Health and Quality of Life Outcomes

RESEARCH Open Access

Optimal DCE design for modelling nonlinear time preferences in EQ-5D-5L valuation studies: exploration of data from Denmark and Peru

Alice Yu[1], Deborah Street[1], Marcel Jonker[2], Sterre Bour[2], Brendan Mulhern[1*], Federico Augustovski[3], Cathrine Jensen[4], Claire Gudex[5], Morten Jensen[6], Romina Tejada[7], Richard Norman[8], Rosalie Viney[1] and Elly Stolk[2,9]

Abstract

Background Discrete choice experiment (DCE) methods are an increasingly popular valuation method, particularly for the EQ-5D-5L. While EQ-5D-5L value sets developed using DCE have traditionally assumed linear time preferences, this assumption has been challenged. This has led to the development of DCE modelling methods that allow for nonlinear time preferences. The aim of this study was to explore the impact of a model that accounts for nonlinear time preferences with DCE choice set formats and design construction methods for EQ-5D-5L value sets.

Methods This study used a four-arm (2 × 2) between-subjects design to investigate the impact of two commonly used DCE choice set formats (i.e. a third option of either immediate death or full health) and two commonly used DCE design construction methods (i.e. generator-developed and efficient designs) on EQ-5D-5L value sets. Mixed logit models that used exponential discounting to account for nonlinear time preferences were estimated in OpenBUGS. This was tested in a sample of respondents from Peru ($n = 942$) and Denmark ($n = 988$).

Results Across all arms and for both countries, discounting was found to be present when modelling explicitly for nonlinear time preferences. Although estimated discount rates varied widely from 1 to 117%, both type of choice set format and type of design construction method influenced the utilities for more severe health states. Choice sets with full health tended to produce a wider range of utility weights, while choice sets with immediate death tended to produce higher estimated discount rates. Generator-developed designs tended to produce the highest and lowest utility weights for health states compared to the efficient designs.

Conclusions This study provides a comparison of DCE choice set format and design construction method when nonlinear time preferences were explicitly modelled. Limitations to this study are discussed including data quality issues with the Peruvian dataset and small sample sizes. Further investigation is needed to confirm the suitability of models that account for nonlinear time preferences in EQ-5D-5L valuation studies.

*Correspondence:
Brendan Mulhern
Brendan.Mulhern@uts.edu.au

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

Keywords Quality of life, EQ-5D-5L, Discrete choice experiment, Nonlinear modelling, Efficient designs, Generator developed designs

Introduction

Time trade off (TTO) has been the key method included in the EQ-5D-5L valuation protocol (EQ-VT) used to estimate value sets for EQ-5D-5L health states [1], but TTO methods are often labour-intensive as they are undertaken as face-to-face interviews. This has led to growing interest in alternative valuation methods, such as discrete choice experiments (DCEs). Like the TTO approach, the DCE method can derive health state values from trade-offs between time and quality of life [2]. DCEs can be administered online and do not require individual interviews. DCEs have gained recognition as a valuation method with an increasing number of valuation studies using DCE methods [3, 4].

One of the challenges of using DCE as a valuation method is how to incorporate time preferences. DCE methods used for EQ-5D value sets have commonly assumed linear time preferences, i.e. respondents are assumed to value time similarly regardless of whether it is closer to the present or further in the future [5–7]. It has also generally been assumed that respondents exhibit constant proportional time trade-off [8, 9]. That is, the trade-off between duration and quality of life gain is a fixed proportion and is independent of the quantity of life years presented to respondents. However, there is evidence that constant proportionality does not hold and that models which include time discounting (i.e. events in the future to be valued less than events closer to the present) provide a better reflection of respondent health preferences [10].

This has led to the introduction of DCE methods that explicitly model potential nonlinear preferences for duration or time [11], i.e. DCE data is modelled assuming time is discounted such that time further away in the future is considered less valuable than time closer to the present. Jonker et al. [11] found that when discount rates are explicitly estimated in the DCE modelling process, there was evidence of nonlinear time preferences in the context of estimating a Dutch value set for the SF-6D. More generally, Jonker and Bliemer [12] demonstrate that for several instruments including the EQ-5D-5L, DCE design specifications that account for nonlinear time preferences tend to have better design efficiencies (based on Bayesian D-error criterion) compared to DCE specifications that assume a linear utility specification.

This favours a movement towards modelling explicitly for nonlinear time preference in the valuation of EQ-5D instruments. However, an issue that remains to be addressed is to understand the impact of DCE design methods when modelling nonlinear time preference. There are currently two commonly used DCE design construction methods—generator-developed [13] and efficient designs [11, 14]. Generator development can be used to construct designs that are D-optimal under the null hypothesis of utility neutrality, that is, all entries in beta are 0. Efficient designs are designs in assumptions are made about the prior probability distribution for beta. An algorithm with an appropriate objective function, such as the determinant of the inverse of the information matrix, is used to evaluate a large number of randomly drawn designs. Of all designs tested the one if kept is the one that best approximates the expected value of the objective function. In practice, due to the random nature of the draws and inability to test all possible designs, it is not possible to know if the final design is actually optimal [15].

There are also two commonly used formats for DCE choice sets—including a third option of either immediate death or full health. In the literature, generator-developed designs have conventionally been paired with DCE choice sets that use immediate death, while efficient designs are commonly paired with DCE choice sets that use the full health option [3, 4, 14, 16]. The development of these two DCE choice set formats are briefly reviewed below.

DCE with duration using immediate death

The DCE valuation method that was developed in 2012 for anchoring values on the QALY scale included duration as an attribute in each option of the choice set [17]. This will be referred as DCE with duration, although it has also been referred to as DCE_{TTO} in the literature. For valuation of the EQ-5D using DCE with duration, each option thus consists of a health state described using the five dimensions of the EQ-5D and an additional attribute for duration, giving six attributes in total. Each respondent's utility is assumed to be defined by the product of the utility of the specified health state and the duration in each option. When duration is zero, the utility is independent of the health state—this is known as the zero-condition assumption [8, 18].

