

The Impact of Different DCE-Based Approaches When Anchoring Utility Scores

Richard Norman¹ · Brendan Mulhern² · Rosalie Viney²

Published online: 31 March 2016
© Springer International Publishing Switzerland 2016

Abstract

Background Discrete choice experiments (DCEs) have been proposed as a method to estimate utility weights for health states within utility instruments. However, the most appropriate method to anchor the utility values on the full health to dead quality-adjusted life year (QALY) scale remains uncertain. We test four approaches to anchoring in which dead is valued at zero and full health at one.

Methods We use data from two DCEs valuing EQ-5D-3L and EQ-5D-5L health states, which presented pairs of health profiles with an associated duration, and a dead option. The approaches to anchoring the results on the required scale were (1) using only preferences between non-dead health profiles; (2) including the dead data, treating it as a health profile with zero duration; (3) explicitly modelling both duration and dead; and (4) using the preferences regarding the dead health state as an external anchor subsequent to the estimation of approach 1.

Results All approaches lead to differences in the scale of utility decrements, but not the ranking of EQ-5D health states. The models differ in their ability to predict preferences around dead health states, and the characteristics of

the value sets in terms of their range and the proportion of states valued as worse than dead.

Discussion Appropriate anchoring of DCEs with or without complementary time trade-off (TTO) data remains unresolved, and the method chosen will impact on health resource allocation decision making employing the value sets.

Key Points

The approach to anchoring discrete choice experiment (DCE) results onto the 0–1 scale required for quality-adjusted life years matters.

The approach to anchoring used in many existing DCE studies leads to a large range of utility scores, relative to the range resulting from other approaches in this study.

Using alternative approaches would impact on resource allocation decisions.

✉ Richard Norman
Richard.Norman@curtin.edu.au

Brendan Mulhern
Brendan.Mulhern@chere.uts.edu.au

Rosalie Viney
Rosalie.Viney@chere.uts.edu.au

¹ School of Public Health, Curtin University, Perth, Australia

² Centre for Health Economics Research and Evaluation (CHERE), University of Technology Sydney, Sydney, Australia

1 Introduction

Discrete choice experiments (DCEs) are becoming an increasingly widely used tool in health economics generally [1]. An area in which they have shown promise in terms of impacting health policy is in the valuation of health states for use in construction of preference-based indexes to calculate quality-adjusted life years (QALY) [2–6]. However, as a relatively novel area, a number of methodological issues require consideration as the DCE approach becomes more widely adopted for this purpose.

An important one is the appropriate way to anchor health states on the full health (1) to dead (0) utility scale required for the QALY model. Bleichrodt et al. [7] identified that relatively few assumptions are required to impose the QALY model on preferences, specifically that constant proportionality of trade-off between survival duration and quality of life holds and that the zero condition, in which a survival duration of zero has a utility of zero (irrespective of quality of life) is asserted. While the constant proportionality assumption is generally imposed by standard valuation methods, it can be explored in the DCE context by varying survival durations in the experiment and then comparing different econometric specifications that impose or relax the constraint that the utility function is linear with respect to duration. The zero condition, while arguably a less contentious restriction, is not as readily testable. It is this latter assumption that we are primarily concerned with. Concerning the zero condition, we first need to impose a model on the data which equalises preferences around health profiles with a duration of zero. Secondly, and more problematically, we want to explore whether a duration of zero is equivalent to the dead health state. The zero condition is important in deriving preference weights that are consistent with the QALY model from stated preference data, because it allows an empirically estimated utility function to be re-scaled (or anchored) on the 0–1 death to full health scale. The issue of how to integrate DCE preferences into the QALY framework (with a health state equivalent to being dead valued at zero and full health equivalent to the best health state described by the measure valued at one) remains unresolved, and is the central focus of this paper.

We conventionally assume that the utility of an alternative i in a choice set C_n to an individual n is given by:

$$U_{in} = V(X_{in}, \beta) + \varepsilon_{in}. \quad (1)$$

The $V(X_{in}, \beta)$ term is the explainable (or systematic) component of utility which is determined by characteristics of the choice or the individual n . However, there is also an error term which differs over alternatives and individuals and makes prediction of choice uncertain. It is assumed that the individual will choose the option if the utility associated with that option is higher than any alternative option. If we assumed there are J items in C_n , the choice is defined as:

$$y_{in} = f(U_{in}) = 1_{\text{iff}} U_{in} = \max_j \{U_{ij}\} \cdot \forall j \neq i \in C_n \quad (2)$$

Alternative i is chosen if and only if:

$$(V_{in} + \varepsilon_{in}) > (V_{jn} + \varepsilon_{jn}) \cdot \forall j \neq i \in C_n \quad (3)$$

which can be rearranged to yield:

$$(V_{in} - V_{jn}) > (\varepsilon_{jn} - \varepsilon_{in}) \cdot \forall j \neq i \in C_n \quad (4)$$

Neither the systematic nor the error components in the utility function are directly observed. Therefore, analysis is reliant on observing choices and inferring the terms from that. In random utility theory (RUT), we assume that the difference in utility between two options (in this case, i and j) is proportional to the frequency that one is chosen over the other [8, 9].

