CHERE

CENTRE FOR HEALTH ECONOMICS RESEARCH AND EVALUATION

UNIVERSITY OF TECHNOLOGY SYDNEY

UTS:CHERE

Deriving utility weights for the EQ-5D-5L using a discrete choice experiment

Working Paper 2012/01

January 2012

A report by the Centre for Health Economics Research and Evaluation

About CHERE

CHERE is an independent research unit affiliated with the University of Technology, Sydney. It has been established since 1991, and in that time has developed a strong reputation for excellence in research and teaching in health economics and public health and for providing timely and high quality policy advice and support. Its research program is policy-relevant and concerned with issues at the forefront of the sub-discipline.

CHERE has extensive experience in evaluating health services and programs, and in assessing the effectiveness of policy initiatives. The Centre provides policy support to all levels of the health care system, through both formal and informal involvement in working parties, committees, and by undertaking commissioned projects. For further details on our work, see www.chere.uts.edu.au.

Project team

Richard Norman
Paula Cronin
Rosalie Viney

Centre for Health Economics Research and Evaluation (CHERE)
University of Technology Sydney, Broadway 2007 New South Wales.

Contact details

Mr Richard Norman
Centre for Health Economics Research and Evaluation (CHERE)
University of Technology, Sydney
City Campus
PO Box 123 Broadway
NSW 2007

Tel: + 61 2 9514 4732
Fax: + 61 2 9514 4730
Email: richard.norman@chere.uts.edu.au

Table of Contents

Abstract

Purpose: To estimate an Australian algorithm for the newly developed 5-level version of the EQ-5D, for use in the economic evaluation of health and healthcare interventions.

Methods: A discrete choice experiment (DCE) was run in an online Australian-representative sample. A random-effects probit model was estimated, and converted to a zero to one scale for use in economic evaluation.

Results: At least one choice set was completed by 944 respondents, of which 932 completed all ten choice sets. The mean and median completion times were 17.9 and 9.4 minutes respectively, demonstrating a highly skewed pattern. Respondents were slightly younger and better-educated than the general Australian population. The regression results broadly reflect the monotonic nature of the EQ-5D-5L. Utility increases in life expectancy, and decreases in higher levels in each dimension of the instrument. A high proportion of respondents found the task clear and relatively easy to complete.

Conclusions: DCEs are a valuable approach in the estimation of utility weights for multi-attribute utility instruments such as the EQ-5D-5L.

Introduction

The EQ-5D-3L (conventionally termed the EQ-5D) is the most widely used multi-attribute utility instrument internationally for use in economic evaluation in healthcare (Richardson, et al., 2011). It was developed by the Euroqol group (www.euroqol.org) as a relatively simple instrument that could be used in clinical studies, and provides valuations of health states for use in economic evaluation. The EQ-5D-3L has five dimensions, intended to represent the major areas in which health changes can manifest: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Within the EQ-5D-3L, each dimension contains three levels, broadly classified as 'No Problems', 'Some Problems', and 'Extreme Problems'. Thus, there are $3^5 = 243$ health states defined within the EQ-5D-3L. For the purposes of economic evaluation (and the construction of the quality-adjusted life year (QALY)), it is necessary to place all 243 health states on a scale where full health is valued at 1, and death is valued at 0.

A major criticism of the EQ-5D-3L is that is highly insensitive to small changes in quality of life. For example, having 'some problems with mobility' encompasses a large range of possible levels of mobility, meaning an improvement within that range, which may represent a large and clinically relevant improvement in mobility, is not captured. A recent study, supported by the Euroqol group, has proposed a new 5-level version of the EQ-5D (Herdman, et al., 2011). This instrument, termed the EQ-5D-5L, contains the same five dimensions, but amends the wording of a number of the level descriptors and adds two additional levels to each dimension. The EQ-5D-5L is reproduced in Table 1.

Table 1: The EQ-5D-5L

Dimension	Level	
Mobility	1	I have no problems in walking about
	2	I have slight problems in walking about
	3	I have moderate problems in walking about
	4	I have severe problems in walking about
	5	I am unable to walk about
Self Care	1	I have no problems washing or dressing myself
	2	I have slight problems washing or dressing myself
	3	I have moderate problems washing or dressing myself
	4	I have severe problems washing or dressing myself
	5	I am unable to wash or dress myself
Usual Activities	1	I have no problems doing my usual activities
	2	I have slight problems doing my usual activities
	3	I have moderate problems doing my usual activities
	4	I have severe problems doing my usual activities
	5	I am unable to do my usual activities
Pain / Discomfort	1	I have no pain or discomfort
	2	I have slight pain or discomfort
	3	I have moderate pain or discomfort
	4	I have severe pain or discomfort
	5	I have extreme pain or discomfort
Anxiety / Depression	1	I am not anxious or depressed
	2	I am slightly anxious or depressed
	3	I am moderately anxious or depressed
	4	I am severely anxious or depressed
	5	I am extremely anxious or depressed

While this is likely to improve the sensitivity relative to the original 3-level version, a major issue is that the many existing valuation studies become largely redundant. Thus, there is significant interest in producing algorithms to value the 5^5 = 3,125 health states contained in the EQ-5D-5L. It is noteworthy that the EQ-5D-5L is likely to be monotonic, in the sense that moving up levels within each dimension (e.g. level 1 to level 2) is expected to make a health profile worse.

