Investigating the relative value of health and social care related quality of life using a discrete choice experiment

Brendan Mulhern[a], Richard Norman[b], Richard De Abreu Lourenco[a], Juliette Malley[c], Deborah Street[a], Rosalie Viney[a]

[a] Centre for Health Economics Research and Evaluation, University of Technology Sydney, 1 – 59 Quay St, Haymarket, Sydney, NSW 2000, Australia

[b] School of Public Health, Curtin University, Kent Street, Bentley, WA 6102, Australia

[c] Personal Social Services Research Unit, London School of Economics and Political Science, Houghton St, London, WC2A 2AE, UK

Author e-mail addresses:

BM: Brendan.mulhern@chere.uts.edu.au

RN: Richard.norman@curtin.edu.au

RDAL: Richard.deabreulourenco@chere.uts.edu.au

JM: j.n.malley@lse.ac.uk

DS: Deborah.street@chere.uts.edu.au

RV: Rosalie.viney@chere.uts.edu.au

Corresponding author:

Brendan Mulhern

Centre for Health Economics Research and Evaluation, University of Technology Sydney, 1 – 59 Quay St, Haymarket, Sydney, NSW 2000, Australia

Tel no: +61 2 9514 4725

E-mail: Brendan.mulhern@chere.uts.edu.au

Abstract

A key outcome in the evaluation of health technologies is the quality adjusted life year (QALY) which is often estimated using health measures such as the EuroQol instruments (EQ-5D-3L and EQ-5D-5L). The impacts of many interventions extend beyond a narrow definition of health to include non-health impacts such as social care related dimensions of quality of life (QoL). This means that there are circumstances where the QALY does not capture the full value of an intervention. In response to this, instruments with a wider measurement framework, such as the Adult Social Care Outcomes Toolkit (ASCOT), which measures social care related QoL, have been developed. Given the range of instruments available, it is important that decision-makers have tools to assess value for money comprehensively and consistently. To date, preference elicitation of different aspects of QoL combined within the same valuation procedure has not been tested. We investigate the relationship between health and social care aspects of QoL when assessed jointly by combining EQ-5D-5L and ASCOT in an online discrete choice experiment (DCE). In July 2016, 975 respondents recruited from internet panels completed 15 choice sets from an underlying design of 300. Conditional logit regression was used to estimate coefficient decrements for each attribute and examine their relative magnitude. Latent class and mixed logit modelling were used to understand preference heterogeneity. The results suggest trading across health and social care aspects indicated by coefficient estimates of differing magnitude. Dimensions with the largest disutility include four from EQ-5D-5L and one from ASCOT. There is evidence of preference heterogeneity at more severe dimension levels. We have used an established method to test the joint valuation of concepts measuring different aspects of QoL. The results have implications for the aspects of QoL that are included in QALY estimation and used in resource allocation decision-making.

Keywords

Australia; Discrete choice experiment; preferences; quality of life; cost utility analysis; QALY; Euroqol

Introduction

A key outcome in the cost utility analysis of new health technologies is the quality adjusted life year (QALY) which combines length and quality of life into a single metric. In estimating QALYs, the quality of life (QoL) weight used is anchored on a scale from full health (1) to dead (0) and is generally measured using multi-attribute utility instruments (MAUIs) with a health focus.

One such health focused MAUI is the EuroQol instrument. The descriptive classification system of the EuroQol instrument measures health related quality of life (HRQoL) across five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), and has a three-response level (EQ-5D-3L; Brooks et al 1996) and a five-response level (EQ-5D-5L; Herdman et al 2011) version. The EQ-5D-3L response levels are none-some-extreme/unable to/confined to bed (for the mobility dimension). The EQ-5D-5L response level descriptors are none-slight-moderate-severe-extreme/unable to. The EQ-5D-3L and EQ-5D-5L are widely used MAUIs (Szende et al 2007; Devlin and Brooks 2017), that are recognised as sources of QoL weights by a number of reimbursement agencies internationally (CADTH, 2017; ISPOR, 2016; NICE, 2013; Pharmaceutical Benefits Advisory Committee, 2015).

Value sets estimating the HRQoL weight for use in the calculation of QALYs are usually developed by eliciting the preferences of the general population for health states described by a MAUI. For example, time trade off (TTO) has been used to develop value sets for the EQ-5D-3L in the UK (Dolan 1997), the US (Shaw et al 2005) and Australia (Viney et al 2011). More recently discrete choice experiments (DCE) have been used to estimate EQ-5D-3L and EQ-5D-5L values internationally, including in the Netherlands (see e.g. Stolk et al 2010; Jonker et al 2017), Canada (Bansback et al 2012), Australia (Viney et al 2014; Norman et al 2013) and the UK (Bansback et al 2014; Mulhern et al 2017a,b).

A challenge that is not addressed by health-related MAUIs is the increasing recognition by

consumers and decision-makers that the impact of many interventions extends beyond a narrow definition of health outcomes to include broader non-health and QoL impacts (Brazier and Tsuchiya 2015; Wildman et al 2016). In a number of population groups this includes social care issues, and therefore the interplay between health and social care is fundamental to providing both longevity and quality of life. This means that there are settings and circumstances where the health-focused QALY is insensitive and does not capture the full value of an intervention or care setting.

Recently, instruments that focus more directly on social outcomes that arise from care interventions have been developed. One such instrument developed in the UK is the Adult Social Care Outcomes Toolkit (ASCOT; Netten et al, 2012). The ASCOT measures social care related quality of life (SRQoL) on eight dimensions (control, personal cleanliness, food and drink, personal safety, social participation, occupation, accommodation cleanliness and comfort, and dignity). Each dimension has four response levels with a variety of severity descriptors used (see Netten et al (2012) for the full descriptions). The ASCOT value set that can be used to estimate a social care QALY was derived using a combination of TTO and best worst scaling in the United Kingdom. The value set ranges from –0.171 to 1, with the anchors in this case being 'dead' (0) and the 'ideal' SRQoL state (1) (Netten et al 2012).

Recent international work has compared the EQ-5D-3L and the ASCOT in a range of different populations. Van Leeuwen et al (2015a) found that the EQ-5D-3L was more strongly associated with physical limitations than ASCOT, but less strongly associated with instruments measuring aspects beyond health in frail older people. Rand et al (2017) found that the ASCOT index score was 'moderately' correlated with the EQ-5D-3L usual activities and anxiety/depression dimensions (with moderate correlations defined as 0.3 - 0.5), but correlated at a lower level (<0.3) with mobility, self-care and pain/discomfort in adults with long-term physical, sensory and mental health conditions. Conversely, in a community-

dwelling sample, Kaambwa et al (2015) found that only the pain/discomfort dimension was moderately correlated with the ASCOT utility score (with moderate in this study defined as 0.4 - 0.6), with correlations between the EQ-5D index score and the ASCOT dimensions ranging from low (< 0.4 for control, cleanliness, occupation and accommodation) to moderate (for food and drink, safety, social participation and dignity). Content validity assessment suggests that respondents considered the items of both instruments valuable, but neither individually provided a comprehensive picture of a patient's QoL (van Leeuwen et al, 2015b). This evidence provides some support for the use of the EQ-5D instruments and ASCOT as complementary outcome measures in the economic evaluations of health and social care interventions.

As yet, however, little work has investigated the potential for unifying instruments focused on health and social aspects of QoL, nor assessed the relationship between instruments when both are valued using the same method. Traditionally, approaches to QoL valuation focus on HRQoL, which potentially disadvantages the assessment of health and social care interventions for aged care, palliative care, disability care and vulnerable populations. Although HRQoL is likely to be affected by the health conditions that these populations experience, the SRQoL impacts of interventions will not be specifically measured, except to the extent that they are reflected in changes in HRQoL. Consequently, SRQoL impacts will not be reflected in the values used for decision-making based on comparisons of the relative performance of interventions. Unifying constructs across instruments such as the EQ-5D-5L and ASCOT could lead to measures and methods that allow decision-makers to assess value for money in an inclusive and consistent way, by capturing the effects of both health and social aspects of QoL on the same utility scale. This would facilitate comparisons between interventions that have impacts predominantly on HRQoL, on SRQoL, or on a combination of both. The utility measured in this framework would be a latent utility that combines preferences for health and social care aspects of living, with the best state measured equivalent to no problems with HRQoL and ideal SRQoL. The utilities are not anchored on the full health to

dead QALY scale, but demonstrate the relative importance of domains of HRQoL and SRQoL that inform such a scale.

