whilst continuing to educate and be guided by decision makers and the public.

Design

Deciding which experimental design to use is an important step and has a key role when performing a DCE. Design defines the coefficients that can be estimated in analysis and can ensure the estimate is as precise as possible for a given sample size. As identified in our review a range of methods have been used by studies in this area. Classical designs such as factorials use orthogonal designs, which apply orthogonal arrays employing different sources, in which the attributes are statistically independent and attribute levels

occur equally [52]. Using a full factorial design requires presentation of all combination of attributes; the Prosser et al. [41] study used a full factorial design but excluded the dominated choices in the design. The authors state that decreasing the number of health states makes the design less efficient. Usually, there are numerous combinations, making presentation of all potential choices unfeasible; therefore, a fractional factorial design is generally more pragmatic to elicit respondent preferences.

In practice, designs which are near-orthogonal are usually also applicable, as the orthogonal designs usually exist for particular combinations. A near-orthogonal design, such as D-efficiency, is often used to measure the relative goodness of a fractional factorial design [53]. Results of this review illustrate the trend towards using D-efficient designs in this area, which reflects a general trend observed in previous DCE methodology reviews [20, 30]. The use of Bayesian efficient designs has also been increasing [47, 50]; this might be due to the use of more random priors parameters instead of fixed priors parameters in the designs to obtain more stable designs. And in the Bayesian efficient designs the priors are considered to be random parameters [54, 55].

In order to design the DCE the researchers can decide whether to include zero priors or non-zero prior information (expected values); there is controversy over using priors in the design. Additional work should be undertaken to understand the changes in the model when zero or non-zero prior information is included, which may then help to inform future studies.

Presenting the DCE choice set and training the participants about choosing health states is usually done by a brief introduction about the DCE at the beginning of the DCEs. Adding a question to the checklist about the method of the presentation of the task within the DCE studies might be beneficial; this can be an area for future work in DCE of MAUI.

Consistency

When designing the DCE, researchers need to consider strategies to identify and prevent high choice inconsistency. Regardless of how choices are presented to respondents, there should be logical consistency across questions. Logical consistency has a key role and is the main principle of choice theory in modern economics. Consistency describes logical ordering of health states in the method [56, 57], and it can be checked by the way the choices are presented to the respondents, for instance, changing the position of the alternatives or order of choice sets [22, 33] or checking if there are respondents who continuously chose the health profiles with the longest duration when duration is involved [58]. However, it is difficult to claim that for a participant who always chooses the health state with the longest duration,

this might not actually be their 'true' preference. Estimated weights should increase when moving form worst health to best health state which can be easily assessed. 'True' preference means that there is no bias in choosing the health state and it is the participants' actual preferred health state.

One of the main reasons for individual's responses to get deleted from DCE data sets is that the respondents are identified as having lexicographic preferences, meaning that the individual makes decisions by ranking attributes and consistently choosing the attributes in the order of their highest priorities; for example, the participant always chooses the option with the highest level of duration [59]. However, based on the aim of DCE it may not be preferable for researchers to delete these respondents [60]. For instance, some studies [39, 59] accomplished the analysis with the complete sample stating that deleting the individuals with dominant preferences will not change the estimated parameters, while the Bansback et al. [24] study deleted observations with lexicographic preferences, explaining that those individuals failed to understand the task. Although future studies could present the results with and without excluding participants and provide justification for selection of an approach, it remains unclear as to which set of results is more indicative of the true underlying preferences based solely on the data collected from the choice set. A valuable avenue for future research may be exploring the best method to check choice stability and consistency in choice studies and the extent to which apparently unstable or inconsistent choices can be identified and subsequently handled within the analysis.

Choice tasks

The de-Bekker Grob's [30] study discussed that the mean number of choices across all health-based DCES is 14, while Bridges et al. [29] suggest that between 8 to 16 choices is a reasonable number in practice. Two studies stated that models using 10 choice pairs are more consistent than 15 pairs [48, 59]. We found that the majority of the studies had between 8 to 12 choice tasks although the maximum number of choice sets was 32.

Having a large or a small number of choice sets has its own advantages and disadvantages. The advantage of having many choices is that it facilitates estimation of mixed logit models; however, if the number of choices is high, the participants might become fatigued and may not give complete attention to the questions and thereby decrease the completion rate [32]. For instance, the Burr et al. [25] study used 32 choices that caused the respondents to take a median of 75 min to answer the DCE questionnaire, and as a result they could not include the individuals with severe case of the disease. Similarly, where the number of choices

is high, participants may revert to heuristic choices, such as lexicographic preferences. In addition, including a low number may result in a lack of information needed for reliable analysis while higher tasks per person could provide one avenue to estimating individual preference algorithms.