However, a concern about the DCE with duration method (in which two or more health state duration combinations are presented to respondents) is that respondents have not been explicitly asked to consider whether a health state is better or worse than being dead. Instead, the position of being dead on the utility scale is inferred through the modelling process [16].

Viney et al. [16] added to the DCE with duration method by introducing immediate death as an option in

each choice set to obtain stronger information about the position of dead. This approach is well established and has been used to obtain value sets for the EQ-5D-5L [16, 19, 20], SF-6D [21, 22], EORTC-QLUC10D [23, 24] and the FACT-8D [25, 26]. This approach typically presents two health state duration combinations (A and B), and a third option (C) described as immediate death. Options A and B are represented by an EQ-5D-5L health state experienced for a specified duration (same for the two health states or may differ). Option C is specified as death (no duration and no health state). Typically, respondents are asked to choose both the best and the worst option of the three in order to obtain a complete ranking of health states from those considered worse than death to full health [27]. This will be called the *immediate death* approach in this study.

DCE with duration using full health

The *full health* approach to DCE with duration was introduced in 2017 [28] in response to concerns about the complexity of tasks given to respondents and whether respondents behave according to the assumptions of the QALY model. DCE with duration responses are modelled under the assumption that utility of a health state is a product of length of life and quality of life, i.e. that respondents treat health and duration multiplicatively. However, in an unconstrained task, there is evidence that most respondents seem to behave in ways that are not consistent with this assumption, and this can bias the results [29]. The *full health* approach implements various constraints on the task to mitigate this problem. Firstly, options A and B always have the same duration, and option C always refers to full health, for a shorter duration compared to options A and B. This approach uses a 'matched pairwise choice' format, meaning that respondents first indicate their preferred option from A and B, and then choose between option B and option C. By allowing the time spent in full health in option C to approach 0 (3 months is the smallest amount of time offered), 'dead' is approximated, but it is not the same as including the immediate death option in the *immediate death* approach.

How to incorporate nonlinear time preferences in EQ-5D-5L valuation studies

The *immediate death* and *full health* approaches represent different ways the DCE with duration can be implemented in EQ-5D valuation studies. Researchers can choose the DCE choice set format, design construction method, implementation, and analysis approach. These choices can affect utility weights [30], as can the way in which 'dead' is treated in the modelling process [20].

Lim, et al. [14] used efficient DCE designs to investigate the sensitivity of the DCE with duration approach to how well the DCE design construction method covers the severity range of health states. They found that the *full health* approach to DCE with duration was sensitive to the DCE design construction method and produced a skewed health state selection, introducing bias to the results. The *immediate death* approach was less impacted by the severity range in the DCE design construction method. The DCE with duration approach thus appeared to be sensitive to the DCE design construction method, although their study used efficient designs only and did not test generator-developed designs.

Roudijk et al. [2] compared the impact of using linear versus nonlinear modelling on the valuation of the EQ-5D-5L. In addition, they used DCE with duration choice sets with full health and immediate death as a 3rd option using the matched pairwise format, although only 3/18 tasks used health state C as immediate death. It was found that the linear and nonlinear models produced consistent and significant estimates. However, this study also only used efficient designs and did not test generator-developed designs. In addition, the focus of analysis was on comparing cTTO and DCE valuation, hence a comparison of different DCE with duration approaches and design construction methods was not the main focus of the study.

Study aims

We used a 2×2 between-subjects design (i.e. a four-arm study) to investigate:

1. The impact of using a model that allows for discounting of time i.e. nonlinear time preferences, on EQ-5D-5L value sets. This is investigated in the context of systematically varying DCE design methods, specifically:
    a. the two DCE with duration approaches of *immediate death* and *full health*, and.
    b. the two DCE choice set construction methods, i.e. generator-developed designs and efficient DCE designs.
2. The consistency of value sets obtained using different DCE design methods by comparing results from using data from two diverse countries (Peru and Denmark).

Methods

The data

This study used data that were collected in Peru and Denmark alongside the national EQ-5D-5L valuation studies reported elsewhere [31–33]. The reason for choosing these two countries was opportunistic. Valuation studies for these two countries were being conducted around the

same time and provided the necessary data to explore the aims of this study.

Data collection: Peru

A detailed account of the data collection of the Peruvian value set can be found in Augustovski et al. [31]. Briefly, a population-based random sample of 1000 adults aged 18–75 years was used. The interviews were conducted at participants' homes and administered as a computer-assisted personal interview (CAPI). Interviews with respondents in Peru were conducted between April 2018 and February 2019. Quality of responses were determined by interviewer compliance with the interview methodology and the face validity of responses [31, 34].

300 respondents were randomly selected to complete 11 cTTO tasks first. All respondents completed 10 DCE latent scale choice sets; that is, a choice between pairs described by the 5 dimensions of the EQ-5D-5L as attributes (no attribute for duration). This was followed by 12 'matched pair' DCE with duration choice sets. The data analysed in this paper are from the responses to the DCE with duration choice sets.

Data collection: Denmark

A detailed account of the data collection in Denmark can be found in Jensen et al. [32]. Similar to the Peruvian data collection, a nationally representative sample of the Danish adult population was used. Individuals who agreed to participate could choose to be interviewed at home or at a nearby public institution. The study was administered using CAPI. Interviews in Denmark were conducted between October 2018 and December 2019.

In the Danish study, all respondents completed 10 cTTO tasks, followed by 7 DCE latent scale choice sets then finally the 12 'matched pair' DCE with duration choice sets. Once again, the focus of the current analysis was on the DCE with duration choice sets.

Arms

The 2 × 2 between-subjects design resulted in a 4-arm study (see Table 1). We compared the two approaches for anchoring the DCE with duration, i.e. the *immediate death* approach and the *full health* approach and then two design construction methods for the choice sets, i.e. generator-developed or the use of an efficient DCE design (see section on choice set construction for further details).