However, the derivation of preferences for different aspects of HRQoL and SRQoL to capture these benefits simultaneously using a valuation method (DCE) employed in the development of health-focused value sets (see e.g. Mulhern et al 2018a) has not been tested. This paper reports on an exploratory DCE study, which collects preference data from an Australian community sample, to investigate the joint valuation of HRQoL, as measured by the EQ-5D-5L, and SRQoL, as measured by ASCOT. This approach has not been attempted previously and as such makes an important contribution to the emerging literature exploring approaches to valuing interventions that go beyond the health focused QALY (e.g. Brazier et al 2018). Investigating relative preferences across the EQ-5D-5L and ASCOT also allows us to explore the use of DCE for the potential development of a combined instrument with a value set informed by aspects of both HRQoL and SRQoL.

1.1. Aims

The aims of the study are twofold:

1. To test the use of DCEs to elicit respondent preferences for different QoL profiles that incorporate aspects of both HRQoL and SRQoL as measured by the EQ-5D-5L and ASCOT; and
2. To investigate the relative magnitude of preferences for different aspects of HRQoL and SRQoL as measured by the EQ-5D-5L and ASCOT.

Methods

DCE valuation task

A DCE was developed based on the dimensions of the EQ-5D-5L and ASCOT Self Completion Tool (SCT-4). The DCE choice sets presented pairs of QoL states comprising dimensions from both the EQ-5D-5L and ASCOT (13 dimensions in total) and asked respondents to choose which profile they preferred. All five EQ-5D-5L dimensions were included. ASCOT

includes nine dimensions, but only eight of these were used by the instrument developers to generate the value set (Netten et al (2012). For consistency the same eight dimensions were included in this study. Because each choice set included 13 dimensions, we simplified the choice task by imposing overlap in the design (Norman et al, 2016; King et al 2018; Jonker et al 2018a; 2018b). That is, we constrained five of the dimensions in each choice set to have the same level of severity in both profiles and allowed levels of the other eight dimensions to vary in each task. The dimensions that differed within choice sets were highlighted with a light grey background (see Figure 1). The use of shading to highlight dimension-level differences within choice sets has been shown to produce similar choice results to those obtained without shading, whilst simplifying the choice task for respondents (Mulhern et al 2018b). The information and instructions provided to respondents prior to presenting the choice sets is reproduced in supplementary Appendix 1. For consistency with the DCE task used to value the EQ-5D-5L within the recommended EQ-VT protocol (Oppe et al 2014), the choice sets did not specify living in the health states for a certain amount of time.

The results of past work into the effectiveness of overlap (Jonker et al 2018b) and formatting (Mulhern et al 2017; Jonker et al 2018a) were used to support the development of the choice set format. An initial pilot launch of the live survey was used to assess the functioning of the survey and the feasibility of the choice set format to respondents. This was measured via multiple choice usability and free-text questions about the difficulty of the task (described in the survey design section) and by initial modelling of the DCE data, where we checked indications of coefficient ordering between the levels of each attribute. The completion times for the pilot survey were used to inform the minimum survey completion time imposed for the responses included for analysis from the full sample.

Study design

As per previous studies employing DCE methods to value QoL instruments (Mulhern et al 2017; Bansback et al 2012; Norman et al 2014), we have included more choice sets in the

design than there are parameters in the model to estimate. The main effects model combining the EQ-5D-5L and ASCOT has 44 parameters to estimate (EQ-5D-5L: (5x (5-1)) = 20; ASCOT (8 x (4-1)) = 24). We constructed 300 choice sets using a modified-Fedorov algorithm with the objective function being to optimise the estimation of a main effects model using the criterion of minimal D-error. This process iteratively improves the set of choice sets included in the design until the D-error does not substantially improve (Cook and Nachtsheim (1980)). This was implemented using the DCE design software Ngene (ChoiceMetrics, 2014). Supplementary appendix 2 includes one block of choice sets from the design, and the associated level descriptors for each dimension. Although there is the potential for implausible combinations of health and social care dimension levels, we did not restrict any combinations in this study. This is because it is difficult to make a judgement, a-priori, that certain combinations are not realistic, particularly as what is considered implausible has been found to be respondent specific (Yang et al 2019). Marten et al (2017) found that EQ-5D-5L level combinations assumed to be implausible (for example no problems with self-care combined with unable to do usual activities) actually appear in general population self-report data. Excluding particular level combinations may also imbalance the design and impact the coefficient estimates derived from subsequent model estimation in non-systematic ways (Viney et al 2014).

The 300 choice sets were separated into 20 blocks of 15 using the blocking functionality available in Ngene. Each of the 20 blocks was included in two versions of the survey that replicated the full design: Version 1 presented the EQ-5D-5L dimensions followed by the ASCOT dimensions and Version 2 presented the ASCOT dimensions followed by the EQ-5D-5L dimensions, with the dimensions within instruments presented in the standard order described in the introduction. Respondents were subsequently randomised to one of 40 survey blocks. Further randomisation of dimensions within the DCE could have been imposed, but the decision was made not to do this to allow respondents to always see concepts related to HRQoL (EQ-5D-5L) followed by SRQoL (ASCOT) or vice versa. Evidence for dimension

order effects in previous health state valuation work is inconclusive (Mulhern et al 2017b; Norman et al 2016).

Survey design

The survey was administered online. The survey comprised background information about the project and ethics approval, followed by an informed consent page, then questions on respondents' demographic characteristics, health and QoL (including EQ-5D-5L and ASCOT). Subsequently, respondents were shown instructions about the task (see supplementary Appendix 1) and were told that they will see two different descriptions of health and social care, a warm up task and the 15 DCE tasks. The order of appearance of each set of tasks within a block was randomized. Finally, multiple choice questions about the difficulty of the tasks (including the overall difficulty, the difficulty imagining the descriptions, and the difficulty telling the difference between them) were included along with a free-text question to understand respondents' opinions of the survey questions and the content in general. These questions provided pilot data relating to the difficulty of the choice sets and the functioning of the survey.

Recruitment and respondents

The study aimed to recruit 1,000 respondents from the Australian general population, targeted to be representative in terms of age and gender quotas. This sample size was targeted to provide approximately 50 observations per DCE choice set (1,000 respondents x 15 observations divided by 300 choice sets) which is in line with other DCE valuation work (Norman et al, 2013; Mulhern et al, 2014). The initial pilot launch recruited approximately 10% of the sample, with the survey reopened following initial assessment of the survey functioning. Respondents were recruited at random from existing internet panels managed by Survey Sampling International, who allocated respondents who were willing to complete questionnaires during the data collection period. To support the generalizability of our findings, the panel company recruited from multiple subpanels. Respondents read the project

information and consented to take part before starting the survey. A small incentive was provided if respondents completed the full requirements of the survey in more than the minimum completion time imposed following the pilot launch. The respondents were not informed about the minimum completion time imposed. The amount and type of incentive differed depending on the procedures of the subpanel from which the respondent was recruited. The study procedure was approved by the Centre for Health Economics Research and Evaluation, University of Technology Sydney, program ethics process [UTS HREC REF NO. 2015000135]. Data collection took place during July 2016.