Anchoring

Health state values can be anchored on a 0–1 scale where one presents the best health and zero represents dead [61]; however, using different descriptive systems, health states worse than dead are possible, which have a negative utility assigned to them [62]. Studies [26, 40] that anchored health states using the best and the worst health states, where health states are scaled from –1 to + 1, need to re-anchor the health states to estimate QALYs. Depending on the design of each study the anchoring might be different, yet the utilities need to be in the interval of zero (dead) to one (full health) scale if the results want to be incorporated into QALY calculations [63].

In DCEs, different models of anchoring health states to provide utility values have been applied. Usually in DCE studies the health states are not compared to death directly and setting dead as zero is a less-direct approach compared to TTO and SG methods. The anchoring for DCE is done mostly by using a combination of DCE and a different but related preference-based method such as TTO, or including duration as an option in the design; however, if a DCE is being applied to overcome the limitations of other preference-based methods, using this method alone to anchor utility values is its own analytical challenges [64].

Based on the studies, we assume that anchoring the DCEs with duration (DCEtto) may be less complicated compared to other methods; however, there is still a debate regarding how to best present the duration attribute in DCE studies in the DCEtto approach [44]. Further research is required to understand how duration may affect the selection of an item without leading to any bias and more broadly, to determine the most appropriate way to anchor DCE data.

Analysis

The use of probit analyses has decreased from 1990 to 2012 [20] while the use of logit analyses has risen in these years. The results of our study agree with this finding; recent papers mostly used logit analyses rather than probit analyses. There was no unique method and approach in all studies. However, most of the studies used conditional logit as their analysis method. The study by Hauber et al. [65] provides a useful guide to explain the statistical methods to derive preference weights from DCEs and choosing the best analysis method based on the aim and the data of their study.

Conclusions

Reflective of the results of this review, there is no gold standard DCE method. Given that there are different methods used to measure health-state values, future studies are required to compare various methods in order to indicate the best method based on the aspects of the value set. As the theoretical literature on DCEs evolves, so too will the methods applied to analyse DCE data. It is evident, however, that DCEs will continue as an approach to elicit preferences for valuing health states.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Kluge, E.-H.W.: Resource allocation in healthcare: implications of models of medicine as a profession. Medscape Gen. Med. **9**(1), 57–57 (2007)
2. Griffin, S., Claxton, K., Sculpher, M.: Decision analysis for resource allocation in health care. J. Health Serv. Res. Policy **13**(3_suppl), 23–30 (2008). https://doi.org/10.1258/jhsrp.2008.008017
3. Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., Torrance, G.W.: Methods for the economic evaluation of health care programmes. Oxford University Press, UK (2015)
4. Kularatna, S., Whitty, J.A., Johnson, N.W., Jayasinghe, R., Scuffham, P.A.: A comparison of health state utility values associated with oral potentially malignant disorders and oral cancer in Sri Lanka assessed using the EQ-5D-3 L and the EORTC-8D. Health Qual. Life Outcomes **14**, 101 (2016). https://doi.org/10.1186/s12955-016-0502-y
5. Whitehead, S.J., Ali, S.: Health outcomes in economic evaluation: the QALY and utilities. Br. Med. Bull. **96**(1), 5–21 (2010). https://doi.org/10.1093/bmb/ldq033
6. Gu, Y., Norman, R., Viney, R.: Estimating health state utility values from discrete choice experiments—a qaly space model approach. Health Econ. **23**(9), 1098–1114 (2014). https://doi.org/10.1002/hec.3066
7. Krabbe, P.F., Devlin, N.J., Stolk, E.A., Shah, K.K., Oppe, M., van Hout, B., Quik, E.H., Pickard, A.S., Xie, F.: Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. Med. Care **52**(11), 935–943 (2014). https://doi.org/10.1097/MLR.0000000000000178

8. Brazier, J., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. J. Health Econ. **21**(2), 271–292 (2002)

9. Bjornson, K., McLaughlin, J.: The measurement of health-related quality of life (HRQL) in children with cerebral palsy. Eur. J. Neurol. **8**(s5), 183–193 (2001)

10. Scuffham, P.A., Whitty, J.A., Mitchell, A., Viney, R.: The use of QALY weights for QALY calculations. PharmacoEconomics **26**(4), 297–310 (2008)