Uptake of DCE methodology in the area has been growing, particularly the inclusion of DCE tasks within the EuroQoL protocol for valuing health states within the recently developed EQ-5D-5L [10, 11]. The question of anchoring is an important one, with one approach using time trade-off (TTO) data as an external anchor. Thus, if a TTO or standard gamble (SG) study values some health state (often the worst) at some value, the DCE will apply a linear transformation such that the DCE and TTO/SG scales coincide both at full health and at the anchor health state. A problem with this approach is that it assumes the value of the anchor is correct; this is problematic if the valuation method through which that health state is valued is biased; that will flow through into the DCE valuations.

Of the studies that have used exclusively DCE data, a common approach is to anchor health states by dividing the coefficients associated with health state dimension movements away from full health by the coefficient on a dead state [12]. For example, in the EQ-5D-3L, health state 22222 would be anchored by summing all the level 2 coefficients, and dividing by the coefficient on dead, then subtracting from one. If the sum of the level 2 coefficients exceeds the coefficient on the dead state, this produces a negative score, so 22222 would be considered as worse than being dead. However, this approach does not easily translate to a DCE where duration is explicitly incorporated in the experiment and modelled in the regression; this is discussed further in the description of approach 3. While including duration as an attribute in the DCE causes this issue, its inclusion has been suggested as appropriate for two key reasons [13]. First, if duration is not included, then respondents may infer something about life expectancy when considering a health profile, and this has been suggested by earlier qualitative work presenting DCE tasks using EQ-5D-5L [14]. For example, if the overall health state is severe, it is plausible that a respondent to a valuation task would infer this means a short life expectancy. For this to be controlled, it is possible to give the same duration in all options within all choice sets. Indeed, this approach allows the analyst to test if QALY weights differ for differing durations. The second reason for including duration is that the trade-off between duration and quality of life is central to the QALY model. To ensure the utility score for a particular health state reflects that trade-off, it is essential for the survey respondent to be asked to consider

trading off between not only the different dimensions of the instrument, but between quality of life and length of life.

A second approach to anchoring is simply to impose the zero condition and re-scale the resulting utility function using an affine transformation such that the utility of full health at any survival duration is the same as the duration (effectively imposing the 0–1 scale on death to full health). This approach has been used in the two previous studies that provided the data that are re-analysed in the current study [3, 6]. In this approach [denoted in this paper as a “DCE(Duration)”], the choices analysed are health profiles described in terms of a health state and an explicit and varying life expectancy. The approach in these studies represents one solution to the problem of anchoring, but this approach may be sub-optimal if it mispredicts respondents preferring being dead to particular combinations of health and life expectancy. Possible evidence for misprediction may be taken from the utility values derived, where a large percentage of the modelled values are negative, or worse than dead.

Another possible option for anchoring is to include dead (or immediate death) as a health state within the DCE; however, this has proven controversial in a DCE setting [15]. Flynn et al. [15] argues that it is problematic to consider death within a RUT framework. Flynn et al.’s argument is that there is a set of respondents who will never acknowledge a health state to be less preferred than immediate death. Of this group, some may just not see a health state they believe to be worse than immediate death. Others, however, might believe that it is “not for humans to decide that death is preferable to a living state, no matter how bad it is.” (p. 3).

If this is the case, Flynn et al. argues that these people violate RUT, under which there is always a non-zero probability of an individual picking an option in a choice set (and these people will never select death). This violation means they should be excluded from the dataset. Is this a valid criticism of including dead as a health state? It is certainly true that there may be people with these preferences and, importantly and more troublesome from an analysis perspective, that it is difficult to identify whether someone who never selects death is in this group or not. A counter-argument is that this type of lexicographic preference is likely to still exist when people respond to a choice set in which there is no death option. In the context of the widely used EQ-5D-3L, it is possible that a respondent might never pick an option which involves being confined to bed (which is the worst level in the Mobility dimension). It is uncertain if Flynn et al.’s position extends to excluding these people from further analysis; however, it is logically difficult to assert this is a different type of lexicographic preference. It can only be asserted that it may be less likely than a refusal to prefer death over some non-death profile.

Therefore, in the analysis to follow, we consider models which both do and do not utilise the dead state, but acknowledge Flynn et al.’s concern.

The primary aim of the study is to test four approaches to anchoring the DCE data such that dead is valued at zero and full health at one. The approaches are compared in terms of consistency with the QALY model, prediction of preferences around worse than death states, consistency in results across methods and in terms of ranking of health states.

2 Methods

The data used in this paper are from two published DCE studies in Australia [3, 6], for the EQ-5D-5L and the EQ-5D-3L, respectively. The details of the studies and the administration of the DCEs are reported elsewhere, but were essentially very similar. Both were administered online in samples of respondents recruited via existing large national online panels, with respondents receiving a modest payment from the panel provider for participation. Evidence suggests online administration of these tasks is similar to computer-assisted personal interviews (CAPIs) in terms of results [16]. The samples in both the EQ-5D-3L and the EQ-5D-5L cases were designed to be population representative. The 3L and 5L studies employed designs of different sizes, with 1620 and 200 choice sets, both constructed using a shift generator approach [17]. The sample sizes were similar between surveys, with 973 respondents answering up to ten choice sets each in the EQ-5D-5L and 1031 answering up to 15 in the EQ-5D-3L.