For the EQ-5D-3L, a Time Trade-Off (TTO) approach (or, more rarely, a Visual Analogue Scale (VAS)) is used to value a selection of these 243 states, and then to impute values for the remainder using regression (Norman, et al., 2009). The use of TTO for valuing EQ-5D states is well described elsewhere (Dolan, 1997, Tsuchiya, et al., 2002). Briefly, for states considered to be preferable to immediate death, a respondent is faced with a choice between ten years of a particular chronic health state defined in EQ-5D space and a period of *x* years in full health. The aim of the TTO is to identify a value of *x* for which the individual is indifferent between the options. The value for the better than death health state is defined as *x* / 10. For states considered worse than death, an alternative choice is presented, which has been well described elsewhere (Tilling, et al., 2010). An important motivator for this paper is that health states better and worse than dead require different choice tasks in the TTO, and there is uncertainty regarding the comparability of weights from the two tasks. A lead-time TTO approach has been proposed, but this does not remove the problem for extremely poor health states (Robinson and Spencer, 2006).

The use of the Time Trade-Off more generally has been criticised for a number of reasons, including the introduction of considerable and uncontrollable bias, and the cognitive difficulty in responding to it (Brazier, et al., 2007, Norman, et al., 2009). Recent studies have considered using Discrete Choice Experiments (DCEs) as an alternative valuation technique (Bansback, et al., 2011, Stolk, et al., 2010, Viney, et al., 2011). The advantages of this approach include that it does not require different questions for health states better and worse than death, it is relatively straightforward to administer electronically, and respondents do not need to express strength of preference (only the relative preference between two options). Additionally, there is a strong and growing literature base outlining the appropriate strategy for the design of these experiments (Street and Burgess, 2007).

In DCEs, the utility of an alternative *i* in a choice set C_n to an individual *n* is given by

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

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

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

Alternative *i* is chosen if and only if

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

which can be rearranged to yield

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

Neither the systematic utility nor the error terms are directly observed. Therefore, analysis is reliant on observing choices and inferring the terms from that. Random Utility Theory (RUT) is the dominant approach to doing this. In RUT, it is assumed 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 (McFadden, 1974, Thurstone, 1927).

The aim of this study was to apply the DCE methodology to the new EQ-5D-5L. The main purpose of this was to demonstrate the feasibility of generating an algorithm which could be of use in economic evaluations using the EQ-5D-5L. In particular, we were interested in whether respondents found the task difficult or unclear, and whether the results were reflective of the monotonic nature of the EQ-5D-5L. The paper is structured as follows. Firstly, we outline how the experiment was designed and implemented online (including sample frame specification). Then, we discuss the econometric techniques for the analysis of the data, and the approach taken to convert these into QALY weights. After presentation of results, we outline some limitations with the approach, and suggest areas in which the approach taken in this study might be extended.

Methods

A discrete choice experiment was developed and administered to a sample of the Australian general population. Respondents were asked to choose between health profiles described in terms of EQ-5D-5L profiles and survival attributes. The health profiles presented could be broadly considered to be similar to those presented in a time trade-off task in that they describe a chronic health state that would be experienced with certainty for a number of years followed by death. The quality of life in each profile was described in terms of an EQ-5D-5L health state, and the options included only quality of life and survival.

The choice task

Each choice set included three options: two health profile options and an immediate death option. Each health profile option in a choice set was defined by five attributes covering the dimensions of the EQ-5D-5L and a survival duration attribute. Four survival durations of 2, 4, 8 and 16 years were included in the experiment.

The third option of immediate death was included to allow for a complete ranking of health profiles over the “worse than death” to “full health” utility space. The task for the respondent was to identify which of the three options was considered the best, and which the worst, thus providing a complete ranking within each choice set. In the analysis presented here, only the choice between the two EQ-5D-5L health profiles is considered. An example of a choice set is provided in Figure 1.