Data analysis – Sample descriptives

We compared the demographic characteristics of the sample to those observed in the Australian population (ABS, 2017). We also calculated the EQ-5D-5L and ASCOT utility scores. For the EQ-5D-5L we used the Australian value set developed by Norman et al (2014). For ASCOT we used the UK value set (Netten et al 2012) as an Australian value set is not available. We assessed the frequency of respondents endorsing each severity level of each of the EQ-5D-5L and ASCOT dimensions.

Data analysis – Conditional logit

Conditional logit regression was used to generate coefficient estimates for each level of the EQ-5D-5L and ASCOT dimensions using the model:

$$u_j = \boldsymbol{\beta} \boldsymbol{x}_j + \varepsilon_j \qquad (1)$$

where $\boldsymbol{\beta}$ is the choice coefficient attached to the attributes $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ is the error term (where the error term is assumed to have a standard Type I extreme value distribution). Robust standard errors were used to take into account repeat observations per respondent. Conditional logit regression does not assess preference heterogeneity but allows for comparison of the overall magnitude (a proxy for importance at the overall level) of the dimensions included. This allowed us to assess the overall rank of the 13 dimensions. We modelled the data for the whole sample (Model 1), and also assessed the consistency of the models for each of the

survey versions separately (Models 2 and 3). To assess the impact of order effects we tested the null hypothesis that the underlying scale (and therefore respondent preferences) does not differ across the two orders using an adaptation of the scale testing approach proposed by Swait and Louviere (1993). This is done using a likelihood ratio test given by $LR = -2(LL_R - LL_U)$ where LL_R is the log-likelihood of a model estimated on the pooled sample which allows for scale differences but assumes that the parameter estimates are the same across the two order groups (Model 4). LL_U is calculated as the sum of the log likelihoods of the models estimated on each order separately (an unrestricted model allowing for preference variation). If the LR statistic is greater than the critical value from a chi square distribution with degrees of freedom equal to the difference in the number of parameters between the restricted and unrestricted models (in this study 61.7) then the null hypothesis may be rejected. The significance of the scale parameter from the pooled restricted model is also considered. Stata 15 (StataCorp 2017) was used for modelling, with the scale model estimated using *clogithet*, a user written Stata module (Hole 2006a; 2006b).

Data analysis – Preference heterogeneity

Exploration of preference heterogeneity was considered to be important given the different QoL aspects included, which may have different impacts and meaning in different population subgroups. To assess preference heterogeneity, both latent class (Greene and Hensher 2002) and mixed logit (Hensher et al 2015; McFadden and Train 2000) regression models were used (Models 5 and 6 respectively).

Latent class analysis looks for groups of respondents with similar patterns of preferences and has been used in other health state valuation research to understand heterogeneity (Tsuchiya et al 2012). For this modelling, the baseline utility function (equation 1) was adapted to allow for heterogeneity at the respondent level (*i*):

$$u_{ij} = \beta_i + \boldsymbol{\lambda}_i' \boldsymbol{x}_j + \varepsilon_{ij} \qquad (2)$$

The results produced are parameter estimates for different preference classes estimated from the overall dataset. Models including between two and six classes were tested. In line with Train (2008), the optimum number of classes was determined using the Bayesian Information Criterion (BIC), a measure of model fit that takes both the number of parameters and the number of observations into account. The model with the lowest BIC is generally preferred (and is reported as Model 5 in the results section). Parameters indicating the class membership of different demographic groups were employed as class delimiters. The demographic groups were entered as binary dummy variables to allow for interpretation of the probability estimates, and included age (18 – 65 years old, and 65 or older), gender, and having a long-term health condition. The Stata package 'lclogit' (Pacifico & Yoo, 2013) was used for this analysis.

Mixed logit (Hensher and Greene, 2002) is a random parameter model that allows an assessment of unobserved preference heterogeneity under violation of the independence of irrelevant alternatives assumption to be made. The mixed logit model builds on the conditional logit model by specifying random parameters in the model. Equation 3 displays the mixed logit model where the utility for individual *i* associated with choice *j* in scenario *s* is:$u_{ijs} == \beta x'_{ijs} + (\eta_i x'_{ijs} + \varepsilon_{ijs})$ (3)

where β is a vector of coefficients and X'_{ijs} is a vector of explanatory variables and η_i is a variability term. The mixed logit model was estimated in Stata using the mixlogit command (Hole 2007). We assumed that each parameter independently followed a normal distribution . In this study all dimension-level EQ-5D-5L and ASCOT parameters were assessed for heterogeneity in a single model (Model 6).

Results

Pilot launch

During the initial pilot launch, 118 respondents completed the survey. The mean (median) time to complete was 23 minutes and 54 seconds (18 minutes and 42 seconds) minutes, with a minimum of 3 minutes and 24 seconds. The initial model tested on the DCE data from the pilot sample indicated that the majority of the dimensions had evidence of monotonicity of coefficient levels – critical for the development of utility scales. The key model performance indicators were as expected and were based on the full pilot sample including people who completed the survey in 3 minutes and 24 seconds or longer. From this we judged that data obtained from this range of completion times would produce a valid data from which to model preferences. We therefore set a minimum completion time of greater than 3 minutes. . Therefore, we set three minutes as the minimum completion time to exclude responses from people who completed the survey very quickly. This also meant that the pilot sample could be retained as part of the main sample.

Regarding the usability questions, we found that only 13% of the sample agreed that they found the task difficult, 17% agreed that it was difficult to imagine the scenarios and 13% agreed that it was difficult to tell the difference between the descriptions. The free-text question did not result in any concerning issues. This evidence was used to support the choice set formats used, and launch the full sample data collection.

Sample

Overall, 1,226 online panel members accessed the survey. Of these, 1,177 (96.0%) consented, 175 (14.3%) dropped out during the survey, 76 (6.2%) completed the survey in less than three minutes, and 975 (79.5%) fully completed the full survey in more than three minutes (this included the 118 respondents from the pilot launch). The mean (median) completion time was 26 minutes and 24 seconds (22 minutes and 12 seconds). Table 1 reports the demographic characteristics of the sample in comparison to the available statistics for the Australian population (ABS 2017). The sample was generally representative of the Australian population in age, gender and income, but respondents were more highly educated

and more likely to be born in Australia. Overall, 44% self-reported a long-term health condition. The ASCOT utility scores were higher than those from the EQ-5D-5L. Figure 2 displays the proportion of respondents answering at each level of each dimension. The respondents are distributed primarily across the first four severity levels of the EQ-5D-5L and the first three of the ASCOT dimensions; few respondents endorsed the most severe level of each dimension across the instruments.

DCE – Conditional logit

Table 2 reports the results for three models. In model 1, all responses across both versions of the survey are pooled (that is regardless of the order of the instrument). Models 2 and 3 are estimated separately for respondents who saw the EQ-5D-5L dimensions first in each choice set and those who saw the ASCOT dimensions first respectively. Non-monotonic coefficients are highlighted in bold, categorization of *p*-values for the difference between the coefficient estimates and the omitted baseline level (1) are indicated by stars, and actual *p*-values for the significance between adjacent levels (relative to the immediately better level) are reported in the 'sig (btwn)' column.

The rank of the dimension (using the overall magnitude of the disutility at the worst level as an indicator of the dimensions on which the most weight is placed) is also reported. The coefficient with the largest decrement is pain/discomfort (models 1 and 3) and mobility (model 2) and the coefficient with the smallest estimate is social participation. For model 1 (the pooled conditional logit), the magnitude of overall dimension-level coefficients can be used to understand the relative weight placed on dimensions. The dimensions with the largest disutility include four dimensions from the EQ-5D-5L (pain/discomfort, mobility, anxiety/depression, self-care) and one from the ASCOT (control). This is followed by two from the ASCOT (food and drink, safety), EQ-5D-5L usual activities and another two ASCOT dimensions (cleanliness, occupation). The remaining three ASCOT dimensions (accommodation, dignity, social participation) have smaller coefficients. The magnitude of the estimates suggests that

dimensions from one of the instruments were not consistently preferred to dimensions from the other. There is some evidence of non-significant disordered levels for three dimensions (usual activities levels 4 (severe) and 5 (unable to), pain/discomfort levels 2 (slight) and 3 (moderate), and accommodation levels 1 (home is as clean and comfortable as I want) and 2 (home is adequately clean and comfortable)). The EQ-5D-5L dimension coefficients increase significantly between levels 3 (moderate) and 4 (severe), and for three of the dimensions (mobility, pain/discomfort, anxiety/depression) the difference between levels 4 (severe) and 5 (extreme/unable to) was also significant. For the ASCOT, the difference between levels 3 and 4 (the two most severe levels with different severity descriptors used for each) for all eight dimensions was significant.