Both studies presented very similarly structured tasks. The choice task for each study presented two health profiles and a death option. Screenshots of the choice task for the studies are presented in “Appendix”. There are small differences in how each experiment presented the choice, but the fundamental structure was unchanged. Each choice set was presented as a triple, with options A and B representing combinations of a health state described within the utility instrument and a time spent in that health state (which is followed immediately by death). Option C, which was in all instances immediate death, was reported at the right-hand side of each choice task. In each task, the respondent was asked to select which of the three options was the best and which was the worst. Thus, each choice set provided a complete ranking over the three options. While each study collected data for preferences between health profiles and death, these data were not used when constructing the utility algorithm that is presented in the published studies. The focus of analysis for those papers was on the choice between the two non-immediate death profiles. This approach imposes the independence of irrelevant alternatives (IIA) assumption such that, if (for

example) option A is preferred to B in a choice set including death, then that profile relative preference would be reported if death were removed from the choice set leaving just the pair of A and B.

Equation 5 sets out the broad approach to the utility specification to allow for estimation of QALY weights. The utility of alternative j in scenario s for individual i is:

$$U_{isj} = \alpha TIME + \beta X'_{isj} TIME + \varepsilon_{isj} \quad (5)$$

In this approach, the marginal rate of substitution for each of the levels of each of the dimensions of each of the utility instruments (other than level 1, which was set as the base in the regression) was estimated using *TIME* (the life expectancy variable) as the numeraire. All data were analysed using a conditional logit. From Eq. 5, the marginal utility of *TIME* is:

$$\frac{\delta U}{\delta TIME} = \alpha + \beta X' \quad (6)$$

To generate QALY weights for health states, we estimate the ratio of the marginal utility of *TIME* for the health state being valued and the marginal utility of *TIME* for full health. Under utility function 5, the $\beta X'$ term drops out of the denominator as full health is the omitted level in each dimension and therefore each X term is zero, meaning the QALY weight for a health state is:

$$\frac{\alpha + \beta X'}{\alpha} \quad (7)$$

Beyond this basic framework, there are a number of decisions in the modelling process that have to be made, and different combinations of these are described below as the four approaches. Each is described in turn.

2.1 Approach 1: The DCE(Duration) Approach

This approach is similar to that taken in the published studies using these data [3, 6], and is often referred to as the DCE_{TTO} approach. In these analyses, the preferences with regard to the dead state are ignored, and we only use the preference between health profiles A and B. As described previously, we assume IIA to identify the relative preference of the respondent between the two non-dead options. Following the general schema outlined above, we impose the QALY framework on the data (although the constant proportional trade-off has been considered and rejected in sensitivity analyses in the EQ-5D-3L paper).

2.2 Approach 2: Treating Dead States as Zero Duration

A limitation of approach 1 is that we are only using half of the stated preference of the respondent. In most cases, we

are discarding the “worst” choice as we can infer whether $A > B$ or $B > A$ when we know which of A, B, or “dead” is best. The exception is when “dead” is best, in which case, we are only considering which of A and B is worst, assuming the respondent would select the other in a head-to-head choice. In terms of amending Eq. 5 to include death, a variety of possible approaches might be taken. First, in approach 2, death is considered as a profile with a duration of zero (thus leaving Eq. 5 unchanged). Since *TIME* is a component of both terms in the systematic component of the utility function presented in Eq. 5, the quality-of-life terms are irrelevant, leaving simply the error term. This approach is attractive as it continues to impose the QALY structure on the data. However, it is based on the assumption that people do consider the dead health state as equivalent to a health profile with a duration of zero, something which may not be the case. If the dead state is worse (better) than would be predicted by approach 2, the resulting utility algorithm would overestimate (underestimate) the proportion of health states valued less than zero.

2.3 Approach 3: Explicitly Modelling Dead States

A further alternative approach is to explicitly model death as part of the regression, in the following way:

$$U_{isj} = \phi DEATH + \alpha TIME + \beta X'_{isj} TIME + \varepsilon_{isj} \quad (8)$$

This is similar to the approach used by Ramos-Goñi et al. [12], but complicated by the inclusion of an explicit duration term. Under this approach, we introduce the potential to move away from the QALY framework. Under the QALY framework, Φ would be equal to zero (due to the zero condition, effectively the model presented in approach 2). The question is how to use these data in the QALY model if Φ is statistically significantly different from zero. It is possible to compare the utility of additional time in some health state (as denoted in Eq. 8) with the disutility of death. However, unlike the Ramos-Goñi et al. approach which did not have duration explicitly modelled, a non-zero coefficient on *DEATH* implies a non-linear utility function with respect to time. As *TIME* tends to zero within some health state, the $\alpha TIME + \beta X'_{isj} TIME$ term also tends to zero. However, when time reaches zero, the Φ coefficient is included in the utility function. This suggests a specific utility (or disutility) for death, which necessarily implies moving away from the QALY framework.