Figure 1: Example Choice Set

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?	○	○	○
Of these three options, which is the worst?	○	○	○

45% complete

prev next

Experimental design

A great strength of DCE methodology is that it allows pre-specification of parameters of interest, which can then be designed for to ensure both precise and unbiased estimates (Street and Burgess, 2007). In the analysis presented below, we were interested in both main effects, and any interaction effects between life expectancy and aspects of quality of life. To estimate these terms, we started with orthogonal main effects plan (OMEP) with five dimensions, each with five levels. This has 25 runs. Each run is then combined with each of 4 durations, specifically 2, 4, 8 and 16 years. To generate the design for the discrete choice experiment, a shifted experiment design was used (Street and Burgess, 2007). The choice of generators is made with the intention of specifying which effects are of interest. For the estimation of the effects of interest in this case, two generators are required, meaning an experimental design of 200 choice sets. This was blocked into 20 equal parts, with each survey respondent answering ten choice sets. Evidence suggests respondents can easily handle 10 of these questions without loss in data quality (Coast, et al., 2006).

A range of demographic questions was collected. This included age, gender, income, country of birth, family composition etc. This is of use for two reasons. Firstly, it is possible to identify if the sample is representative of the Australian population.

Secondly, it is interesting to explore whether different types of people have different patterns of responses.

Sample Recruitment

The respondents were recruited from an existing Australian online panel, administered by Pure Profile Pty Ltd. Respondents were paid by the panel administrators for completion of the survey a small sum dependent on the time they spent answering the choice sets (approximately $15) to complete the survey. Each respondent used a web link to access the survey, so was able to self-complete at their convenience. A thorough description of the task was provided at the beginning of the survey and a help button was available throughout the task. Each respondent was familiarised with the EQ-5D-5L by being asked to describe their own health using the tool, and then completed the task for the 10 choice sets.

Analysis

A common approach to specifying a utility function in the analysis of choice data is to include a main effect for each level of each attribute (with one level in each attribute omitted to avoid over-identification). This approach to the utility function can be extended to consider interactions between levels of attributes. In the case of health gain, considering quantity of life separately from characteristics of those extra years (such as quality of life), and including simple interactions between them is unlikely to be adequate in capturing this inter-relatedness. The reason for this is because it is necessary to impose the zero-condition specified by Bleichrodt *et al.* as being a necessary condition for the QALY model. Specifically, the utility of a health profile with zero life expectancy is zero irrespective of the quality of life in that (non-) period (Bleichrodt, et al., 1997).

The utility function for all models is assumed to be linear with respect to time, <u>and the characteristics of that time enters not as a main effect, but as an interaction with the *TIME* variable</u>. Therefore, the proposed specification of the utility function was that proposed by Bansback *et al.* (2011)). Thus, the utility of alternative *j* in scenario *s* for individual *i* was

$$U_{isj} = \alpha TIME_{isj} + \beta X'_{isj} TIME_{isj} + \nu_i + \varepsilon_{isj} \quad \text{Equation 5}$$

where X'_{isj} was a set of dummy relating to the levels of the EQ-5D health state presented in option j. The error term ($\nu_i + \varepsilon_{isj}$) consisted of a person-specific error term distributed independent and identically distributed (iid) normal and a conventional random error term distributed iid normal, thus the model adopted a random-effects probit. This model imposed the structure of the QALY model on to the data. An important point to note is that the impact of moving away from level 1 of each dimension is investigated through two-factor interaction terms rather than through the main effect.

The marginal utility of time in this approach, which will be used later, is therefore

$$\frac{\delta U_{isj}}{\delta TIME_{isj}} = \alpha + \beta X'_{isj}$$ Equation 6

A major advantage of this approach is that, if TIME is set at zero, the systematic component of the utility function is zero. Thus, the choice between two profiles with TIME set to zero is random irrespective of the levels of the other parameters. In terms of the constraints the QALY model places on individual preferences, it is noteworthy that Equation 1 features both conditions specified by Bleichrodt, Wakker and Johannesson (1997) as being jointly sufficient for the QALY model.

A number of variations on this specification were considered. In Model A, X'_{isj} included only the twenty main effects, these being the movement from level 1 to one of level 2,3,4 or 5 in each of the 5 dimensions of the EQ-5D-5L. In most EQ-5D-3L algorithms, interactions were captured using the N3 variable; this being a dummy variable equal to one if and only if at least one of the five dimensions is at the worst level. Similarly, we introduced an N5 variable which is equal to one if and only if at least one of the dimensions is at level 5. Therefore, Model B replicates Model A, but includes this N5 variable.

As the EQ-5D-5L is intended to be monotonic, Models C and D repeat Models A and B, but combining levels as necessary to ensure monotonicity. This was done by combining levels with non-monotonic ordering, re-estimating and checking that no additional non-monotonic orderings occurred.