Table 1 The four arms in the study to compare DCE with duration data in Peru and Denmark

Anchoring approach to DCE with duration	Construction method of choice sets	
	Generator-developed	Efficient DCE design
Immediate death*	death_gen	death_eff
Full health**	fullhealth_gen	fullheatlh_eff

* Respondents saw immediate death as the third option (health state C) ** respondents saw full health for a shorter duration compared to health state B as the third option (health state C)

The DCE with duration choice set 'matched pair' format

Each DCE with duration choice set used the 'matched pair' format where each choice set included three options and was answered as a two-stage task. Respondents were first asked to choose between health state options A and B, and then between health state options B and C. Depending on the arm to which respondents were assigned, health state C was shown either as *immediate death* or as *full health* for a shorter duration than health state B. Participants completed a practice choice set followed by 12 choice sets used for data analysis. An English-language mock-up of the choice sets is provided in Appendix A.

Choice set construction method

The generator-developed designs used in this study were constructed in Mathematica using the approach in Street and Burgess [10], with the additional constraint that two of the health state attributes were to be at the same level in options A and B for the full health design. Further details on the construction process can be found in Appendix B. The final design of 105 choice sets was 86% efficient relative to the set of all choice sets. For both arms, duration was obtained from an initial coding of levels. In the arms anchored to full health, the full health duration was chosen randomly from an allowable set of shorter durations. The option to be compared to full health in each set was fixed, i.e. health state option B was always compared to health state option C.

The efficient DCE designs, i.e. used in arms *death_eff* and *fullhealth_eff*, were constructed using the Time Preferences Corrected QALY Design (TPC-QD) software with customised code that has been used in previous studies [11, 28]. These designs explicitly allow for the measurement of nonlinear time preferences [12]. As priors are needed to optimise this design, a pilot round of data collection was used to improve upon the initially selected priors and thus improve the DCE design efficiency. A pilot design was developed and administered to the first batch of respondents in arms *death_eff* and *fullhealth_eff*. The results were used as priors to develop a second set of choice sets, which were then given to a second batch of respondents in arms *death_eff* and *fullhealth_eff*. Final analyses were conducted with all data. Choice sets were blocked into 10 blocks of 12 choice sets each. Respondents were randomly assigned to one of the blocks of 12 choice sets. The order of health states A and B was randomised.

Data analysis

Modelling nonlinear time preferences and estimating discount rates

The mixed logit (MXL) model was estimated using exponential discounting to reflect nonlinear time preferences. This paper uses one specification of discounting for time as the focus of the study was on DCE design comparison rather than exploring optimal modelling of the discounting for time. Introducing further comparison between different discounting specifications would add complexity which does not necessarily address the aims of this paper. Exponential discounting was chosen over hyperbolic or power specifications as it is better identified which was considered desirable for the relatively small sample size per arm in this study.

Analyses were conducted in OpenBUGS. The model specification is based on Jonker et al. [11], which allows for linear time preference as a special case. For the MXL model, utility U for individual i for alternative n in choice set j is specified as:

$$\begin{gathered} U_{inj} = (\beta_i \ \mathrm{X}_{inj})\, NPV_{inj} + \varepsilon_{inj}\,, \\ \beta_i \ \sim \ MVN \ (\beta\,, \sum) \end{gathered} \tag{1}$$

where β_i are the preference parameters associated with individual i that are multivariate normal distributed with population mean β and covariance matrix, $\sum$. X_{inj} are the attribute levels faced by individual i in alternative n of choice set j. Net present value NPV_{inj} is the sum of the present value of future life years ($TIME_{inj}$). In this case, net present value, NPV_{inj}, is discounted using the standard exponential function. The standard exponential function allows for linear time preference as a special case when the discount rate (r) is equal to zero. This can be expressed as:

$$\begin{gathered} NPV_{inj} = TIME_{inj}, \ \text{if } r = 0, \\ NPV_{inj} = \left(\left(1 - e^{(-r)TIME_{inj}}\right)\right)/(e^r - 1)), \ \text{if } r \neq 0. \end{gathered} \tag{2}$$

.

Table 2 Number of respondents and choice sets

Arm	Peru		Denmark	
	No. of choice sets	No. of respondents	No. of choice sets	No. of respondents
death_gen	125	221	125	253
fullhealth_gen	105	241	105	228
death_eff	240	229	240	251
fullhealth_eff	228	251	240	256

Reporting and comparison of results

After MXL model parameters were estimated, the QALY scale parameter estimates were obtained by dividing the estimate of the mean, $\widehat{\beta}$, by the first element of $\widehat{\beta}$, i.e. the perfect health/duration intercept, $\widehat{\beta}_1$. This can be expressed as:

$$QALY_{decrement} = \widehat{\beta} \,/\, \widehat{\beta}_1 \tag{3}$$

The significance of the QALY scale parameter estimates compared to level 1, i.e. baseline, and any disordering was examined by arm and dataset. Significance is indicated by the non-inclusion of 0 in the 95% credible intervals. To compare model performance across arms, the mean absolute error (MAE) was calculated, with a lower score indicating better model performance.

For the purposes of calculating the utility weights associated with health states, QALY scale estimates were reordered, and are summarised in Appendix C. If dimension level parameter estimates were inconsistent in order and not significant at the 5% level (i.e. the 95% credible intervals did not comprise 0), then it was assigned the parameter estimate of the previous dimension level. For instance, the parameter estimate for the dimension of Mobility at level 2 (i.e. MO2) was 0.01 and not significant at the 5% level, i.e. the 95% credible interval comprises of 0. Therefore, its value was constrained to the previous dimension level, i.e. Mobility at level 1 (MO1), which is 0, i.e. no difference in utility to MO1. In contrast, MO3 was also not significant at the 5% level, but because its value of -0.03 exhibited consistent ordering (i.e. decrement to MO2), its parameter estimate was left unchanged. Utility weights of selected health states were used for comparison purposes across arms, datasets and also against the published Peruvian and Danish datasets. Estimated discount rates were examined for significance ($p < 0.05$) and compared across arms.

Results

Data cleaning and screening of respondents

Table 2 outlines total number of respondents included in the analysis and the number of choice sets in each arm. During the data cleaning process in the main valuation study, 27 respondents in the Peruvian data set and 36 respondents in the Danish data set were dropped due to low quality responses. An additional 11 respondents each from the Peruvian data and the Danish data were excluded in the current study due to incomplete or missing choice data. Respondents were only included for analysis if they completed all 24 assigned choice sets.