Comparing Models 2 and 3 suggests there is some evidence of an order effect in that the order of dimension presentation matters for preference estimation. The EQ-5D-5L dimensions have a larger decrement when presented first (Model 2). When ASCOT is presented first (Model 3), the magnitude of the disutility of control becomes larger in comparison to the EQ-5D-5L dimensions, and there is increased non-monotonicity, in particular for the EQ-5D-5L dimensions; non-monotonicity occurred in only one dimension of the EQ-5D-5L (anxiety/depression) in Model 2, but occurred in three (across self-care, usual activities, pain/discomfort) in Model 3. The overall pattern of ranking of the coefficients is reasonably similar. Dimensions ranked one to five in Model 2 are also ranked one to five in Model 3 (with four of the coefficients in a different order). Dimensions Ranked six to eight are also ranked six to eight (in a different order) and those ranked nine to 12 are in the same order across both models. Therefore ten of the 13 estimates are ranked in the same order across the models.

To test the pattern further using scale testing, Model 4 reports the coefficients from the restricted model, and the scale parameter controlling for the presentation order. Comparing this model with the two unrestricted models for each order separately gives an LR statistic of 118, and the scale parameter is significant at the 0.001 level, hence the null hypothesis of

preference homogeneity is rejected. This indicates that the order of measure presentation has an impact on the results.

DCE – Assessing heterogeneity

For the latent class analysis, the two-class model (Model 5) displayed the lowest BIC (18,093) of the models including up to six classes, and the resulting estimates are shown in Table 3. The first class includes 55% of the sample who demonstrate strong ordered preferences across many of the EQ-5D-5L and ASCOT dimensions. The second class (45%) includes a less clear pattern of preferences, with evidence of disordering across dimension levels. Those in class 1 are more likely to be aged over 60 and have a long-term condition than those in class 2.

Model 6 is the mixed logit allowing for heterogeneity across all dimension levels, with significant standard deviations indicating dimensions with heterogeneity. Variation in preferences is apparent for the more severe levels of all five EQ-5D-5L dimensions, and for seven of the ASCOT dimensions (all but food and drink which has evidence of heterogeneity at level 2).

Discussion

This research investigated the relative magnitude of health and social care related quality of life dimensions included in the EQ-5D-5L and ASCOT using a stated preference approach in Australia. We find that respondents make choices reflecting their trade-offs between diverse dimensions of quality of life within an overall utility framework combining preferences for certain aspects of HRQoL and SRQoL. The estimates demonstrate that the HRQoL and SRQoL outcomes included in the EQ-5D-5L and ASCOT have different levels of importance (using coefficient size as a proxy for importance) to a large and generally representative pool of respondents. The magnitude of preferences is generally higher for EQ-5D-5L HRQoL

dimensions, with some exceptions, and there is evidence of preference heterogeneity.

The overall coefficient magnitude is an indicator of which particular QoL aspects within the classification system respondents prefer, and suggest that two measures with different perspectives can be perceived as describing a broader concept within which people trade across different aspects. This pattern of preferences has implications for decision-making using a conventional QALY (focusing on HRQoL), that does not include wider areas of SCRQoL. The use of an instrument generating a value set combining HRQoL and SCRQoL would be driven more by the needs of particular groups, for example people with long-term conditions or frailty. These values could be used to assess the integration of services for people with multiple and complex conditions for whom maintaining QoL within a social care setting may be a more important consideration than improving health.

There is also evidence of preference heterogeneity across all dimensions to different degrees, with most heterogeneity apparent at the more severe dimension levels. This means that preferences for the dimensions of both instruments differ in different groups of respondents in terms of which dimension they would most want to avoid. This has implications for the sensitivity of decision-making, as a value set applied to data from different population groups may not accurately reflect the preferences of that population or patient group. Value sets representing general population preferences at the overall level are preferred by many decision makers, but population specific value sets taking into account heterogeneity of preferences could be considered for sensitivity analysis. Interactions between HRQoL and SRQoL dimension levels included in the EQ-5D-5L and ASCOT might also be an important area for further investigation. For example, an interaction between ASCOT social participation and EQ-5D-5L anxiety/depression would help understand the link between, and preferences for avoiding, mental health problems resulting from possible social isolation. Our study did not allow for assessment of interactions as it was beyond the scope of this exploratory work.

It is possible to compare the disutility of the coefficients for each instrument with those from valuation studies of the EQ-5D-5L and ASCOT separately. The order of magnitude of the coefficients we estimated within each individual instrument shows some consistency with other published value sets produced using a range of preference elicitation methods. Regarding EQ-5D-5L, the order of the coefficients is similar to the Australian value set derived using DCE methods (Norman et al 2013), with mobility, pain/discomfort and anxiety/depression having the largest decrements. The order of the ASCOT dimensions is consistent with the UK value set (Netten et al 2012) for the dimensions with the largest and smallest overall impact on ASCOT utility. One key difference for the ASCOT preferences is the weight placed on control compared to the other dimensions, which is higher in this study. This could be linked to the preferences of the Australian population, or an effect of presenting control alongside the EQ-5D-5L dimensions, where control over life might be considered differently when presented alongside specific health aspects.

Other work has assessed the relationship between preferences elicited for both instruments in the same study. Stevens et al (2018) estimated an exchange rate between EQ-5D-3L and the ASCOT using TTO valuations of each measure separately in the same respondents. The exchange rate suggested that health outcomes as measured by the EQ-5D-3L were more valued than ASCOT outcomes, but the gradient was close to 1 and the intercept was also small. The study did not allow for respondents to express preferences for different dimensions measured by the instruments within the same framework, as the health and social care descriptions were valued separately. In this work we combined the two aspects and show generally higher preferences for HRQoL aspects included in EQ-5D-5L with some exceptions. The studies provide complementary information from different preference elicitation approaches about how people consider HRQoL and SRQoL as measured by the EQ-5D-5L and ASCOT but are difficult to directly compare given the different methodologies. A key point of difference is the use of survival as the unit of trading across different QoL states in the TTO approach which changes the cognitive nature of the task. TTO also produces direct values for

particular states in comparison to DCE that produces binary choices for QoL states aggregated at the overall level to estimate values.

Another factor that may impact upon our results is the wording of the dimensions included. This clearly differs between the instruments, but qualitatively the wording of the EQ-5D-5L is more consistent internally than the ASCOT as it uses the same severity descriptors across dimensions. This difference in the wording of the descriptions and the severity levels may influence valuations and potentially mask the importance of certain domains (that are important to the respondent but the way they are worded means that they do not have the same overall severity perception). It would be useful to examine the relationship between the dimensions and the wording in a more systematic way. This could lead to further insight regarding the relationship between the different dimensions using consistent descriptions and severity levels. It may also lead to the development of dimension descriptors that are quite different to those in the original instruments, but further work could test whether homogenizing the wording influences both self-report and trading across dimensions. However, the amount of work required to develop new instruments and associated descriptions is extensive. The developmental work conducted for the EQ-5D-5L (Herdman et al 2011) and ASCOT (Netten et al 2012) is comprehensive, and therefore using existing instruments is a key starting point in understanding how respondents trade across dimensions.