Generating measures of welfare

Using marginal rates of substitution as a method for deriving a welfare measure for the utility associated with a changing attribute is an easy to apply and an intuitive method, and similar to the one proposed in this project. There are a large number of applications of the approach in the health economics literature (Gyrd-Hansen and Sogaard, 2001, McIntosh and Ryan, 2002, Scott, 2001). The *MRS* is calculated by partially differentiating an indirect utility function *V* with respect to one attribute x_1, and then with respect to another attribute x_2, then calculating a ratio, i.e.,

$$MRS_{x_1,x_2} = \frac{\delta V / \delta X_1}{\delta V / \delta X_2}.$$ Equation 7

Thus, the numerator is the marginal utility of X_1, and the denominator is the marginal utility of X_2. Using the ratio puts the marginal utility of X_1 in the units of X_2. If X_2 is a price, the *MRS* represents a marginal willingness to pay for a change in X_1. In a main effects model, this term is usually interpreted as a ratio of coefficients (although Lancsar *et al.* (2007) have shown that Equation 7 is a more general expression which continues to be applicable under different specifications of the utility function).

This can be adapted to the framework established in equation 5. The need for adaptation from the conventional *MRS* approach is that we are interested in the value of a health profile relative to some other health profile (e.g. the value of quality of life in a particular health state relative to full health). Therefore, what is needed is the ratio of marginal utilities (*RMU*) associated with the two profiles. The *RMU* between two alternatives health profiles *j* and *j** can therefore be estimated as

$$RMU_{j,j^*} = \frac{\alpha + \beta X'_{isj}}{\alpha + \beta X'_{isj^*}}. \qquad \text{Equation 8}$$

This would provide a value for a health state with characteristics defined by the $\beta X'_{isj}$ term, relative to another health state defined by characteristics $\beta X'_{isj^*}$. If this latter term is set to full health (i.e. the best level of each dimension in the EQ-5D-5L), the value generated using equation 4 is a QALY weight. This is analogous to the approach taken by Bansback *et al.* (2011) who mock up a TTO to derive QALY weights from the regression results of a similar DCE investigating the EQ-5D-3L.

To estimate confidence intervals around the QALY weights, the *wtp* command in STATA was employed with the default delta method (Hole, 2007), in which a first-order Taylor expansion around the mean value of the variables, and then calculating the variance of the resulting expression. The delta method has been shown to perform well, and to provide similar results to other competing approaches such as the Fieller or the Krinsky Robb methods (Hole, 2007).

Results

The number of individuals who entered the survey was 1,638, of which 665 were screened out as they were outside of the sampling frame. Of the remaining 973 individuals, 944 completed at least one choice, with 932 completing all choice sets. Of the 932 completing all choice sets, 930 completed the demographics section. Individuals completing at least one choice set were included in the analysis set; no attempt was made to exclude on the basis of rationality.

Mean completion time was 17.9 minutes. However, this was heavily skewed data, with 95% of respondents completing within 26.3 minutes. The median completion time was 9.4 minutes. The demographics of the respondents are reported in Table 2. The sample closely matches the Australian population on these key characteristics. Nevertheless, the sample is slightly younger and better educated than the general population.

Table 2: Demographic profile of the sample and the Australian general population

Characteristic	Value / Range	Sample	Population[2]
Gender	Female	53.70%	56.09%
Country of Birth	Born in Australia	79.05%	76.00%
Age (years)	16-29	23.26%	21.33%
	30-44	25.80%	23.98%
	45-59	25.16%	22.40%
	60-74	15.33%	14.00%
	75+	10.46%	18.29%
Highest level of education	Primary/Secondary	41.35%	60.51%
	Trade certificate	34.16%	22.24%
	Bachelor's degree or above	24.49%	17.26%
Gross household income[1]	<$20,000	14.94%	15.77%
	$20,000 - $40,000	27.40%	23.02%
	$40,001 - $60,000	19.42%	17.64%
	$60,001 - $80,000	14.57%	13.87%
	$80,001 - $100,000	9.96%	11.03%
	$100,001 +	13.70%	18.67%

[1] 128 individuals chose to not disclose income
2 Data for the Australian general population sourced from the Australian Bureau of Statistics (Australian Bureau of Statistics, 2002, Australian Bureau of Statistics, 2005, Australian Bureau of Statistics, 2006, Australian Bureau of Statistics, 2007a) (Australian Bureau of Statistics, 2007b)

The self-assessed health of the respondents as defined by the EQ-5D-5L is presented in Table 3. As our sample was general population, it is unsurprising that the modal response in each dimension is Level 1, with the exception of Pain / Discomfort. Equally, the worse levels are sparsely populated, particularly for Self Care.