The number of respondents in the *full health* approach exceeded those in the *immediate death* approach in Peru.

For Denmark, the highest number of respondents was in the *fullhealth_eff* arm, with the lowest in the *fullhealth_gen* arm. For Peru, there were just over 200 respondents in each arm. The number of choice sets were higher in the arms using efficient designs, as two batches of respondents were used (see Methods: Choice set construction method for further details).

Comparison of MXL model results

Table 3 summarises the MXL model QALY scale estimates. For comparison purposes, parameter estimates that were not significant at the 5% level in comparison to level 1, i.e. the baseline and/or misordered were highlighted in red. The Peruvian dataset had more non-significant estimates compared to the Danish dataset. Consistently across both datasets, the *death_gen* arms and the *death_eff* had the highest and lowest number of

Table 3 MXL model QALY scale estimates with 95% credible intervals (95% CI)

Peru	**death_gen**	**death_eff**	**fullhealth_gen**	**fullhealth_eff**
MO2	0.01 (-0.04, 0.05)	-0.03 (-0.08, 0.01)	0.02 (-0.03, 0.08)	-0.03 (-0.07, 0.01)
MO3	-0.03 (-0.10, 0.05)	-0.12 (-0.16, -0.07)	-0.11 (-0.17, -0.05)	-0.09 (-0.13, -0.04)
MO4	-0.21 (-0.27, -0.14)	-0.19 (-0.24, -0.14)	-0.2 (-0.28, -0.14)	-0.24 (-0.29, -0.19)
MO5	-0.41 (-0.49, -0.32)	-0.37 (-0.43, -0.31)	-0.56 (-0.68, -0.46)	-0.48 (-0.55, -0.41)
SC2	0.02 (-0.05, 0.07)	-0.02 (-0.06, 0.03)	0 (-0.06, 0.06)	0 (-0.04, 0.04)
SC3	-0.03 (-0.08, 0.02)	-0.07 (-0.12, -0.02)	0 (-0.06, 0.06)	-0.02 (-0.06, 0.02)
SC4	-0.13 (-0.18, -0.08)	-0.14 (-0.19, -0.09)	-0.17 (-0.24, -0.11)	-0.13 (-0.18, -0.08)
SC5	-0.25 (-0.32, -0.20)	-0.22 (-0.27, -0.17)	-0.34 (-0.42, -0.26)	-0.29 (-0.35, -0.24)
UA2	-0.03 (-0.09, 0.03)	-0.03 (-0.07, -0.02)	-0.04 (-0.10, 0.02)	-0.07 (-0.12, -0.03)
UA3	-0.1 (-0.15, -0.04)	-0.05 (-0.10, 0.00)	-0.08 (-0.14, -0.01)	-0.08 (-0.13, -0.04)
UA4	-0.18 (-0.24, -0.13)	-0.13 (-0.18, -0.08)	-0.29 (-0.37, -0.22)	-0.22 (-0.27, -0.17)
UA5	-0.39 (-0.45 -0.33)	-0.32 (-0.38, -0.26)	-0.54 (-0.66, -0.45)	-0.47 (-0.55, -0.40)
PD2	0 (-0.04, 0.05)	-0.04 (-0.08, 0.00)	-0.02 (-0.07, 0.03)	-0.04 (-0.08, 0.00)
PD3	-0.08 (-0.12, -0.04)	-0.05 (-0.09, -0.01)	-0.04 (-0.10, 0.01)	-0.1 (-0.15, -0.06)
PD4	-0.2 (-0.26, -0.14)	-0.17 (-0.22, -0.13)	-0.28 (-0.36, -0.21)	-0.21 (-0.26, -0.17)
PD5	-0.44 (-0.53, -0.36)	-0.37 (-0.44, -0.30)	-0.56 (-0.69, -0.45)	-0.49 (-0.57, -0.42)
AD2	-0.01 (-0.06, 0.06)	-0.05 (-0.09, -0.00)	0.01 (-0.05, 0.06)	-0.02 (-0.06, 0.02)
AD3	-0.04 (-0.09, 0.05)	-0.06 (-0.10, -0.02)	-0.07 (-0.12, -0.01)	-0.04 (-0.08, 0.00)
AD4	-0.15 (-0.20, -0.09)	-0.16 (-0.20, -0.11)	-0.14 (-0.21, -0.09)	-0.17 (-0.22, -0.13)
AD5	-0.32 (-0.38, -0.25)	-0.31 (-0.37, -0.25)	-0.37 (-0.47, -0.29)	-0.31 (-0.37, -0.25)
Denmark	**death_gen**	**death_eff**	**fullhealth_gen**	**fullhealth_eff**
MO2	-0.03 (-0.07, 0.01)	-0.01 (-0.04, 0.03)	-0.02 (-0.05, 0.01)	-0.01 (-0.03, 0.01)
MO3	-0.07 (-0.11, -0.03)	-0.04 (-0.08, -0.01)	-0.11 (-0.14, -0.08)	-0.06 (-0.09, -0.04)
MO4	-0.12 (-0.17, -0.08)	-0.12 (-0.15, -0.09)	-0.16 (-0.20, -0.13)	-0.14 (-0.17, -0.12)
MO5	-0.19 (-0.23, -0.15)	-0.17 (-0.21, -0.14)	-0.23 (-0.27, 0.20)	-0.21 (-0.24, -0.18)
SC2	-0.01 (-0.06, 0.04)	-0.02 (-0.05, 0.02)	-0.04 (-0.06, 0.00)	0 (-0.03, 0.02)
SC3	-0.04 (-0.08, 0.00)	-0.05 (-0.08, -0.01)	-0.07 (-0.09, -0.04)	-0.02 (-0.04, 0.00)
SC4	-0.09 (-0.13, -0.05)	-0.11 (-0.14, -0.07)	-0.18 (-0.21, -0.14)	-0.11 (-0.14, -0.09)
SC5	-0.16 (-0.20, -0.11)	-0.19 (-0.22, -0.15)	-0.2 (-0.24, -0.17)	-0.2 (-0.23, -0.17)
UA2	-0.02 (-0.07, 0.03)	-0.02 (-0.05, 0.02)	-0.02 (-0.05, 0.00)	0.01 (-0.01, 0.03)
UA3	-0.07 (-0.11, -0.03)	-0.05 (-0.08, -0.01)	-0.09 (-0.12, -0.06)	-0.05 (-0.07, -0.03)
UA4	-0.18 (-0.22, -0.14)	-0.15 (-0.19, -0.12)	-0.17 (-0.20, -0.14)	-0.15 (-0.17, -0.12)
UA5	-0.22 (-0.27, -0.17)	-0.22 (-0.25, -0.18)	-0.25 (-0.29, -0.21)	-0.21 (-0.24, -0.18)
PD2	-0.04 (-0.08, 0.00)	-0.04 (-0.07, -0.01)	-0.06 (-0.10, -0.03)	-0.03 (-0.05, -0.01)
PD3	-0.11 (-0.15, -0.06)	-0.13 (-0.17, -0.10)	-0.11 (-0.14, -0.08)	-0.11 (-0.14, -0.09)
PD4	-0.35 (-0.40, -0.30)	-0.33 (-0.38, -0.29)	-0.37 (-0.42, -0.32)	-0.36 (-0.39, -0.32)
PD5	-0.59 (-0.66, -0.53)	-0.52 (-0.58, -0.47)	-0.54 (-0.61, -0.48)	-0.59 (-0.65, -0.54)
AD2	-0.01 (-0.05, 0.03)	-0.11 (-0.14, -0.08)	-0.08 (-0.11, -0.05)	-0.06 (-0.08, -0.04)
AD3	-0.07 (-0.12, -0.02)	-0.22 (-0.25, -0.18)	-0.18 (-0.21, -0.15)	-0.17 (-0.20, -0.14)
AD4	-0.35 (-0.41, -0.30)	-0.46 (-0.51, -0.41)	-0.44 (-0.50, -0.39)	-0.42 (-0.46, -0.38)
AD5	-0.55 (-0.63, -0.48)	-0.7 (-0.78, -0.63)	-0.69 (-0.77, -0.62)	-0.64 (-0.70, -0.59)