11. McCabe, C., Brazier, J., Gilks, P., Tsuchiya, A., Roberts, J., O'Hagan, A., Stevens, K.: Using rank data to estimate health state utility models. J. Health Econ. **25**(3), 418–431 (2006). https://doi.org/10.1016/j.jhealeco.2005.07.008

12. Ryan, M., Netten, A., Skåtun, D., Smith, P.: Using discrete choice experiments to estimate a preference-based measure of outcome—an application to social care for older people. J. Health Econ. **25**(5), 927–944 (2006). https://doi.org/10.1016/j.jhealeco.2006.01.001

13. 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 **35**(3), 328–340 (2015). https://doi.org/10.1177/0272989X14559542

14. Lancsar, E., Wildman, J., Donaldson, C., Ryan, M., Baker, R.: Deriving distributional weights for QALYs through discrete choice experiments. J. Health Econ. **30**(2), 466–478 (2011). https://doi.org/10.1016/j.jhealeco.2011.01.003

15. Ali, S., Ronaldson, S.: Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. Br. Med. Bull. **103**(1), 21–44 (2012)

16. Stolk, E.A., Oppe, M., Scalone, L., Krabbe, P.F.: Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value Health **13**(8), 1005–1013 (2010). https://doi.org/10.1111/j.1524-4733.2010.00783.x

17. Brazier, J., Rowen, D., Yang, Y., Tsuchiya, A.: Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. Eur. J. Health Econ. **13**(5), 575–587 (2012). https://doi.org/10.1007/s10198-011-0352-9

18. Ludwig, K., von der Schulenburg, J.G., Greiner, W.: Valuation of the EQ-5D-5L with composite time trade-off for the German population—an exploratory study. Health Qual. Life Outcomes **15**(1), 39 (2017). https://doi.org/10.1186/s12955-017-0617-9

19. Craig, B.M., Busschbach, J.J.V.: Towards a more universal approach in health valuation. Health Econ. (2011). https://doi.org/10.1002/hec.1650

20. Clark, M.D., Determann, D., Petrou, S., Moro, D., de Bekker-Grob, E.W.: Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics **32**(9), 883–902 (2014)

21. Mulhern, B., Norman, R., Street, D.J., Viney, R.: One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. PharmacoEconomics **37**(1), 29–43 (2019). https://doi.org/10.1007/s40273-018-0714-6

22. Mulhern, B., Norman, R., Lorgelly, P., Lancsar, E., Ratcliffe, J., Brazier, J., Viney, R.: Is dimension order important when valuing health states using discrete choice experiments including duration? PharmacoEconomics (2016). https://doi.org/10.1007/s40273-016-0475-z

23. King, M.T., Viney, R., Simon Pickard, A., Rowen, D., Aaronson, N.K., Brazier, J.E., Cella, D., Costa, D.S.J., Fayers, P.M., Kemmler, G., McTaggart-Cowen, H., Mercieca-Bebber, R., Peacock, S., Street, D.J., Young, T.A., Norman, R., On behalf of the, M.C.: 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 **36**(2), 225–238 (2018). https://doi.org/10.1007/s40273-017-0582-5

24. Bansback, N., Brazier, J., Tsuchiya, A., Anis, A.: Using a discrete choice experiment to estimate health state utility values. J. Health Econ. **31**(1), 306–318 (2012). https://doi.org/10.1016/j.jhealeco.2011.11.004

25. Burr, J.M., Kilonzo, M., Vale, L., Ryan, M.: Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optom. Vis. Sci. **84**(8), 797–808 (2007). https://doi.org/10.1097/OPX.0b013e3181339f30

26. Ratcliffe, J., Brazier, J., Tsuchiya, A., Symonds, T., Brown, M.: Using DCE and ranking data to estimate cardinal values for health states for deriving a preference-based single index from the sexual quality of life questionnaire. Health Econ. **18**(11), 1261–1276 (2009)

27. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. **151**(4), 264–269 (2009)

28. Xie, F., Pickard, A.S., Krabbe, P.F., Revicki, D., Viney, R., Devlin, N., Feeny, D.: A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). PharmacoEconomics **33**(8), 867–877 (2015). https://doi.org/10.1007/s40273-015-0292-9

29. Bridges, J.F., Hauber, A.B., Marshall, D., Lloyd, A., Prosser, L.A., Regier, D.A., Johnson, F.R., Mauskopf, J.: Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health **14**(4), 403–413 (2011). https://doi.org/10.1016/j.jval.2010.11.013