Table 3: Respondent Self-Assessed Health

	Level 1	Level 2	Level 3	Level 4	Level 5
Mobility	624 (66.0%)	185 (19.6%)	97 (10.3%)	32 (3.4%)	8 (0.9%)
Self Care	823 (87.0%)	83 (8.8%)	28 (3.0%)	11 (1.7%)	1 (0.1%)
Usual Activities	626 (66.2%)	179 (18.9%)	88 (9.3%)	43 (4.6%)	10 (1.1%)
Pain / Discomfort	336 (35.5%)	356 (37.6%)	166 (17.6%)	70 (7.4%)	18 (1.9%)
Anxiety / Depression	507 (53.6%)	249 (26.3%)	141 (14.9%)	32 (3.4%)	17 (1.8%)

Regarding clarity and difficulty of the task, 8.17% of the respondents found the task either unclear or very unclear, and 10.86% of respondents found the task either difficult or very difficult. Overall, 15.70% fell into at least one of these categories.

The results from the random-effects probit models are presented in Table 4.

Table 4: Estimation Results

Mean (SE)				
Model	A	B	C	D
TIME	0.135(0.007)***	0.143(0.007)***	0.135(0.007)***	0.139(0.007)***
TIME x MO2	-0.010(0.003)***	-0.010(0.003)***	-0.010(0.003)***	-0.010(0.003)***
TIME x MO3	-0.014(0.003)***	-0.012(0.003)***	-0.013(0.003)***	-0.013(0.003)***
TIME x MO4	-0.038(0.003)***	-0.039(0.003)***	-0.038(0.003)***	-0.038(0.003)***
TIME x MO5	-0.045(0.003)***	-0.039(0.004)***	-0.045(0.003)***	-0.042(0.004)***
TIME x SC2	-0.009(0.003)***	-0.011(0.004)***	-0.009(0.003)***	-0.010(0.004)***
TIME x SC3	-0.011(0.003)***	-0.012(0.003)***	-0.010(0.003)***	-0.011(0.003)***
TIME x SC4	-0.030(0.003)***	-0.031(0.003)***	-0.030(0.003)***	-0.030(0.003)***
TIME x SC5	-0.044(0.003)***	-0.040(0.004)***	-0.044(0.003)***	-0.042(0.004)***
TIME x UA2	-0.017(0.003)***	-0.016(0.003)***	*-0.017(0.003)****	-0.016(0.003)***
TIME x UA3	-0.017(0.003)***	-0.017(0.003)***	*-0.017(0.003)****	-0.017(0.003)***
TIME x UA4	-0.042(0.003)***	-0.040(0.003)***	*-0.041(0.003)****	*-0.039(0.003)****
TIME x UA5	-0.041(0.003)***	-0.036(0.004)***	*-0.041(0.003)****	*-0.039(0.003)****
TIME x PD2	-0.010(0.003)***	-0.012(0.004)***	-0.010(0.003)***	-0.011(0.004)***
TIME x PD3	-0.012(0.003)***	-0.013(0.003)***	-0.012(0.003)***	-0.012(0.003)***
TIME x PD4	-0.036(0.003)***	-0.037(0.003)***	-0.036(0.003)***	-0.036(0.003)***
TIME x PD5	-0.049(0.003)***	-0.044(0.004)***	-0.049(0.003)***	-0.046(0.004)***
TIME x AD2	-0.019(0.003)***	-0.019(0.003)***	-0.019(0.003)***	-0.019(0.003)***
TIME x AD3	-0.034(0.003)***	-0.033(0.003)***	-0.035(0.003)***	-0.034(0.003)***
TIME x AD4	-0.058(0.003)***	-0.059(0.003)***	*-0.057(0.003)****	*-0.055(0.003)****
TIME x AD5	-0.055(0.003)***	-0.049(0.004)***	*-0.057(0.003)****	*-0.055(0.003)****
TIME x N5		-0.015(0.005)***		-0.008(0.004)*
Constant	0.031(0.019)	0.055(0.021)***	0.03(0.019)	0.044(0.02)**
Log likelihood	-5763	-5758	-5763	-5761
AIC	11571	11564	11567	11567
BIC	11736	11736	11710	11724
Lnsig2u	-1.823	-1.820	-1.822	-1.820
Σu	0.402	0.403	0.402	0.402
ρ	0.139	0.139	0.139	0.139
Degrees of freedom	23	24	20	22

Note: Coefficients in *italics* refer to dimensions that have been combined to ensure monotonicity of results

Model fit was similar between models, with only small divergence between models in terms of log-likelihood or Information Criteria results. We provisionally selected Model D as our preferred algorithm for use in cost-utility analyses. However, given the uncertainty in this decision, the QALY weights associated with each of the models are presented in Table 5.