Note: Parameter estimates that were not significant at the 5% level (i.e. the 95% credible interval does not include 0) in comparison to the baseline, were highlighted in red

Table 4 Model performance: mean absolute errors

Mean absolute error	death_gen	death_eff	fullhealth_gen	fullhealth_eff
Peru	0.45	0.5	0.4	0.34
Denmark	0.4	0.29	0.23	0.31

Table 5 Utilities of select health States

Peru	death_gen	death_eff	fullhealth_gen	fullhealth_eff
21111	1	0.97	1	0.97
12111	1	0.98	1	1
11211	0.97	0.97	0.96	0.93
11121	1	0.96	0.98	0.96
11112	0.99	0.95	1	0.98
11111	1	1	1	1
11131	0.92	0.95	0.96	0.9
11232	0.88	0.87	0.92	0.81
21231	0.89	0.89	0.92	0.8
22222	0.96	0.83	0.94	0.84
31331	0.79	0.78	0.77	0.73
32332	0.78	0.71	0.77	0.71
33333	0.72	0.65	0.7	0.67
44444	0.13	0.21	-0.08	0.03
55555	-0.81	-0.59	-1.37	-1.04
%<0	11%	6%	27%	19%
Denmark	**death_gen**	**death_eff**	**fullhealth_gen**	**fullhealth_eff**
21111	0.97	0.99	0.98	0.99
12111	0.99	0.98	0.96	1
11211	0.98	0.98	0.98	1
11121	0.96	0.96	0.94	0.97
11112	0.99	0.89	0.92	0.94
11111	1	1	1	1
11131	0.89	0.87	0.89	0.89
11232	0.86	0.74	0.79	0.83
21231	0.84	0.84	0.85	0.88
22222	0.89	0.8	0.78	0.9
31331	0.75	0.78	0.69	0.78
32332	0.73	0.65	0.57	0.72
33333	0.64	0.51	0.44	0.59
44444	-0.09	-0.17	-0.32	-0.18
55555	-0.71	-0.8	-0.91	-0.85
%<0	15%	22%	29%	21%

Note: Utility weights of health states were based on QALY scale parameter estimates with forced ordering, different from Table 3. QALY scale parameter estimates with forced ordering is available in Appendix C

non-significant parameters, respectively. The arms that anchored choice sets using full health tended to be relatively consistent, with the same number of non-significant parameter estimates in each dataset, regardless of choice set construction method.

Table 4 summarises mean absolute error (MAE) by arm, with a lower score indicating better model performance. In terms of MAE, there was no clear construction method that performed better than the other. For DCE with duration anchoring approach, the *full health* arms had lower MAEs in the Peruvian dataset. For Denmark, the *death_gen* arm had the highest MAE, with the lowest MAE in the *fullhealth_gen* arm.

Comparison of health state utilities

Table 5 illustrates a broad range of health states and their utilities. Most of the health states shown are from the 33 most common health states in the general population [29]. For 'milder' health states (health states with dimension levels 3 or lower, e.g. 21231) the range of utilities tend to be quite stable across approach to the DCE with duration choice set and construction method. That is, 'milder' health states were not as sensitive to the approach and construction method of the DCE with duration choice set.

For more 'severe' health states (health states with dimension levels 3 or higher, e.g. 44444), utilities were impacted by the anchoring approach to the DCE with duration choice set. The *full health* arms seemed to produce a wider range of utilities compared to the *death* arms, regardless of construction method (from 1 to -1.37 and to -0.91 for the Peruvian and Danish respondents, respectively). More severe health states were also given worse utility weights in the *full health* compared to the *death* arms. The construction method of DCE with duration choice sets also appeared to have an impact on health state utilities. The arms with generator-developed designs tended to produce the highest and lowest utility values for a particular health state.