30. de Bekker-Grob, E.W., Ryan, M., Gerard, K.: Discrete choice experiments in health economics: a review of the literature. Health Econ. **21**(2), 145–172 (2012)

31. Selivanova, A., Buskens, E., Krabbe, P.F.M.: Head-to-head comparison of EQ-5D-3L and EQ-5D-5L health values. PharmacoEconomics **36**(6), 715–725 (2018). https://doi.org/10.1007/s40273-018-0647-0

32. Norman, R., Viney, R., Aaronson, N., Brazier, J., Cella, D., Costa, D., Fayers, P., Kemmler, G., Peacock, S., Pickard, A., Rowen, D., Street, D., Velikova, G., Young, T., King, M., Aaronson, N.K., Brazier, J.E., Costa, D.S.J., Fayers, P.M., Pickard, A.S.: Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Qual. Life Res. **25**(3), 637–649 (2016). https://doi.org/10.1007/s11136-015-1115-3

33. Norman, R., Kemmler, G., Viney, R., Pickard, A.S., Gamper, E., Holzner, B., Nerich, V., King, M.: Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. Value Health **19**(8), 1033–1038 (2016). https://doi.org/10.1016/j.jval.2016.07.003

34. Gamper, E.-M., Holzner, B., King, M.T., Norman, R., Viney, R., Nerich, V., Kemmler, G.: Test-retest reliability of discrete choice experiment for valuations of QLU-C10D health states. Value Health **21**(8), 958–966 (2018). https://doi.org/10.1016/j.jval.2017.11.012

35. Huynh, E., Coast, J., Rose, J., Kinghorn, P., Flynn, T.: Values for the ICECAP-Supportive Care Measure (ICECAP-SCM) for use in economic evaluation at end of life. Soc. Sci. Med. **189**, 114–128 (2017)

36. Xie, F., Pullenayegum, E., Pickard, A.S., Ramos Goñi, J.M., Jo, M.W., Igarashi, A., Jo, M.-W.: Transforming latent utilities to health utilities east does not meet west. Health Econ. **26**(12), 1524–1533 (2017). https://doi.org/10.1002/hec.3444

37. Krabbe, P.F.M., Stolk, E.A., Devlin, N.J., Xie, F., Quik, E.H., Pickard, A.S.: Head-to-head comparison of health-state values derived by a probabilistic choice model and scores on a visual analogue scale. Eur. J. Health Econ. **18**(8), 967–977 (2017)

38. Pullenayegum, E., Xie, F.: Scoring the 5-level EQ-5D: can latent utilities derived from a discrete choice model be transformed to health utilities derived from time tradeoff tasks? Med. Decis. Making 33(4), 567–578 (2013). https://doi.org/10.1177/0272989X13475718

39. Scalone, L., Stalmeier, P.F., Milani, S., Krabbe, P.F.: Values for health states with different life durations. Eur. J. Health Econ. 16(9), 917–925 (2015). https://doi.org/10.1007/s10198-014-0634-0

40. Ratcliffe, J., Flynn, T., Terlich, F., Stevens, K., Brazier, J., Sawyer, M.: Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. PharmacoEconomics 30(8), 713–727 (2012). https://doi.org/10.2165/11597900-000000000-00000

41. Prosser, L.A., Payne, K., Rusinak, D., Shi, P., Messonnier, M.: Using a discrete choice experiment to elicit time trade-off and willingness-to-pay amounts for influenza health-related quality of life at different ages. PharmacoEconomics 31(4), 305–315 (2013). https://doi.org/10.1007/s40273-013-0029-6

42. Purba, F., Hunfeld, J., Iskandarsyah, A., Fitriana, T., Sadarjoen, S., Ramos-Goñi, J., Passchier, J., Busschbach, J., Purba, F.D., Hunfeld, J.A.M., Fitriana, T.S., Sadarjoen, S.S., Ramos-Goñi, J.M., Busschbach, J.J.V.: The Indonesian EQ-5D-5L Value Set. PharmacoEconomics 35(11), 1153–1165 (2017). https://doi.org/10.1007/s40273-017-0538-9

43. Eckert, J., Lichters, M., Piedmont, S., Vogt, B., Robra, B.P.: Case vignettes based on EQ-5D to elicit stated preferences for health services utilization from the insurees' perspective. BMC Health Serv. Res. 15, 481 (2015). https://doi.org/10.1186/s12913-015-1143-2