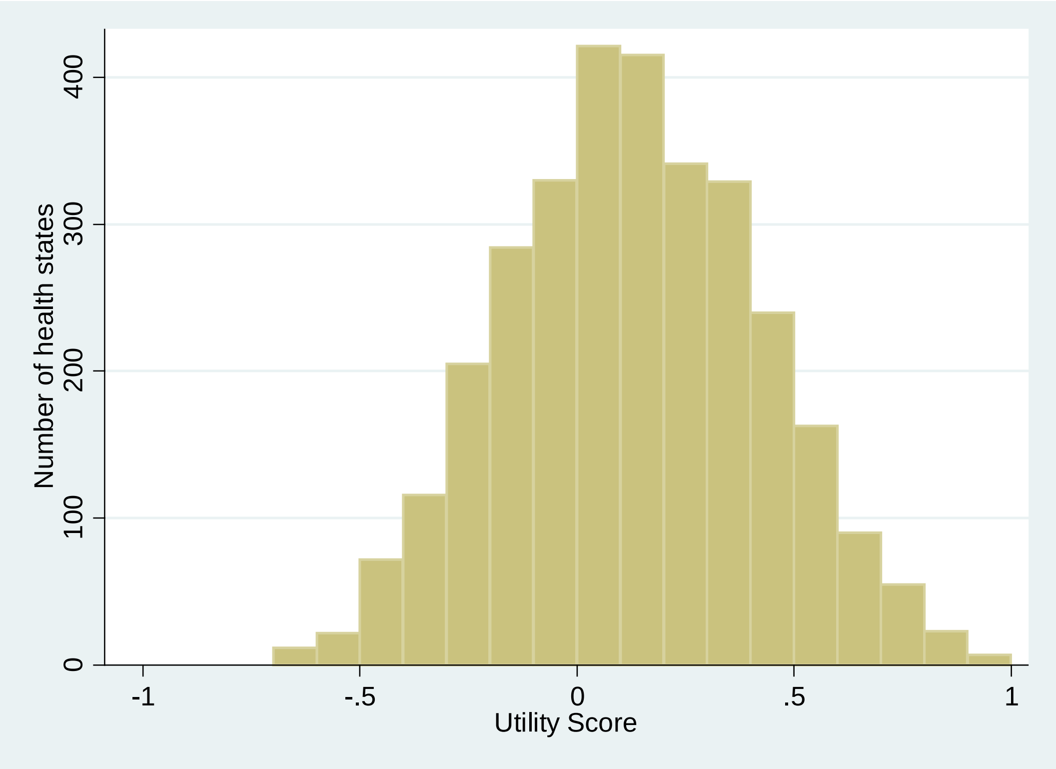
Table 5: Utility weights (confidence intervals) under the four models

	Model A	Model B	Model C	Model D
MO2	0.074 (0.027-0.122)	0.070 (0.025-0.115)	0.075 (0.027-0.122)	0.072 (0.026-0.119)
MO3	0.100 (0.056-0.144)	0.086 (0.043-0.129)	0.100 (0.056-0.144)	0.091 (0.047-0.135)
MO4	0.284 (0.241-0.328)	0.272 (0.230-0.313)	0.284 (0.240-0.327)	0.276 (0.233-0.319)
MO5	0.330 (0.283-0.378)	0.273 (0.217-0.328)	0.332 (0.285-0.380)	0.302 (0.247-0.356)
SC2	0.069 (0.020-0.117)	0.080 (0.034-0.125)	0.067 (0.019-0.115)	0.072 (0.025-0.119)
SC3	0.079 (0.033-0.124)	0.083 (0.040-0.126)	0.078 (0.032-0.123)	0.079 (0.035-0.123)
SC4	0.222 (0.177-0.267)	0.218 (0.176-0.26)	0.221 (0.177-0.266)	0.218 (0.175-0.261)
SC5	0.328 (0.281-0.374)	0.277 (0.225-0.330)	0.328 (0.282-0.375)	0.301 (0.248-0.353)
UA2	0.128 (0.081-0.175)	0.111 (0.066-0.157)	0.124 (0.085-0.162)	0.116 (0.069-0.163)
UA3	0.122 (0.078-0.167)	0.119 (0.077-0.162)	0.124 (0.085-0.162)	0.120 (0.076-0.163)
UA4	0.308 (0.263-0.353)	0.278 (0.232-0.324)	0.304 (0.265-0.343)	0.283 (0.240-0.326)
UA5	0.300 (0.254-0.346)	0.253 (0.201-0.304)	0.304 (0.265-0.343)	0.283 (0.240-0.326)
PD2	0.076 (0.028-0.124)	0.086 (0.040-0.131)	0.074 (0.026-0.122)	0.079 (0.032-0.126)
PD3	0.088 (0.043-0.134)	0.092 (0.049-0.135)	0.088 (0.042-0.133)	0.089 (0.045-0.133)
PD4	0.265 (0.220-0.309)	0.258 (0.216-0.300)	0.264 (0.219-0.309)	0.259 (0.216-0.303)
PD5	0.361 (0.314-0.408)	0.308 (0.254-0.362)	0.362 (0.314-0.409)	0.333 (0.279-0.386)
AD2	0.141 (0.094-0.188)	0.134 (0.089-0.178)	0.142 (0.096-0.189)	0.140 (0.094-0.185)
AD3	0.254 (0.210-0.299)	0.233 (0.189-0.277)	0.256 (0.212-0.301)	0.246 (0.202-0.291)
AD4	0.431 (0.384-0.478)	0.411 (0.366-0.457)	0.419 (0.378-0.461)	0.398 (0.353-0.443)
AD5	0.404 (0.354-0.453)	0.342 (0.284-0.400)	0.419 (0.378-0.461)	0.398 (0.353-0.443)
N5		0.107 (0.042-0.172)		0.059 (0.001-0.117)