For instance, looking at the Danish utilities for health state 33,333, the arms using generator-developed designs had the highest utility value at 0.64 in the *death_gen* arm and the lowest utility value of 0.44 in the *fullhealth_gen* arm. A similar pattern was found for the Peruvian utilities for the same health state and in most of the other health states, including the worst health state 55,555.

Comparison of health state utilities with published value sets

The health state utilities of the four arms of the study were also compared against the published value sets for Peru and Denmark [1]. This has been summarised in Table 6.

The utilities of the published Peruvian data set were quite different to those from the four arms. Most notably, the utility of 22222 (0.441) is much lower in the published value set compared to the 4 arms in this study. The differences may, in part, reflect the different data sets used; the published Peruvian value set was based on the cTTO data only. The published Danish value set, which was based on

Table 6 Comparison of health state utilities with published value sets

		Health State Utilities				
Country	**Arm**	**11111**	**22222**	**33333**	**44444**	**55555**
Peru	**Published***	1	0.441	0.148	−0.28	−1.073
	death_gen	1	0.96	0.72	0.13	-0.81
	death_eff	1	0.83	0.65	0.21	-0.59
	fullhealth_gen	1	0.98	0.7	-0.08	-1.37
	fullhealth_eff	1	0.84	0.67	0.03	-1.04
Denmark	**Published****	1	0.771	0.571	−0.251	−0.757
	death_gen	1	0.89	0.64	-0.09	-0.71
	death_eff	1	0.8	0.51	-0.17	-0.8
	fullhealth_gen	1	0.78	0.44	-0.32	-0.91
	fullhealth_eff	1	0.9	0.59	-0.18	-0.85

*The Peruvian published value set is based on cTTO data only **The Danish published value set is based on a combination of cTTO and DCE data

Table 7 Estimated MXL model discount rates

Mean (95% CI)	Peru	Denmark
Fullheath_eff	0.25 (0.21–0.32)	0.09 (0.07–0.11)
Fullhealth_gen	0.15 (0.10–0.22)	0.01 (-0.01-0.04)
Death_eff	0.80 (0.56-1.00)	0.28 (0.23–0.35)
Death_gen	1.17 (1.06–1.27)	0.27 (0.21–0.34)

a combination of the cTTO and DCE data, is relatively more consistent with the results of the four arms.

Comparison of discount rates

The estimated discount rates are summarised in Table 7. Except for the *fullhealth_gen* arm from the Danish data set, the 95% credible intervals were all above 0, suggesting that when modelling explicitly for nonlinear time preferences, there was evidence that discounting was present. A sensitivity analysis was also conducted on the treatment of immediate death as '0', see Appendix D. Apart from the *death_gen* arm in the Peruvian data, all the other arms had a death dummy estimate that included 0 in the 95% credible interval. This suggests the assumption of immediate death as represented by 0 years in duration does hold in most cases, at least when nonlinear modelling is used.

For the Peruvian data, discount rates varied widely from 15 to 117%, which lacks face validity. In contrast, the variation in the Danish data was much less, from 1 to 28%. It was also noted that discount rates were sensitive to the approach to the DCE with duration choice set. The *death* arms tended to produce higher discount rates compared to the *full health* arms, irrespective of construction method. In both datasets, the *full health* arms consistently outperformed the *death* arms based on face validity of discount parameters.

A summary of the key study findings is provided in Table 8.

Discussion

In this four-arm study, EQ-5D-5L value sets were estimated with an exponential discount function to account for nonlinear time preferences. It was found that DCE with duration approach, DCE design construction

Table 8 Summary of key findings

Generator-developed designs • No consistent pattern in terms of impact on significance of QALY parameter estimates • No consistent pattern in terms of model performance (MAE) • Produces more 'extreme' health state utilities, designs tended to produce the highest and lowest utility values for a particular health state • No consistent pattern in terms of impact on discount rates • Combination of generator-developed design with the DCE with duration using immediate death produced greatest number of non-significant QALY parameter estimates, and worse model performance (MAE) compared to other arms	DCE with duration using immediate death • Produced more 'extremes', e.g. the death_gen arm had the highest and the death_eff arm the lowest number of non-significant parameters • No consistent pattern in terms of model performance (MAE) • Fewer health states considered worse than death • Discount rate estimates were consistently higher, and discount rates lacked face validity • Combination of generator-developed design with the DCE with duration using immediate death produced greatest number of non-significant QALY parameter estimates, and worse model performance (MAE) compared to other arms
Efficient designs • Consistent pattern in terms of number of significant QALY parameter estimates • Consistent pattern in terms of model performance (MAE) • Consistent results in terms of health state utilities • No consistent pattern in terms of impact on discount rates	DCE with duration using full health • Consistent pattern in terms of number of significant QALY parameter estimates • Consistent results in terms of model performance (MAE) • Produces a wider range of utilities, and many more health states considered worse than death (compared to the immediate death approach) • Better face validity of discount rate estimates

method design and data source affected the value sets. Compared to the DCE with duration approach using *immediate death*, the approach using *full health* better supported the estimation of a model that accounts for nonlinear time preferences in terms of discount rates and inconsistencies of model estimates. The arms using the *full health* approach also showed better model performance in the Peruvian dataset, although it was less clear for the Danish data. Regarding DCE design construction method, arms with the generator-developed designs tended to produce more 'extreme' ranges of utility values (producing either the widest or narrowest range) compared to arms using the efficient DCE designs where results were more consistent.

A consistent finding across arms was the presence of discounting, although the estimated rates varied quite substantially. Discount rates were generally more volatile in the Peruvian data set, which had more inconsistent or non-significant parameter estimates. Karim et al. [33] found evidence for potential bias in the Peruvian data set induced by cTTO being completed first by a portion of respondents. This could have contributed to the variability seen in the current study and demonstrates how nonlinear modelling methods might be sensitive to data quality in terms of parameter estimation and discount rate estimates.