In the models based on the survey ordering, the relative magnitude of some of the dimensions from the instrument that appears first is increased, and the scale testing is consistent with there being different patterns of preferences across the subgroups. When the ASCOT dimensions appear first, there is evidence of increased non-monotonicity, particularly for the EQ-5D-5L dimensions. However, the overall pattern of dimension ranking for each model has similarities. This provides some evidence for an order effect that may be aggregated for the overall model. We did not impose full dimension order randomization in the design of this study as in other DCEs the impact of dimension ordering has not been pronounced (Norman

et al 2016; Mulhern et al 2017). It may be that the difference between this work and previous studies is that here we have moved clusters of dimensions rather than individual ones.

This work has a number of limitations and associated opportunities for further work. Firstly, as this was exploratory work to understand respondent trading across dimensions, we did not include any form of anchoring to the latent scale values, for example by including duration (see e.g. Bansback et al 2012; 2014; Norman et al 2013; 2014; Viney et al 2014). Therefore, we do not have estimates on the full health – dead utility scale. Incorporating duration would be a key development for further work in this area to allow the values produced to be used as inputs for the estimation of QALYs. Secondly, as with all online studies, it is difficult to assess respondent engagement with completing the DCE task. To support completion, we introduced overlap across five dimensions and used shading, both strategies that have been shown to enhance respondent completion rates in previous studies (Mulhern et al 2018b; Jonker et al 2018b). We do not have detailed qualitative information about whether imposing overlap and shading supported respondent completion, but the general ordering and interpretability of the results, and the responses to the self-reported difficulty questions in the pilot launch suggest that the format used was acceptable to respondents.

This study is able to draw inferences about the relationship between HRQoL and SRQoL as measured by the EQ-5D-5L and ASCOT. Arguably, the methods we apply in this research could reasonably be extended to forming other joint valuation indices that incorporate other conceptualizations of QoL. For example, further work could apply the methods we have used to assess the relationship between HRQoL and capabilities (as measured by the ICEpop CAPability measure for Adults) (ICECAP-A; Al-Janabi et al 2012) or wellbeing. There are also other measures of HRQoL, that include both different and overlapping dimensions from those used in our research (for example the Short Form-6 Dimension (SF-6D; Brazier et al 2002; Brazier and Roberts 2004) describes HRQoL differently from the EQ-5D-5L (Herdman et al 2011)). Thus, our work provides evidence that the DCE valuation approach can be used to

estimate joint indices of QoL. This provides significant scope to expand on such indices to produce QALY valuations that are more sensitive to the impacts of care services across a wide range of patient groups and settings. It could also be hypothesized that there are cross-cultural differences in preferences for different aspects of quality of life depending on a number of factors such as the characteristics of the healthcare system, and different attitudes towards aspects of health and social care. Further research could repeat this work in different countries. Finally, other valuation tasks such as best worst scaling (Krucien et al 2017) may provide preference information from a different perspective about how populations trade between different impacts of health conditions.

In conclusion, we have used an established valuation methodology (DCE) to demonstrate that it is possible to value concepts measuring different (and in some cases overlapping) aspects of quality of life (health and social-care related) on the same underlying scale. We have also demonstrated that respondents trade between the different concepts. This has implications for decision-making around the funding and use of interventions that have a wider impact on quality of life than just that captured by a more narrowly focused health-related QALY metric.

References

Al-Janabi, H., Flynn, T., Coast, J., 2012. Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. Qual Life Res. 21, 167-176.

Australian Bureau of Statistics. 2017. 2016 Census Quick Stats. Available at http://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/036.

Bansback, N., Brazier, J., Tsuchiya, A., Anis, A. 2012. Using a Discrete Choice Experiment to Estimate Health State Utility Values. J Health Econ. 31(1), 306-318.

Bansback, N., Hole, A.R., Mulhern, B., Tsuchiya, A. 2014. Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. Soc Sci Med. 114, 38-48.

Brazier, J., Roberts, J., Deverill, M. 2002. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 21(2), 271-92.

Brazier, J.E., Roberts, J. 2004. Estimating a preference-based index from the SF-12. Med Care. 42(9), 851-59.

Brazier J.E., Rowen D., Lloyd A., Karimi M. Future Directions in Valuing Benefits for Estimating QALYs: Is Time Up for the EQ-5D? Value Health 22(1), 2019 62-68.

Brazier J, Tsuchiya A. 2015. Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. Applied Health Economics and Health Policy. 13(6), 557–565.

Brooks, R. 1996. EuroQol: The current state of play. Health Policy. 37, 53-72.

Canadian Agency for Drugs and Technologies in Health. 2006. Guidelines for the Economic Evaluation of Health Technologies: Canada. 3rd edition. Ottawa: Canadian Agency for Drugs and Technologies in Health.

ChoiceMetrics. 2014. Ngene 1.1.1. User Manual & Reference Guide. ChoiceMetrics: Australia.

Cook, R.D., Nachtsheim, C.J. 1980. A Comparison of Algorithms for Constructing Exact D-Optimal Designs. Technometrics. 22(3), 315-324.

Devlin, N., Brooks, R. 2017. EQ-5D and the EuroQol Group: Past, Present and Future. Applied Health Economics and Health Policy. 15(2), 127-37.

Dolan, P. 1997. Modeling valuations for EuroQol health states. Med Care. 35(11), 1095-108.

Greene, W.H., Hensher, D.A. 2002. A Latent Class Model for Discrete Choice Analysis: Contrasts with Mixed Logit. Institute for Transport Studies, University of Sydney: Working Paper ITS-WP-02-08.

Hensher, D.A., Green, W.A. 2002. The Mixed Logit Model: The State of Practice. University of Sydney: Working Paper ITS-WP-02-01.

Hensher, D.A., Rose, J.M., Greene, W.H. 2015. Applied Choice Analysis 2nd Edn. Cambridge: University Press.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin D., Bonsel, G., Badia, X. 2011. Development and preliminary testing of the new five level version of EQ-5D (EQ-5D-

5L). Qual Life Res. 20(10), 1727-36.

Hole, A.R., 2006a. CLOGITHET: Stata Module to Estimate Heteroscedastic Conditional Logit Models. Statistical Software Components S456737. Boston College. Department of Economics.

Hole, A.R., 2006b. Small-sample properties of tests for heteroscedasticity in the conditional logit model. Econ. Bull. 3 (18), 1e14.

Hole, A.R. 2007. Fitting mixed logit models by using maximum simulated likelihood. The Stata Journal. 7(3), 388-401.

ISPOR. 2016. 'Pharmacoeconomic Guidelines Around the World'. ISPOR Connections, 10.

Jonker, M., Attema, A., Donkers, B., Stolk, E., Versteegh, M. 2017. 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. 26(12), 1534-47.

Jonker, M.F., Donkers, B., deBekker-Grob, E., Stolk, E. 2018a. The Effect of Level Overlap and Color Coding on attribute Non-attendance in Discrete Choice Experiments. Value Health. 21, 767-71.

Jonker, M.F., Donkers, B., deBekker-Grob, E., Stolk, E. 2018b. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Economics. DOI: 10.1002/hec.3846.

Kaambwa, B.C., Gill, L., McCaffrey, N.M., Lancsar, E., Cameron, I.D., Crotty, M., et al. (2015). An empirical comparison of the OPQoL-Brief, EQ-5D-3L and ASCOT in a community dwelling population of older people. Health and Quality of Life Outcomes, 13(164).

King M.T., Viney R., Pickard A.S., Rowen D., Aaronson N., Brazier J.E. et al. Australian Utility Weights for the EORTC QLU-C10D, a Multi Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. PharmacoEconomics, 36, 2018, 225–238.

Krucien, N., Watson, V., Ryan, M. 2017. Is best worst scaling suitable for health state valuation? A comparison with discrete choice experiments. Health Econ. 26(12), e1-16.