One solution is to consider values for a particular duration, say 10 years. In this case, the value of moving from dead to full health for those 10 years is $10\alpha - \Phi$ (which will be greater than 10α since Φ is negative). The

beta terms can then be scaled according to their magnitude relative to the magnitude of that figure. As the betas are interacted with duration, we have to multiply the coefficient on each beta by ten. To be explicit, the utility decrement associated with (for example) Mobility level 2 is $10\beta_{MO2}/(10\alpha - \phi)$. However, this means the utility decrement will be dependent on *TIME*, something not true in the other approaches considered here. In the re-analysis of the existing data, we present the utility decrement for a period of 10 years. Assuming Φ is negative, the impact of moving to shorter (longer) durations will be to reduce (increase) the absolute size of the utility decrements.

A point to note under this approach is that Φ relates to the immediacy of death, rather than death being in the profile at all. All options in all choice sets include death, but only in option C is death immediate. While this may be seen as placing an undue emphasis on immediacy, if people do place a particular value on avoiding this immediacy, then it is of value to explore how such preferences might be included within QALY calculations.

2.4 Approach 4: Using Dead Preferences for Scaling

An alternative approach to using the dead state preferences is to exclude them from the regression, then to use them to anchor the results of other regressions on to the required 0–1 scale. In the context of the TTO, Rand-Hendriksen et al. [18] note the value of information regarding whether respondents place a health state better, worse, or equal to dead, and assert that the TTO procedure does not provide substantial additional information. While we do not pursue analysis of DCE data using simply the preference between dead and health states (although the data would allow such an analysis), approach 4 uses the dead preferences to anchor regression results from approach 1.

Thus, the regression from approach 1 can be used as a base analysis. Using this algorithm, each health state within each instrument can be valued using the approach described above. Then a supplementary regression can be undertaken identifying the utility value at which health states are no longer preferred to dead by the mean respondent.

$$P(HS_1 > DEAD) = f(U(HS)) + \varepsilon \quad (9)$$

Specifically, our analysis undertook a logit using each comparison between a non-dead state and a dead state. The dependent variable in this supplementary regression is whether an individual believed the health state to be better than dead, and the independent variables are the dimensions of the instrument (so, for instance, ten parameters for the EQ-5D-3L and 20 for the 5L). For simplicity, we assume the mutual utility independence between survival duration and quality of life, thus do not include duration in the independent variables. Using these regression results, we

can identify the relative position of health states and the dead health state. A linear transformation can then be applied to all health states in the utility algorithm derived in approach 1 to match the data with the dead state preferences. This is done by setting the utility of the health state with a 50 % chance of being preferred to dead as zero.

The advantage of this type of approach is that it uses all of the data, it predicts preferences with respect to a dead health state, and it produces a value set independent of duration, something which may diverge from true preferences at an individual level [19], but is likely to be necessary for economic evaluation of health technologies.

2.5 Empirical Specification

Before we present the results under each of the four approaches, one point concerning the specification of the utility function should be noted. The published studies have slightly different terms in the utility algorithm. The EQ-5D-5L is main effects only, while the EQ-5D-3L study employs two-factor interactions between each pairwise combination of dimensions at the worst level ($MO3 \times SC3$, $MO3 \times UA3$, etc.). For this secondary analysis, we have estimated results to allow main effects only in the utility algorithm to ensure comparability across studies.

In addition to the presentation of the utility decrements associated with movements away from the least severe level in each dimension, we also estimated Spearman rank coefficients comparing the similarity or difference in the rank order of the health states in both of the health state instruments under each of the four approaches. To be explicit, this meant that, for each instrument, we estimated six Spearman rank coefficients (approach 1 vs. approach 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 2 vs. 4, and 3 vs. 4).

3 Results

The regressions for the three secondary analyses of the EQ-5D-3L and the EQ-5D-5L are presented in Tables 1 and 2, respectively, which also include the minimum estimated value for states 33333 (EQ-5D-3L) and 55555 (EQ-5D-5L) and the percentage of states modelled as worse than dead for each approach, which provides information about the relative position of the health state equivalent to dead, or zero, across the modelling approaches. Figures 1 and 2 show the corresponding utility algorithm under each of the approaches for the three instruments.

3.1 The EQ-5D-3L

For all analyses of the 3L, Mobility has the largest coefficient decrement, followed by pain/discomfort, anxiety/

Table 1 Regression and utility algorithm (EQ-5D-3L)

	Approach 1		Approach 2		Approach 3 (<i>TIME</i> = 10 years)		Approach 4
	Coefficient (s.e.)	Utility decrement	Coefficient (s.e.)	Utility decrement	Coefficient (s.e.)	Utility decrement	Utility decrement
Duration	0.271 (0.007)***		0.283 (0.005)***		0.218 (0.005)***		
MO2 × dur	−0.033 (0.004)***	−0.122	−0.03 (0.003)***	−0.111	−0.029 (0.003)***	−0.097	−0.086
MO3 × dur	−0.142 (0.004)***	−0.524	−0.12 (0.003)***	−0.444	−0.116 (0.003)***	−0.387	−0.371
SC2 × dur	−0.032 (0.005)***	−0.119	−0.027 (0.003)***	−0.099	−0.026 (0.003)***	−0.087	−0.084
SC3 × dur	−0.077 (0.005)***	−0.286	−0.065 (0.003)***	−0.241	−0.062 (0.003)***	−0.207	−0.202
UA2 × dur	−0.026 (0.005)***	−0.095	−0.023 (0.003)***	−0.085	−0.022 (0.003)***	−0.073	−0.067
UA3 × dur	−0.052 (0.005)***	−0.191	−0.046 (0.003)***	−0.169	−0.043 (0.003)***	−0.143	−0.135
PD2 × dur	−0.030 (0.004)***	−0.110	−0.027 (0.003)***	−0.099	−0.026 (0.003)***	−0.087	−0.078
PD3 × dur	−0.134 (0.004)***	−0.496	−0.119 (0.003)***	−0.438	−0.114 (0.003)***	−0.380	−0.351
AD2 × dur	−0.037 (0.004)***	−0.136	−0.032 (0.003)***	−0.117	−0.029 (0.003)***	−0.097	−0.096
AD3 × dur	−0.100 (0.004)***	−0.371	−0.088 (0.003)***	−0.325	−0.084 (0.003)***	−0.280	−0.263
Death					−0.818 (0.016)***		
Mean		0.183		0.291		0.387	0.422
(minimum)		(−0.868)		(−0.617)		(−0.397)	(−0.322)
P(WTD)		0.305		0.185		0.095	0.062