To produce a QALY weight for a health state in the EQ-5D-5L, the relevant coefficients are subtracted from one. For example, under Model B, health state 31245 would be valued as 1 – (0.086 + 0.111 + 0.258 + 0.342 + 0.107) = 0.096 i.e. picking up the four relevant coefficients for the non-level 1 dimensions plus the coefficient on the N5 term.

The distribution of weights under Model D is presented in Figure 1.

Figure 1: Distribution of EQ-5D-5L Utility Weights (Model D)



Discussion

The EQ-5D-5L algorithm developed in this work is, to our knowledge, the first such algorithm for this novel multi-attribute utility instrument. It was developed using a discrete choice experiment which is becoming more common in the field, and offers a number of attractive characteristics relative to other methods of preference elicitation. Importantly, these results now allow economic evaluations in Australia to utilise the EQ-5D-5L, which we expect to supersede the 3-level version in coming years.

However, there are a number of limitations to the work presented here. The consideration of interaction effects is limited to the inclusion or exclusion of the N5 term in the modelling. This is analogous to the approach taken to modelling TTO data in much of the EQ-5D-3L literature (Dolan, 1997). The DCE approach has been shown to be capable of exploring interactions in a less blunt approach (Viney, et al., 2011), and this type of extension would be a valuable next step in this area. The problem with designing DCEs to allow unbiased and precise estimation of these interaction terms is that the size of the experiment (and hence the sample size) can increase dramatically.

This is the reason why we did not attempt to explore this issue in this exploratory study, but is an area the work presented here can be adapted to move into.

One potential limitation of the approach adopted in this study is the possibility that an on-line panel is not representative of the Australian population, which may limit the applicability of the weights to the Australian population overall. An on-line panel was used because it is a very cost-effective means of collecting these data, and these approaches are increasingly used in valuation studies (Wittenberg and Prosser, 2011) . While the panel respondents can be selected to enhance the representativeness on observable characteristics (such as those reported in Table 2), there is still the potential concern that these respondents differ from the general population in some unobservable dimension (such as psychological attributes that may influence attitudes to health and death). However, we would argue that this criticism applies to some degree to all approaches to survey administration that might be used. The second criticism is that it has been argued that the mode of administration can significantly affect the quality of data collected (Bowling, 2005). Potentially, the nature of online respondents make this approach particularly susceptible to poor data quality in that they are not being observed while responding and it is not possible to identify how carefully they are considering the choices. However, it is likely that the weight of this more general criticism depends on the nature of the task. In a DCE, answering on criteria above and beyond the levels and dimensions presented (at extreme this might be answering randomly, or answering all A’s for example) does not bias the results assuming that some basic design properties have been considered. Thus, it is important to limit any unconsidered responses, but the impact of these in a DCE is to reduce the effective sample size rather than systematically bias the conclusions drawn. This contrasts with the TTO in which unconsidered responses do systematically bias conclusions, an example of which has been described elsewhere (Norman, et al., 2010).

A final limitation is that the method of analysis presented here (i.e. the random-effects probit) involves a simple approach to modelling of heterogeneity. While the panel nature of the data is reflected in the composite error term that is employed, there are a number of superior methods for the modelling of heterogeneity. For instance, a recent study (Fiebig, et al., 2010) has proposed the generalised multinomial logit model, which accounts for both the certainty with which respondents answer questions (termed scale heterogeneity) and the degree of divergence in preference regarding different areas of the presented profiles, in this case the dimensions within the EQ-5D-5L (termed preference heterogeneity). It has been argued that this modelling can significantly improve the fit of the model, but it remains to be seen whether values derived from the regression results (such as utility weights or willingness to pay values) are significantly impacted by this improved modelling of the error term (Greene and Hensher, 2011).