Marten, O., Mulhern, B., Bansback, N., Tsuchiya, A. 2017. Modelling implausible EQ-5D-5L

health states: prevalence in the general public and it's effects on health state valuation – preliminary results. EuroQol Plenary: Barcelona.

McFadden, D., Train, K. 2000. Mixed MNL models for discrete response. J Appl Econ.15, 447-470.

Mulhern, B., Bansback, N., Brazier, J., et al. 2014. Preparatory Study for the Revaluation of the EQ-5D Tariff: Methodology Report. Health Technol Assess. 18(12).

Mulhern, B., Bansback, N., Hole, A.R., Tsuchiya, A. 2017. Using Discrete Choice Experiments with duration to model EQ-5D-5L health state preferences: Testing experimental design strategies. Med Decis Mak. 37(3):285-297.

Mulhern, B., Norman, R., Lorgelly, P., Lancsar, E., Ratcliffe, J., Brazier, J., Viney, R. 2017. Is dimension order important when valuing health states using Discrete Choice Experiments including duration? Pharmacoeconomics. 35(4), 439-51.

Mulhern, B., Norman, R., Shah, K., Bansback, N., Longworth, L., Viney, R. 2018b. How should DCE with duration choice sets be presented for the valuation of health states? Med Decis Mak. 38(3), 306-318.

Mulhern, B., Norman, R., Street, D., Viney, R. 2018a. One method, many methodological choices: A structured review of Discrete Choice Experiments for health state valuation. Pharmacoeconomics, doi: 10.1007/s40273-018-0714-6.

National Institute of Health and Care Excellence. 2013. Methods for health technology assessment. London: NICE.

Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.M., Brazier, J., Flynn, T., Forder, J. 2012. Outcomes of social care for adults: developing a preference-weighted measure. Health Technol Assess. 16, 1-166.

Norman, R., Cronin, P., Viney, R. 2013. A Pilot Discrete Choice Experiment to Explore Preferences for EQ-5D-5L Health States. Applied Health Economics and Health Policy. 11(3), 287-298.

Norman, R., Kemmler, G., Viney, R., Pickard, A.S., Gamper, E., Holzner, B., Nerich, V., King, M. 2016. Order of presentation of dimensions does not systematically bias utility weights from

a discrete choice experiment. Value Health. 19, 1033-38.

Norman, R., Viney, R., Brazier, J., et al. 2014. Valuing SF-6D Health States Using a Discrete Choice Experiment. Med Decis Mak. 34(6), 773-86.

Oppe, M., Devlin, N.J., Van Hout, B., Krabbe, P., De Charro, F. 2014. A Program of Methodological Research to Arrive at the New International EQ-5D-5L Valuation Protocol. Value Health. 17(4), 445-53.

Pacifico, D., Yoo, H. 2013. 'lclogit: a Stata command for fitting latent-class conditional logit models via the expectation-maximization algorithm.', Stata journal. 13(3), 625-639.

Pharmaceutical Benefits Advisory Committee. 2015. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Canberra: Commonwealth of Australia.

Rand, S., Malley, J., Towers, A.M., Netten, A., Forder, J. 2017. Validity and test-retest reliability of the self-completion adult social care outcomes toolkit (ASCOT-SCT4) with adults with long-term physical, sensory and mental health conditions in England. Health and Quality of Life Outcomes. 15:163.

Shaw, J.W., Johnson, J.A., Coons, S.J. 2005. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 43(3), 203-20.

StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LP

Stevens, K., Brazier, J., Rowen, D. 2018. Estimating an exchange rate between the EQ-5D-3L and ASCOT. Eur J Health Econ, 19(5): 653-61.

Stolk, E., Oppe, M., Scalone, L., Krabbe, P. 2010. Discrete choice modelling for the quantification of health states: The case of the EQ-5D. Value Health. 13(8), 1005-13.

Swait, J., Louviere, J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. J. Mark. Res. 30 (3), 305e314.

Szende, A., Oppe, M., Devlin, N. 2007. EQ-5D value sets: Inventory, comparative review and user guide. Springer: Rotterdam.

Train, K. 2008. EM Algorithms for Nonparametric Estimation of Mixing Distributions. Journal of Choice Modelling. 1(1), 40-69.

Tsuchiya A., Mulhern B., Bansback N., Hole A. Further exploration of discrete choice experiments with a duration attribute to value health states. EuroQol Group Plenary, 2012, Rotterdam

van Leeuwen, K.M., Bosmans, J.E., Jansen, A.P., Hoogendijk, E.O., van Tulder, M.W., van der Horst, H.E., Ostelo, R.W. 2015a. Comparing measurement properties of the EQ-5D-3L, ICECAP-O, and ASCOT in frail older adults. Value Health. 18(1), 35-43.

Van Leeuwen, K.M., Jansen, A.P., Muntinga, M., Bosmans, J., Westerman, M., van Tulder, M., van der Horst, H. 2015b. Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. BMC Health Services Research, 15:201.

Viney, R., Norman, R., Brazier, J., et al. 2014. An Australian discrete choice experiment to value EQ-5D health states. Health Econ. 23, 729-42.

Viney, R., Norman, R., King, M.T., Cronin, P., Street, D.J., Knox, S., Ratcliffe, J. 2011. Time trade-off derived EQ-5D weights for Australia. Value Health. 14(6), 928-36.

Wildman, J., Meekin, P., Grieve, E., Briggs, A. 2016. Economic evaluation of integrated new technologies for health and social care: Suggestions for policy makers, users and evaluators. Soc Sci Med. 169, 141-48.

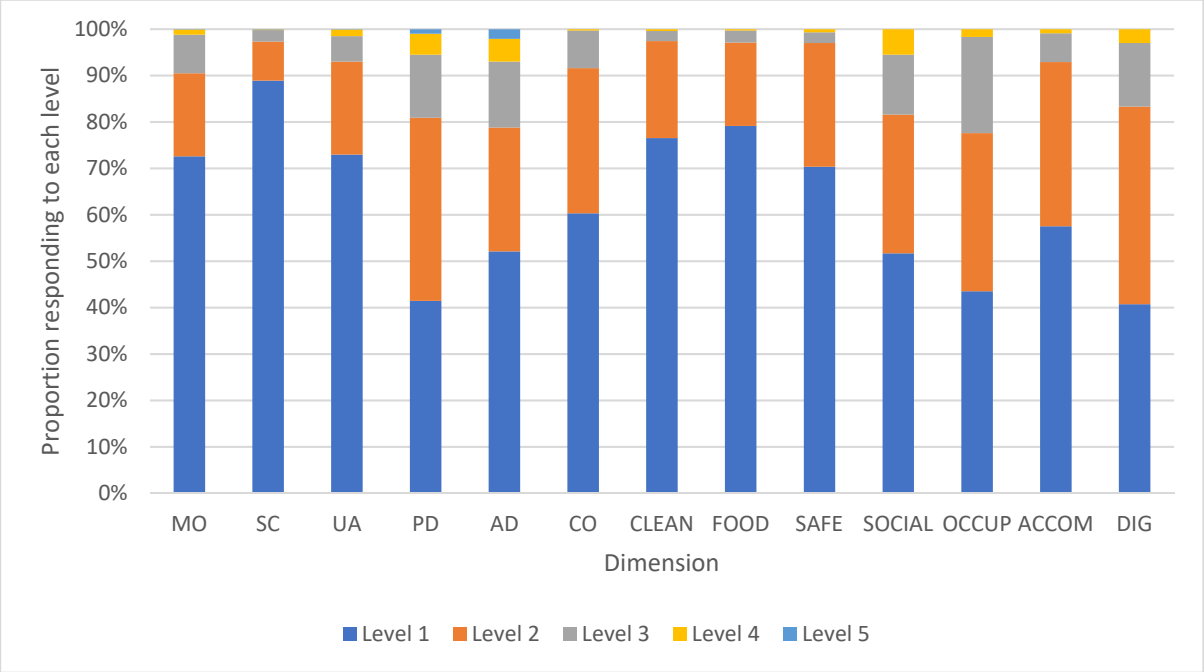
Yang Z, Feng Z, Busschbach J, Stolk E, Luo N. 2019. How Prevalent Are Implausible EQ-5D-5L Health States and How Do They Affect Valuation? A Study Combining Quantitative and Qualitative Evidence. Value Health. https://doi.org/10.1016/j.jval.2018.12.008

Figure 1: Example DCE choice set

Please consider and imagine living with the two health descriptions below. Then tell us which description you would prefer to live in.