In all tables, statistical significance is noted at the 1 % level (***)

AD anxiety/depression, *dur* duration, *MO* mobility, *PD* pain/discomfort, *SC* Self-Care, *s.e.* standard error, *TIME* the life expectancy variable, *UA* Usual Activities, *P(WTD)* proportion of health states valued as being worse than dead (i.e. valued below 0)

depression, self-care, and usual activities. This can be seen as an indicator of the overall importance of the dimension in comparison to the others included. Under approach 1, health state 33333 (the pits state for the EQ-5D-3L) is valued at -0.868 . Under approach 2, introducing the dead data, and treating the dead state as a duration of zero reduces the scale of the utility algorithm. In this, the value for the pits state rises to -0.617 , suggesting that dead is being treated as worse than a duration of zero. It should be noted that the difference between the two algorithms is almost solely one of scale (the relative importance of different levels is almost identical).

If death is explicitly modelled in the utility function (approach 3), we observe a large and negative coefficient on the dead state. As noted in the “Methods” section, if death were simply considered to be equivalent to a duration of zero, this coefficient would be zero. For a duration of 10 years, the scale of scoring in the value set is reduced relative to approaches 1 and 2, with only 9.5% of health states valued as being worse than dead. Under approach 4, in which the preferences between health profiles and the dead state are used to anchor the preferences between the two non-dead states, the decrements in the value set reduce considerably. The value of the pits state reduces in absolute terms to -0.322 . While this last approach does involve a more complicated two-stage procedure, the advantage of it is that it reflects and more accurately predicts the preferences with regard to the dead state. The Spearman

coefficients suggested the rank of health states is almost identical under each of the approaches. All of these coefficients exceeded 0.99.

3.2 The EQ-5D-5L

The EQ-5D-5L data show the same pattern as that described for the 3L. The monotonic structure of the instrument is again reflected, other than small disorderings in the worst levels of Usual Activities and Anxiety/Depression. The impact of modelling the dead state as a duration of zero is the same as for the EQ-5D-3L. For the 5L, the pits state (health state 55555) is valued at -0.728 if preferences around dead are ignored (i.e. approach 1), and -0.432 if dead is included as a duration of zero (i.e. approach 2). The reduction in the scale of the scores within the value set is higher in the 5L than in the 3L. Under approach 3, the scale in the value drops further, with the mean health state valued at over 0.4, with only 69 of the 3125 health states (2 %) valued below zero. The reduction in scale of the value set seen for the EQ-5D-3L is similarly seen in the EQ-5D-5L data under approach 4. In this approach, the worst value across all health states is -0.086 (due to a slight non-monotonicity between levels 4 and 5 of Usual Activities, this is 0.003 lower than the value of the pits state). As with the EQ-5D-3L, the health state valuations under the approaches showed almost perfect agreement in terms of the rank of health states (all coefficients again exceeded 0.99).

Table 2 Regression and utility algorithm (EQ-5D-5L)