44. Jonker, M.F., Donkers, B., de Bekker-Grob, E.W., Stolk, E.A.: Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value Health 21(8), 993–1001 (2018). https://doi.org/10.1016/j.jval.2018.01.016

45. Craig, B.M., Rand, K.: Choice defines QALYs: a US valuation of the EQ-5D-5L. Med. Care 56(6), 529–536 (2018). https://doi.org/10.1097/MLR.0000000000000912

46. Ramos-Goni, J.M., Rivero-Arias, O., Errea, M., Stolk, E.A., Herdman, M., Cabases, J.M.: Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L heath states. Eur. J. Health Econ. 14(Suppl 1), S33–42 (2013). https://doi.org/10.1007/s10198-013-0511-2

47. Ramos-Goñi, J.M., Pinto-Prades, J.L., Oppe, M., Cabasés, J.M., Serrano-Aguilar, P., Rivero-Arias, O.: Valuation and modeling of EQ-5D-5L health states using a hybrid approach. Med. Care 55(7), e51–e58 (2017)

48. Bansback, N., Hole, A.R., Mulhern, B., Tsuchiya, A.: Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Soc. Sci. Med. 114, 38–48 (2014). https://doi.org/10.1016/j.socscimed.2014.05.026

49. Mulhern, B., Shah, K., Janssen, M.F., Longworth, L., Ibbotson, R.: Valuing health using time trade-off and discrete choice experiment methods: does dimension order impact on health state values? Value Health 19(2), 210–217 (2016). https://doi.org/10.1016/j.jval.2015.11.005

50. Shiroiwa, T., Ikeda, S., Noto, S., Igarashi, A., Fukuda, T., Saito, S., Shimozuma, K.: Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. Value Health 19(5), 648–654 (2016). https://doi.org/10.1016/j.jval.2016.03.1834

51. Versteegh, M.M., Brouwer, W.B.F.: Patient and general public preferences for health states: a call to reconsider current guidelines. Soc. Sci. Med. 165, 66–74 (2016). https://doi.org/10.1016/j.socscimed.2016.07.043

52. Huber, J., Zwerina, K.: The importance of utility balance in efficient choice designs. J. Market. Res. 307–317 (1996).

53. Vanniyasingam, T., Cunningham, C.E., Foster, G., Thabane, L.: Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. BMJ Open 6(7), e011985 (2016). https://doi.org/10.1136/bmjopen-2016-011985.

54. Hensher, D.A., Rose, J.M., Greene, W.H.: Applied choice analysis, 2nd edn. Cambridge University Press, UK (2015)

55. Kitamura, R., Yoshii, T., Yamamoto, T.: Expanding sphere of travel behaviour research: selected papers from the 11th international conference on travel behaviour research. Emerald Group Publishing, (2009)

56. De Martino, B., Harrison, N.A., Knafo, S., Bird, G., Dolan, R.J.: Explaining enhanced logical consistency during decision making in autism. J. Neurosci. 28(42), 10746–10750 (2008)

57. Arrow, K.J.: Risk perception in psychology and economics. Econ. Inq. 20(1), 1–9 (1982)

58. Norman, R., Cronin, P., Viney, R.: A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl. Health Econ. Health Policy 11(3), 287–298 (2013). https://doi.org/10.1007/s40258-013-0035-z

59. Mulhern, B., Bansback, N., Hole, A.R., Tsuchiya, A.: Using discrete choice experiments with duration to model EQ-5D-5L health state preferences. Med. Decis. Making 37(3), 285–297 (2017). https://doi.org/10.1177/0272989X16670616

60. Lancsar, E., Louviere, J.: Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ. 15(8), 797–811 (2006). https://doi.org/10.1002/hec.1104

61. Hakim, Z.: Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. Health Econ. 8(2), 103–116 (1999)

62. Lamers, L.M.: The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med. Care 45(3), 238–244 (2007)

63. Robinson, A., Spencer, A., Moffatt, P.: A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. Med. Decis. Making 35(3), 341–350 (2015)

64. Attema, A.E., Edelaar-Peeters, Y., Versteegh, M.M., Stolk, E.A.: Time trade-off: one methodology, different methods. Eur. J. Health Econ. 14(Suppl 1), 53–64 (2013). https://doi.org/10.1007/s10198-013-0508-x

65. Hauber, A.B., González, J.M., Groothuis-Oudshoorn, C.G.M., Prior, T., Marshall, D.A., Cunningham, C., Ijzerman, M.J., Bridges, J.F.P.: Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. Value Health 19(4), 300–315 (2016). https://doi.org/10.1016/j.jval.2016.04.004