Health description A	Health description B
You have no control over your daily life	You have no control over your daily life
You feel adequately clean and presentable	You feel clean and are able to present yourself the way you like
You don't always get adequate or timely food and drink	You don't always get adequate or timely food and drink
You feel as safe as you want	You feel as safe as you want
You have some social contact with people, but not enough	You have little social contact with people and feel socially isolated
You are able to spend time as you want, doing things you value or enjoy	You are able to do enough of the things you value or enjoy with your time
Your home is adequately clean and comfortable	Your home is adequately clean and comfortable
The way you are helped and treated completely undermines the way you think and feel about yourself	The way you are helped and treated does not affect the way you think or feel about yourself
You are unable to walk about	You have moderate problems in walking about
You have slight problems washing and dressing yourself	You have severe problems washing and dressing yourself
You have moderate problems doing your usual activities	You have slight problems doing your usual activities
You have no pain or discomfort	You have extreme pain or discomfort
You are moderately anxious or depressed	You are moderately anxious or depressed
Which do you prefer?	
Health description A	Health description B

Figure 2: Percentage of respondents answering at each level of each dimension



Note – MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression; CO: control; CL: cleanliness; FD: food and drink; SA: safety; SP: social participation; OC: occupation; AC: accommodation; DI: dignity

Table 1: Sample demographics

Characteristic	N(%)	Approx. Aust population (%)
Female	495 (50.8)	51
Age Group (years)		
Mean (sd)	46.0 (16.6)	
18 – 24	115 (11.7)	12
25 – 34	176 (18.1)	18
35 – 44	182 (18.6)	19
45 – 54	176 (18.1)	18
55 – 64	148 (15.2)	15
65+	178 (18.3)	18
Marital status		
Married/partner	585 (60.0)	48
Single/widowed/separated	390 (40.0)	52
Highest education level		
Secondary school highest level	275 (28.2)	61
Further education	700 (71.2)	39
Income (Australian $)[1]		
0 to 80,000	643 (75.8)	70
80,001 +	205 (24.2)	30
Prefer not to say	127 (13.0)	
Country of birth		
Australia	748 (76.7)	67
Other	227 (23.3)	33
Number of children		
0	554 (56.8)	N/A
1	141 (14.5)	N/A
2	163 (16.7)	N/A
3+	117 (12.0)	N/A
Health status		
Excellent	107 (11.0)	N/A
Very good	343 (35.2)	N/A
Good	339 (34.8)	N/A
Fair	154 (15.8)	N/A
Poor	32 (3.3)	N/A
Has long term condition	431 (44.2)	N/A
EQ-5D-5L utility score (m(sd))	0.773 (0.23)	N/A
ASCOT utility score (m(sd))	0.846 (0.16)	N/A
Hospitalised in last 12 months	257 (26.4)	N/A
Ever experienced serious illness:		
In self	285 (29.2)	N/A
In family	427 (43.8)	N/A
In caring for others	213 (21.9)	N/A

1 Australian dollar = 0.72 US dollars as of Oct 2018; sd: Standard Deviation; EQ-5D-5L utility calculated using Norman et al (2013); ASCOT utility calculated using Netten et al (2013); Demographics taken from the Australian Bureau of Statistics; N/A: Not available

Table 2: Conditional Logit models for overall sample, and by measure order

Parameter	Model 1: Overall model				Model 2: EQ-5D-5L first				Model 3: ASCOT first				Model 4: Restricted pooled for CLogithet	
	Coef (p)	SE	Sig (btwn)	Rank	Coef (p)	SE	Sig (btwn)	Rank	Coef (p)	SE	Sig (btwn)	Rank	Coef (p)	SE
MO2	-0.112*	0.046	0.015	2	-0.195**	0.067	0.003	1	-0.041	0.065	0.521	4	-0.102*	0.044
MO3	-0.246***	0.046	0.005		-0.417***	0.068	0.001		-0.098	0.064	0.390		-0.227***	0.045
MO4	-0.599***	0.045	<0.001		-0.795***	0.065	<0.001		-0.424***	0.062	<0.001		-0.561***	0.047
MO5	-0.799***	0.049	<0.001		-1.061***	0.072	<0.001		-0.566***	0.068	0.025		-0.747***	0.053
SC2	0.000	0.044	0.870	5	-0.044	0.063	0.404	4	**0.044**	0.061	0.534	5	0.001	0.041
SC3	-0.196***	0.045	<0.001		-0.193***	0.065	0.034		-0.189**	0.062	<0.001		-0.187***	0.043
SC4	-0.479***	0.045	<0.001		-0.465***	0.063	<0.001		-0.501***	0.062	<0.001		-0.457***	0.043
SC5	-0.516***	0.046	0.433		-0.545***	0.066	0.233		**-0.496*** **	0.064	0.939		-0.489***	0.045
UA2	-0.019	0.048	0.620	9	-0.039	0.070	0.560	9	0.000	0.067	0.917	9	-0.017	0.046
UA3	-0.024	0.047	0.925		-0.069	0.068	0.636		**0.017**	0.066	0.788		-0.020	0.045
UA4	-0.278***	0.048	<0.001		-0.206**	0.070	0.034		**-0.349*** **	0.067	<0.001		-0.267***	0.046
UA5	**-0.271*** **	0.047	0.875		-0.261***	0.068	0.397		**-0.281*** **	0.065	0.272		-0.258***	0.045
PD2	-0.275***	0.048	<0.001	1	-0.195**	0.059	0.006	2	-0.348***	0.068	<0.001	1	-0.265***	0.046
PD3	**-0.256*** **	0.050	0.705		-0.228**	0.072	0.644		**-0.281*** **	0.070	0.338		-0.245***	0.048
PD4	-0.694***	0.048	<0.001		-0.612***	0.070	<0.001		-0.779***	0.068	<0.001		-0.663***	0.049
PD5	-0.848***	0.046	0.002		-0.787***	0.066	0.015		-0.908***	0.064	0.065		-0.807***	0.048
AD2	-0.038	0.046	0.342	3	**0.014**	0.066	0.996	3	-0.087	0.064	0.191	3	-0.039	0.043
AD3	-0.199***	0.048	0.001		-0.164*	0.070	0.008		-0.234***	0.068	0.025		-0.189***	0.046
AD4	-0.574***	0.047	<0.001		-0.440***	0.069	<0.001		-0.703***	0.066	<0.001		-0.550***	0.047
AD5	-0.710***	0.048	0.004		-0.613***	0.069	0.011		-0.812***	0.067	0.094		-0.679***	0.048
CO2	-0.160***	0.042	<0.001	4	-0.122*	0.061	0.038	5	-0.202***	0.058	0.001	2	-0.155***	0.040
CO3	-0.247***	0.042	0.035		-0.164**	0.061	0.482		-0.333***	0.059	0.023		-0.238***	0.041
CO4	-0.667***	0.041	<0.001		-0.503***	0.060	<0.001		-0.834***	0.058	<0.001		-0.640***	0.042
CL2	-0.111*	0.044	0.015	8	-0.191**	0.064	0.004	8	-0.033	0.062	0.604	=6	-0.102*	0.042
CL3	-0.188***	0.043	0.077		-0.246***	0.062	0.381		-0.133*	0.060	0.102		-0.176***	0.041
CL4	-0.294***	0.044	0.010		-0.265***	0.064	0.749		-0.318***	0.062	0.002		-0.280***	0.043
FD2	-0.102*	0.045	0.035	6	-0.128	0.065	0.059	6	-0.080	0.062	0.260	8	-0.095*	0.043
FD3	-0.259***	0.043	<0.001		-0.258	0.062	0.034		-0.266***	0.060	0.002		-0.247***	0.041
FD4	-0.361***	0.046	0.033		-0.428	0.068	0.014		-0.307***	0.065	0.542		-0.341***	0.045
SA2	-0.058	0.043	0.160	7	-0.040	0.063	0.525	7	-0.084	0.061	0.134	=6	-0.056	0.041
SA3	-0.127**	0.042	0.111		-0.112	0.061	0.253		-0.153**	0.059	0.249		-0.122**	0.040
SA4	-0.330***	0.046	<0.001		-0.357***	0.061	<0.001		-0.318***	0.059	0.005		-0.313***	0.041
SP2	-0.009	0.043	0.816	13	-0.060	0.062	0.350	13	**0.042**	0.060	0.490	13	-0.007	0.041
SP3	-0.046	0.045	0.410		-0.100	0.066	0.529		**0.007**	0.063	0.573		-0.041	0.043
SP4	-0.146***	0.041	0.017		**-0.085**	0.060	0.809		-0.198***	0.058	0.001		-0.140***	0.040
OC2	-0.080	0.046	0.108	10	-0.048	0.066	0.507	10	-0.116	0.064	0.089	10	-0.078	0.044
OC3	**-0.072**	0.044	0.861		-0.102	0.063	0.378		**-0.050**	0.061	0.268		-0.068	0.042
OC4	-0.234***	0.043	<0.001		-0.201***	0.062	0.102		-0.272***	0.060	<0.001		-0.224***	0.041
AC2	**0.035**	0.043	0.437	11	**0.044**	0.062	0.451	11	**0.029**	0.060	0.688	11	0.033	0.041