	Approach 1		Approach 2		Approach 3 (<i>TIME</i> = 10 years)		Approach 4
	Coefficient (s.e.)	Utility decrement	Coefficient (s.e.)	Utility decrement	Coefficient (s.e.)	Utility decrement	Utility decrement
Duration	0.227 (0.011)***		0.246 (0.008)***		0.167 (0.008)***		
MO2 × dur	-0.016 (0.005)***	-0.072	-0.015 (0.004)***	-0.064	-0.012 (0.004)***	-0.047	-0.046
MO3 × dur	-0.022 (0.005)***	-0.098	-0.020 (0.004)***	-0.089	-0.019 (0.004)***	-0.073	-0.063
MO4 × dur	-0.066 (0.005)***	-0.290	-0.052 (0.004)***	-0.227	-0.049 (0.004)***	-0.187	-0.185
MO5 × dur	-0.077 (0.006)***	-0.340	-0.059 (0.004)***	-0.261	-0.058 (0.004)***	-0.221	-0.217
SC2 × dur	-0.018 (0.006)***	-0.080	-0.014 (0.004)***	-0.063	-0.013 (0.004)***	-0.051	-0.051
SC3 × dur	-0.023 (0.005)***	-0.102	-0.016 (0.004)***	-0.071	-0.016 (0.004)***	-0.061	-0.065
SC4 × dur	-0.054 (0.006)***	-0.239	-0.042 (0.004)***	-0.183	-0.040 (0.004)***	-0.154	-0.153
SC5 × dur	-0.076 (0.006)***	-0.334	-0.061 (0.004)***	-0.268	-0.059 (0.004)***	-0.227	-0.213
UA2 × dur	-0.028 (0.006)***	-0.124	-0.019 (0.004)***	-0.084	-0.019 (0.004)***	-0.071	-0.079
UA3 × dur	-0.031 (0.005)***	-0.134	-0.021 (0.004)***	-0.094	-0.020 (0.004)***	-0.076	-0.086
UA4 × dur	-0.071 (0.006)***	-0.312	-0.052 (0.004)***	-0.227	-0.050 (0.004)***	-0.191	-0.199
UA5 × dur	-0.070 (0.006)***	-0.306	-0.050 (0.004)***	-0.218	-0.049 (0.004)***	-0.189	-0.195
PD2 × dur	-0.018 (0.006)***	-0.081	-0.011 (0.004)***	-0.050	-0.011 (0.004)***	-0.041	-0.052
PD3 × dur	-0.021 (0.005)***	-0.094	-0.014 (0.004)***	-0.061	-0.013 (0.004)***	-0.049	-0.06
PD4 × dur	-0.060 (0.006)***	-0.265	-0.050 (0.004)***	-0.220	-0.048 (0.004)***	-0.185	-0.169
PD5 × dur	-0.079 (0.006)***	-0.346	-0.067 (0.004)***	-0.293	-0.064 (0.004)***	-0.246	-0.221
AD2 × dur	-0.032 (0.006)***	-0.140	-0.032 (0.004)***	-0.141	-0.029 (0.004)***	-0.110	-0.089
AD3 × dur	-0.054 (0.005)***	-0.237	-0.047 (0.004)***	-0.207	-0.044 (0.004)***	-0.169	-0.151
AD4 × dur	-0.097 (0.006)***	-0.425	-0.087 (0.004)***	-0.383	-0.083 (0.004)***	-0.317	-0.271
AD5 × dur	-0.091 (0.006)***	-0.402	-0.080 (0.004)***	-0.351	-0.077 (0.004)***	-0.294	-0.257
Death					-0.945 (0.024)***		
Mean		0.116		0.289		0.408	0.436
(minimum)		(-0.757)		(-0.432)		(-0.202)	(-0.121)
P(WTD)		0.345		0.118		0.022	0.009

In all tables, statistical significance is noted at the 1% level (***)

AD anxiety/depression, *dur* duration, *MO* mobility, *PD* pain/discomfort, *SC* self-care, *s.e.* standard error, *TIME* the life expectancy variable, *UA* usual activities, *P(WTD)* proportion of health states valued as being worse than dead (i.e. valued below 0)

4 Discussion

This paper uses existing datasets to explore different ways of anchoring DCE-derived utility weights. Choice sets in both datasets include both combinations of health states and duration, and dead states. The inclusion of duration allows exploration of trade-offs between length of life and quality of life; however, it introduces a complicating factor if death is to be specifically incorporated in the model (as in the third approach presented in this analysis).

The pattern in results between the four approaches presented here is consistent. Approach 1, which has been most commonly used in the DCE_{TTO} literature to date, leads to the largest range of utility scores in the resultant value set. Employing a utility algorithm with a relatively wide spread of scores will, *ceteris paribus*, cause a

relatively large focus on quality of life rather than life expectancy. Using dead states is potentially controversial [15], but their inclusion provides a means to anchor the utility function on the 0–1 death to full health scale. Without this, it is necessary to impose further assumptions regarding an average respondent's utility function. We have presented three ways of anchoring using dead health states, each of which leads to a different range of utility scores within the value set.

Approach 2, in which the dead state is considered as a duration of zero, has the advantage over the standard approach 1 of using all of the data that were collected in both studies. It better predicts choices with respect to dead, but may not adequately capture what respondents think of when they are asked whether dead is actually preferable to a living state. Approach 3 also uses all of the data, and has

Fig. 1 Utility decrements associated with movements from “No Problems”, by anchoring approach (EQ-5D-3L). *AD* anxiety/depression, *MO* mobility, *PD* pain/discomfort, *SC* self-care, *UA* usual activities