We noted that more severe health states were sensitive to both DCE with duration approach and DCE design construction method, whereas this mattered less for milder health states. Regardless of country, respondents who saw the *full health* approach were more likely to give worse valuations than respondents who saw the *immediate death* approach for the same health states. Anchoring to immediate death might have encouraged respondents to be more willing to accept living in worse health states. For the Peruvian data, the wider utility range of the *full health* approach makes the DCE with duration results better align with the published value set. DCE with duration approach mattered les in the Danish data, with arms generally aligned with the published value set.

The *immediate death* approach, which is often analysed in the literature using a linear model, has exhibited a notable sensitivity to how the "dead" option is integrated into the analysis [20]. This sensitivity can result in a lack of generalisability across different segments of the task, but despite the use of a nonlinear model in the current study, disparities persisted between anchoring approaches. Jonker and Norman [29] offered a possible explanation, highlighting that the modelling assumptions of the QALY do not hold in the unconstrained DCE with duration paired with the death option. Consequently, researchers and end-users of utility estimates produced from DCE with duration tasks must carefully evaluate the implications of the chosen anchor and its impact on the validity and reliability of their findings.

Regarding DCE design construction method, we found more consistent results with the efficient DCE designs. However, it is difficult to decide which construction method is the preferred one. Generator-developed construction methods are easier to implement as they do not require priors to inform the design and, when priors are used, they can be robust to misspecified priors compared to efficient DCE designs depending on the assumed distribution [15]. On the other hand, efficient DCE construction methods can be tailored to the needs of even the most demanding models; design optimisation software is available for the EQ-5D instruments [12], but these designs are difficult to generate for other instruments.

Methodological limitations and directions for future research

There are several limitations to this study. The 'matched pair' format was used for both approaches to the DCE with duration —this is not the conventional format used for the *immediate death* approach. Future studies comparing these two methods could use the "best- worst" format instead to investigate how these impact on value sets derived using nonlinear modelling methods.

There were also twice the number of choice sets used in the arms with efficient designs compared to the generator-developed designs. This was due to the efficient designs using two rounds of data collection, with the first data collection informing choice sets for the second round. Future studies comparing these two DCE design construction methods could compare the performance of generator-developed designs versus efficient designs when a similar or equivalent number of choice sets are used for each.

Although the treatment of immediate death as '0' was generally supported, we did find that respondents valued health states differently when comparing against full health or immediate death. In particular, comparisons against full health tended to produce a wider range of utilities. The issue of how respondents interpret and value 'immediate death' warrants qualitative exploration in future studies.

Although the aims of this study were focused on the impact of methodological variations on value sets, it must be acknowledged that estimated discount rates varied quite widely. It is possible that larger sample sizes are needed as there were only 220–250 respondents per arm. Indeed, other studies that have used nonlinear time preference modelling methods had much larger sample sizes, e.g. Jonker et al. [11] included 1775 respondents for their analyses. Future studies could include more respondents to see if this can reduce variability seen in discount rates.

It is also important to keep in mind that estimated empirical discount rates are usually unrelated to discount rates in a health technology assessment setting [11, 35]. This study also only used an exponential function to account for nonlinear time preferences and did not test whether other discounting functions may have provided a better fit. This was beyond the scope of the current study, and the findings should be interpreted in the context of this limitation. Future studies could explore other discounting functions and its impact on estimated discount rates.

In general, it was observed that the model to account for nonlinear time preferences in this study tended to perform better on data from Denmark compared to Peru. Future research could explore methods to improve DCE methods for health state valuation, particularly if data quality may be an issue. The current study also only focused on the EQ-5D-5L, and future studies should explore whether nonlinear modelling is appropriate for other instruments, particularly for younger age groups such as the EQ-5D-Y-3L and EQ-5D-Y-5L.

Conclusions

As far as the authors are aware, this is the first study to provide a head-to-head comparison of the two commonly used DCE with duration approaches and associated construction methods in valuation studies that account for nonlinear time preferences, including a comparison of data from two diverse countries. Several methodological issues have been raised, and directions for future work have been suggested.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12955-025-02391-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Author contributions

Abbreviations = Last name then first nameConcept and design: S.D., J.M., M. B., A.F., G.C., N.R., V.R., S.E. Acquisition of data: J.C., F.A., T. R., V.R. Analysis and interpretation of data: Y. A., S. D., S. B., J. M. Drafting of the manuscript: Y. A., M. B. Critical revision of manuscript: all authorsStatistical Analysis: Y. A., S. D., B. S., J.M., J. M. Provision of study materials or patients: J. C., G. C., T. R., A. F. Obtaining funding: Elly StolkOther: S. D. (design of DCE with duration choice sets). J. M. (design of DCE with duration choice sets)

Funding

This work was partly funded by the EuroQol Research Foundation. It was also completed as part of author Alice Yu's PhD. She received a PhD scholarship (University of Technology Sydney Chancellors Research Scholarship).

Data availability

No datasets were generated during the current study.

Declarations

Competing interests

The authors declare no competing interests.

Ethics and consent to participate declarations

Not applicable.

Author details

[1]Centre for Health Economics Research and Evaluation (CHERE), Faculty of Health, University of Technology Sydney (UTS), Sydney, NSW, Australia
[2]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, Netherlands
[3]Health Technology Assessment and Health Economics Department of the Institute for Clinical Effectiveness and Health Policy (IECS-CONICET), University of Buenos Aires, Buenos Aires, Argentina
[4]Danish Center for Health Services Research, Department of Clinical Medicine, Aalborg University, Aalborg, Gistrup, Denmark
[5]Department of Clinical Research, University of Southern Denmark, Odense, Denmark
[6]Department of Economics and Business Economics, Business and Social Sciences, Aarhus University, Aarhus, Denmark
[7]Division of Cancer Epidemiology, Gerald Bronfman Department of Oncology, McGill University, Montreal, Canada
[8]School of Population Health, Curtin University, Perth, Australia
[9]EuroQol Research Foundation, Rotterdam, The Netherlands

Received: 15 November 2024 / Accepted: 2 June 2025
Published online: 15 June 2025