References

Australian Bureau of Statistics. *Education and training indicators, Australia, 2002.* Canberra, 2002.

Australian Bureau of Statistics. 3222.0 - population projections, Australia, 2004 to 2101 2005.

Australian Bureau of Statistics. *Tobacco smoking in Australia: A snapshot, 2004-05* Australian Bureau of Statistics: Canberra, 2006.

Australian Bureau of Statistics. *Household income and income distribution, Australia, 2005-06.* Canberra, 2007a.

Australian Bureau of Statistics. *Migration 2005-6 (3412.0).* ABS: Canberra, 2007b.

Bansback N, Brazier J, Tsuchiya A, Anis A. 2011. Using a discrete choice experiment to estimate societal health state utility values. *Journal of Health Economics*.

Bleichrodt N, Wakker P, Johannesson M. 1997. Characterizing QALYs by risk neutrality. *Journal of Risk and Uncertainty* **15**: 107-114.

Bowling A. 2005. Mode of questionnaire administration can have serious effects on data quality. *J Public Health (Oxf)* **27**: 281-291.

Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. 2007. *Measuring and valuing health benefits for economic evaluation*. Oxford University Press: Oxford.

Coast J, Flynn TN, Salisbury C, Louviere J, Peters TJ. 2006. Maximising responses to discrete choice experiments: A randomised trial. *Appl Health Econ Health Policy* **5**: 249-260.

Dolan P. 1997. Modelling valuations for EuroQol health states. *Medical Care* **35**: 1095-1108.

Fiebig D, Keane M, Louviere J, Wasi N. 2010. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* **29**: 393-421.

Greene WH, Hensher DA. 2011. Does scale heterogeneity across individuals matter? An empirical assessment of alternative logit models. *Transportation*.

Gyrd-Hansen D, Sogaard J. 2001. Analysing public preferences for cancer screening programmes. *Health Economics* **10**: 617-634.

Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5l). *Quality of Life Research*.

Hole AR. 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics* **16**: 827-840.

Lancsar E, Louviere J, Flynn T. 2007. Several methods to investigate relative attribute impact in stated preference experiments. *Social Science and Medicine* **64**: 1738-1753.

McFadden D. 1974. Conditional logit analysis of qualitative choice behaviour. In *Frontiers in econometrics,* Zarembka P, (ed.). New York Academic Press: New York.

McIntosh E, Ryan M. 2002. Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications for discontinuous preferences. *Journal of Economic Psychology* **23**: 367-382.

Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. 2009. International comparisons in valuing EQ-5D health states: A review and analysis. *Value in Health* **12**: 1194-1200.

Norman R, King M, Clarke D, Viney R, Cronin P, Street D. 2010. Does mode of administration matter? Comparison of on line and face-to-face administration of a time trade-off task. *Quality of Life Research* **19**: 499-508.

Richardson J, McKie J, Bariola E. *Review and critique of health related multi attribute utility instruments. Research paper 64.* Centre for Health Economcs, Monash University: Melbourne, 2011.

Robinson A, Spencer A. 2006. Exploring challenges to TTO utilities: Valuing states worse than dead. *Health Economics* **15**: 393-402.

Scott A. 2001. Eliciting gps' preferences for pecuniary and non-pecuniary job characteristics. *Journal of Health Economics* **20**: 329-347.

Stolk EA, Oppe M, Scalone L, Krabbe PFM. 2010. Discrete choice modeling for the quantification of health states: The case of the EQ-5D. *Value in Health* **13**: 1005-1013.

Street DJ, Burgess L. 2007. *The construction of optimal stated choice experiments: Theory and methods*. Wiley: Hoboken, New Jersey.

Thurstone LL. 1927. A law of comparative judgment. *Psychological Review* **34**: 273-286.

Tilling C, Devlin N, Tsuchiya A, Buckingham K. 2010. Protocols for time tradeoff valuations of health states worse than dead: A literature review. *Medical Decision Making* **30**: 610-619.

Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, et al. 2002. Estimating an EQ-5D population value set: The case of Japan. *Health Economics* **11**: 341-353.

Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, et al. 2011. An Australian discrete choice experiment to value EQ-5D health states. *Unpublished Manuscript*.

Wittenberg E, Prosser LA. 2011. Ordering errors, objections and invariance in utility survey responses: A framework for understanding who, why and what to do. *Applied Health Economics and Health Policy* **9**: 225-241.