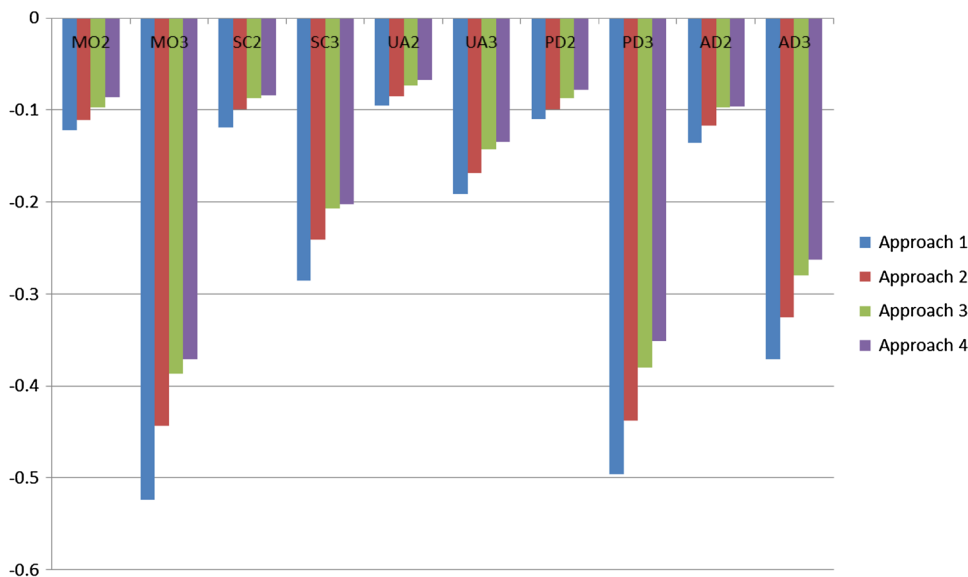
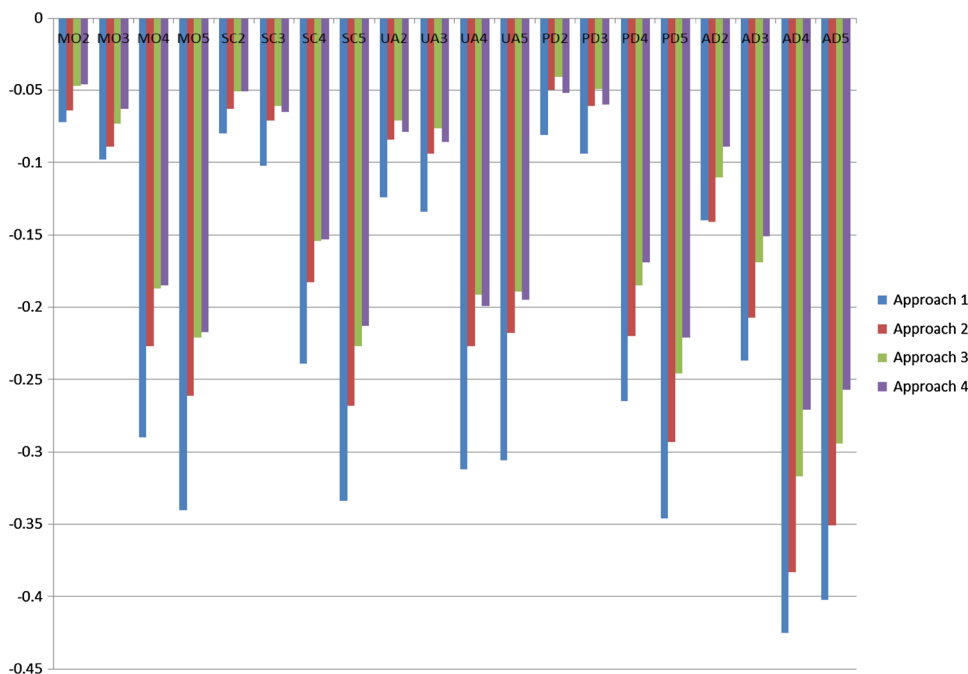


Fig. 2 Utility decrements associated with movements from “No Problems”, by anchoring approach (EQ-5D-5L). *AD* anxiety/depression, *MO* mobility, *PD* pain/discomfort, *SC* self-care, *UA* usual activities



the advantage of explicitly modelling the dead health state. The advantage of this approach is that it is likely to fit the data well. The disadvantage of it lies in issues regarding the construction of a value set for use in economic evaluation. We have not yet identified a solution under this approach which produces a value set that is independent of duration. This is a major issue for economic evaluation, where different value sets for different durations raise a number of very difficult new problems for economic modelling of treatment pathways and interventions. For example, while it may be relatively straightforward under this approach to apply different utility weights to the same chronic health

state under different durations, it is computationally much more challenging to apply these weights to parts of health profiles surrounded by other health states, and further imposes additive separability of duration in the utility function. For example, if a hypothetical individual is in an acute health state for a small proportion of the time they spend in an economic model, do we apply the value for that short duration which may be the consequence of the person doing the valuation believing the state refers to a close to dead health state? Given the intractability of these problems, it is unlikely that this approach would be helpful for calculation of QALYs in economic models.

Approach 3 includes dead as a health state. The assumption that dead can be valued as if it were a health state is questionable, and indeed our results suggest that quantitatively, respondents did not treat dead as simply being a health state with a duration of zero. The question of how respondents consider the health state of being dead is one which is likely to be better addressed in a qualitative setting. We are unaware of such work, and believe it would be a useful extension of the quantitative analysis presented here.

Approach 4 is a hybrid approach. It uses a two-stage approach, in which the choices over the health state profiles are used on their own to produce an ordinal scale with interval properties, and then the choices between health profiles and death are used to anchor this scale in the second stage. Thus, it is likely to predict preferences regarding health states being better or worse than dead, uses all of the data, and produces a single value set for economic evaluation.

It should be noted that the minimum values under each of the four approaches are below the minimum value in the corresponding Australian EQ-5D-3L TTO weights [19]. While comparisons across methods are problematic, the worst health state (33333) is valued at -0.217 in the TTO-derived value set. The impact of this on economic evaluation is potentially large. Value sets with a more disperse range will relatively value quality of life over life expectancy. This applies to all of these DCE-derived weights relative to the pre-existing TTO weights. A valuable avenue for future research is the exploration of the relative

predictive power of TTO- and DCE-derived algorithms to identify their capacity for predicting preferences between health states over the ranges of severity and duration.