References

1. Devlin N, Roudijk B, Ludwig K, editors. Value sets for EQ-5D-5L: A compendium, comparative [internet]eview & user guide [Internet]. Cham (CH): Springer; 2022.
2. Roudijk B, Jonker MF, Bailey H, et al. A direct comparison between discrete choice with duration and composite time trade-off methods: do they produce similar results? Value Health. 2024;27:1280–8.
3. Mulhern B, Norman R, Street DJ, et al. One method, many methodological choices: A structured review of Discrete-Choice experiments for health state valuation. PharmacoEconomics. 2019;37:29–43.
4. Wang H, Rowen DL, Brazier JE, et al. Discrete choice experiments in health state valuation: A systematic review of progress and new trends. Appl Health Econ Health Policy. 2023;21:405–18.
5. Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ. 2018;27:7–22.
6. Versteegh M, Vermeulen M, Evers KMAA. Dutch tariff for the Five-Level version of EQ-5D. Value Health. 2016;19:343–52.
7. Mulhern B, Bansback N, Hole AR, et al. Using discrete choice experiments with duration to model EQ-5D-5L health state preferences: testing experimental design strategies. Med Decis Making. 2016;37:285–97.
8. Miyamoto JM, Wakker PP, Bleichrodt H, et al. The zero-condition: A simplifying assumption in QALY measurement and multiattribute utility. Manage Sci. 1998;44:839–49.
9. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status. Oper Res. 1980;28:206–24.
10. Craig BM, Rand K, Bailey H, et al. Quality-Adjusted Life-Years without constant proportionality. Value Health. 2018;21:1124–31.
11. Jonker MF, Donkers B, de Bekker-Grob EW, et al. Advocating a paradigm shift in Health-State valuations: the Estimation of Time-Preference corrected QALY tariffs. Value Health. 2018;21:993–1001.
12. Jonker MF, Bliemer MCJ. On the optimization of bayesian D-Efficient discrete choice experiment designs for the Estimation of QALY tariffs that are corrected for nonlinear time preferences. Value Health. 2019;22:1162–69.
13. Street DJ, Burgess L. The construction of optimal stated choice experiments: theory and methods. Wiley; 2007.
14. Lim S, Jonker MF, Oppe M, et al. Severity-Stratified discrete choice experiment designs for health state evaluations. PharmacoEconomics. 2018;36:1377–89.
15. Traets F, Sanchez DG, Vandebroek M. Generating optimal designs for discrete choice experiments in R: the Idefix package. J Stat Softw. 2020;96:1–41.

16. Viney R, Norman R, Brazier J et al. An Australian discrete choice experiment to value EQ-5D health states Health Economics. 2014; 23: 729– 42.
17. Bansback N, Brazier J, Tsuchiya A, et al. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31:306–18.
18. Tsuchiya A, Dolan P. The QALY model and individual preferences for health States and health profiles over time: A systematic review of the literature. Med Decis Making. 2005;25:460–67.
19. Norman R, Mulhern B, Lancsar E, et al. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. PharmacoEconomics. 2023;41:427–38.
20. Norman R, Mulhern B, Viney R. The impact of different DCE-Based approaches when anchoring utility scores. PharmacoEconomics. 2016;34:805–14.
21. Norman R, Viney R, Brazier J, et al. Valuing SF-6D health States using a discrete choice experiment. Med Decis Making. 2013;34:773–86.
22. Mulhern BJ, Bansback N, Norman R et al. Valuing the SF-6Dv2 classification system in the united Kingdom using a Discrete-choice experiment with duration. Med Care. 2020; 58.
23. Shiroiwa T, King MT, Norman R, et al. Japanese value set for the EORTC QLU-C10D: A multi-attribute utility instrument based on the EORTC QLQ-C30 cancer-specific quality-of-life questionnaire. Qual Life Res. 2024;33:1865–79.
24. Finch AP, Gamper E, Norman R, et al. Estimation of an EORTC QLU-C10 value set for Spain using a discrete choice experiment. PharmacoEconomics. 2021;39:1085–98.
25. King MT, Revicki DA, Norman R, et al. United States value set for the functional assessment of Cancer Therapy-General eight dimensions (FACT-8D), a Cancer-Specific Preference-Based quality of life instrument. PharmacoEconomics - Open. 2024;8:49–63.
26. McTaggart-Cowan H, King MT, Norman R, et al. The FACT-8D, a new cancer-specific utility algorithm based on the functional assessment of Cancer Therapies-General (FACT-G): a Canadian valuation study. Health Qual Life Outcomes. 2022;20:97.
27. 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. 2013;11:287–98.
28. Jonker MF, Attema AE, Donkers B, et al. Are health state valuations from the general public biased?? A test of health state reference dependency using Self-assessed health and an efficient discrete choice experiment. Health Econ. 2017;26:1534–47.
29. Jonker MF, Norman R. Not all respondents use a multiplicative utility function in choice experiments for health state valuations, which should be reflected in the elicitation format (or statistical analysis). Health Econ. 2022;31:431–39.
30. Mulhern B, Norman R, Shah K, et al. How should discrete choice experiments with duration choice sets be presented for the valuation of health states?? Med Decis Mak. 2018;38:306–18.
31. Augustovski F, Belizán M, Gibbons L, et al. Peruvian valuation of the EQ-5D-5L: A direct comparison of time Trade-Off and discrete choice experiments. Value Health. 2020;23:880–88.
32. Jensen CE, Sørensen SS, Gudex C, et al. The Danish EQ-5D-5L value set: A hybrid model using cTTO and DCE data. Appl Health Econ Health Policy 2021 Appl Health Econ Health Policy. 2021;19:579–91.
33. Karim S, Craig BM, Tejada RA, et al. Preference heterogeneity in health valuation: a latent class analysis of the Peru EQ-5D-5L values. Health Qual Life Outcomes. 2023;21:1.
34. Ramos-Goñi JM, Oppe M, Slaap B, et al. Quality control process for EQ-5D-5L valuation studies. Value in health. J Int Soc Pharmacoeconomics Outcomes Res. 2017;20:466–73.
35. Brouwer WB, Niessen LW, Postma MJ, et al. Need for differential discounting of costs and health effects in cost effectiveness analyses. BMJ. 2005;331:446–48.

Publisher's note