The approaches considered here are by no means an exhaustive set that might be taken to anchor DCE with duration data to provide QALY weights. What our analysis does show is that the range of the scores in a value set is highly dependent on the method of anchoring, and that this has the potential to influence the conclusions from economic evaluations using these scores. If we are to use this type of DCE in future (i.e. with duration and a dead health state), we recommend further work to identify a preferred method, or suite of methods, for anchoring on the 0–1 scale.

Compliance with Ethical Standards

Conflict of interest The authors (Norman, Mulhern, and Viney) declare no conflict of interests. Data collection was funded by a National Health and Medical Research Council Project Grant (403303).

Author contributions Norman conceived and undertook the analysis, and was primarily responsible for drafting the manuscript. Mulhern helped to develop the approaches to anchoring explored in the analysis, and commented on and amended the draft manuscript. Viney was responsible for the design and collection of the underpinning data, and commented on and amended the draft manuscript.

Appendix

Screenshots of the tasks in the two studies.

If you had to choose between the following scenarios:

Scenario A	Scenario B	Scenario C
<ul style="list-style-type: none"> You have no problem in walking about You have some problems washing and dressing yourself You have no problems with performing your usual activities You have moderate pain or discomfort You are extremely anxious or depressed 	<ul style="list-style-type: none"> You have some problems in walking about You have no problems with self-care You are unable to perform your usual activities You have extreme pain or discomfort You are not anxious or depressed 	Death
You will live in this state for 4 years, then die.	You will live in this state for 4 years, then die.	

which is best?

which is worst?

next

SurveyENGINE

Consider the following options. If you had to choose between them, which of the three is the best, and which is the worst?

	State 1	State 2	Immediate death
Mobility	You have slight problems in walking about	You have no problems in walking about	
Self-Care	You have no problem with washing or dressing yourself	You have slight problems with washing or dressing yourself	
Usual Activities (e.g. work, study, housework, family or leisure activities)	You are unable to do your usual activities	You have severe problems doing your usual activities	
Pain / Discomfort	You have severe pain or discomfort	You have extreme pain or discomfort	
Anxiety / Depression	You are moderately anxious or depressed	You are slightly anxious or depressed	
You will live in this health state for this period of time, then die	4 years	8 years	
Of these three options, which is the best?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Of these three options, which is the worst?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

45% complete

prev next

© 2011 SurveyEngine P/L

References

- de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ.* 2012;21(2):145–72.
- Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate societal health state utility values. *J Health Econ.* 2012;31:306–18. doi:10.1016/j.jhealeco.2011.11.004.
- 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(3):287–98.
- Norman R, Viney R, Brazier J, Burgess L, Cronin P, King MT, et al. Valuing SF-6D health states using a discrete choice experiment. *Med Decis Making.* 2014;34(6):773–86.
- Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health.* 2010;13(8):1005–13.
- Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. An Australian discrete choice experiment to value eq-5d health States. *Health Econ.* 2014;23(6):729–42. doi:10.1002/hec.2953.
- Bleichrodt N, Wakker P, Johannesson M. Characterizing QALYs by risk neutrality. *J Risk Uncertain.* 1997;15(2):107–14.
- McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, editor. *Frontiers in econometrics.* New York: New York Academic Press; 1974. p. 105–42.
- Thurstone LL. A law of comparative judgment. *Psychol Rev.* 1927;34:273–86.
- Devlin NJ, Krabbe PF. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ.* 2013;14(Suppl 1):S1–3. doi:10.1007/s10198-013-0502-3.
- Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–53. doi:10.1016/j.jval.2014.04.002.
- Ramos-Goñi JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabases JM. Dealing with the health state ‘dead’ when using discrete choice experiments to obtain values for EQ-5D-5L health states. *Eur J Health Econ HEPAC Health Econ Prev Care.* 2013;14(Suppl 1):S33–42. doi:10.1007/s10198-013-0511-2.
- Flynn T. Using conjoint analysis to estimate health state values for cost-utility analysis: issues to consider. *Pharmacoeconomics.* 2010;28(9):711–22.
- Mulhern B, Bansback N, Brazier J, Buckingham K, Cairns J, Devlin N et al. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technol Assess.* 2014;18(12):vii–xxvi, 1–191. doi:10.3310/hta18120.
- Flynn TN, Louviere JJ, Marley AA, Coast J, Peters TJ. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Popul Health Metrics.* 2008;6:6.
- Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value Health.* 2013;16(1):104–13. doi:10.1016/j.jval.2012.09.001.
- Street DJ, Burgess L. The construction of optimal stated choice experiments: theory and methods. *Wiley series in probability and statistics.* New Jersey: Wiley; 2007.
- Rand-Hendriksen K, Augestad LA, Dahl FA, Kristiansen IS, Stavem K. A shortcut to mean-based time tradeoff tariffs for the EQ-5D? *Med Decis Making.* 2012;32(4):569–77. doi:10.1177/0272989X11431607.
- 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(4):460–7. doi:10.1177/0272989X05276854.