AC3	-0.055	0.041	0.028		**0.015**	0.059	0.622		-0.119*	0.058	0.009		-0.055	0.039
AC4	-0.199***	0.044	<0.001		-0.157**	0.063	0.003		-0.240***	0.061	0.032		-0.191***	0.042
DI2	-0.006	0.041	0.880	12	-0.018	0.059	0.792	12	**0.005**	0.056	0.972	12	-0.004	0.038
DI3	-0.064	0.044	0.162		**0.002**	0.064	0.741		-0.128**	0.062	0.024		-0.063	0.042
DI4	-0.159***	0.042	0.023		-0.114	0.061	0.051		-0.206***	0.059	0.176		-0.151***	0.040
Order													0.094*	0.047
No obs	14,625				6,975				7,650				14,625	
LL	-8,949				-4,280				-4,608				-8,947	

Coef: Coefficient estimate; p-values for the difference between the coefficient estimate and the omitted baseline level (1) are indicated by stars: ***significant at 0.001, ** significant at 0.01; *significant at 0.05; sig(btwn): Significance between adjacent levels relative to the immediately better level; Rank: rank defined by the magnitude of the worst level.

Note – MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression; CO: control; CL: cleanliness; FD: food and drink; SA: safety; SP: social participation; OC: occupation; AC: accommodation; DI: dignity

Table 3: Models assessing heterogeneity

	Model 5: Latent class		**Model 6: Mixed logit**	
Parameter	**Class 1**	**Class 2**	**Coef**	**SD**
MO2	-0.085	-0.100	-0.176**	0.234
MO3	-0.383	-0.135	-0.344***	0.011
MO4	-0.807	-0.440	-0.834***	0.443***
MO5	-1.098	-0.592	-1.099***	0.805***
SC2	**0.048**	-0.025	-0.013	0.053
SC3	-0.269	-0.162	-0.228***	0.499***
SC4	-0.879	-0.251	-0.658***	0.469***
SC5	-1.151	**-0.054**	-0.718***	0.816***
UA2	-0.044	**0.010**	-0.065	0.208
UA3	-0.052	-0.005	**-0.036**	0.290*
UA4	-0.505	-0.099	-0.398***	0.204
UA5	-0.572	**-0.028**	**-0.389*** **	0.295**
PD2	-0.433	-0.191	-0.367***	0.298
PD3	**-0.425**	**-0.176**	**-0.349*** **	0.040
PD4	-1.382	-0.197	-0.939***	0.650***
PD5	-1.541	-0.373	-1.158***	0.766***
AD2	-0.134	**0.001**	-0.052	0.398***
AD3	-0.420	-0.065	-0.288***	0.377**
AD4	-1.439	**0.092**	-0.799***	0.782***
AD5	-1.590	-0.086	-1.002***	0.821***
CO2	-0.228	-0.126	-0.170**	0.106
CO3	-0.234	-0.254	-0.321***	0.347**
CO4	-1.102	-0.298	-0.891***	0.791***
CL2	-0.264	-0.043	-0.148**	0.096
CL3	-0.408	-0.079	-0.253***	0.324**
CL4	-0.547	-0.154	-0.388***	0.295*
FD2	-0.068	-0.150	-0.153**	0.319**
FD3	-0.378	-0.181	-0.337***	0.235
FD4	-0.435	-0.327	-0.486***	0.311
SA2	-0.234	**0.063**	-0.077	0.222
SA3	-0.410	**0.070**	-0.164**	0.087
SA4	-0.693	-0.077	-0.451***	0.364***
SP2	-0.118	**0.073**	-0.013	0.174
SP3	-0.118	**0.012**	-0.065	0.192
SP4	-0.369	**0.042**	-0.216***	0.355***
OC2	-0.157	-0.057	-0.113	0.262
OC3	-0.191	**-0.018**	-0.126*	0.292
OC4	-0.473	-0.099	-0.317***	0.482***
AC2	**0.025**	**0.038**	**0.033**	0.108
AC3	-0.141	**0.016**	-0.078	0.368***
AC4	-0.315	-0.113	-0.274***	0.285*
DI2	-0.039	**0.000**	-0.015	0.295**
DI3	-0.183	**0.018**	-0.080	0.069
DI4	-0.361	-0.004	-0.208***	0.437***
Demographic				
Age Cat (18-60 and 60+)	0.758	0.000		
Gender	0.746	0.000		

Has long term condition	0.192	0.000		
Class Share	0.550	0.450		
N obs	14,625		14,625	
LL	-8,730		-8,809	
BIC	18,093			

Note – MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression; CO: control; CL: cleanliness; FD: food and drink; SA: safety; SP: social participation; OC: occupation; AC: accommodation; DI: dignity

Supplementary Appendix 1: Information provided to respondents during the survey

DCE choice set instruction page (presented prior to starting the questions)

Title: Section B – Making choices between options
You will now be presented with 16 questions.

In this set of questions you will see two different descriptions of health and social care. Your task is to imagine living with the problems described. Then tell us which of the descriptions you would prefer to live in.

Please remember that there are no right or wrong answers. Some of the health descriptions may be difficult for you to imagine - just do the best you can. We are interested in your views, because it will help us to understand what aspects of quality of life are most important to people.

Let's start the questions now.

Text above DCE choice sets (all tasks)

Please consider and imagine living with the two health descriptions below. Then tell us which description you would prefer to live in.

Appendix 2: One block of choice sets from the design

Pair no	Pair code	EQ 1	EQ 2	EQ 3	EQ 4	EQ 5	ASCO T 1	ASCO T 2	ASCOT 3	ASCO T 4	ASCO T 5	ASCO T 6	ASCO T 8	ASCO T 9
5	A	5	2	3	1	3	4	2	3	1	3	1	2	4
5	B	3	4	2	5	3	4	1	3	1	4	2	2	2
6	A	3	3	5	3	4	1	4	4	3	2	3	4	3
6	B	4	3	2	3	3	4	2	2	4	2	2	4	3
50	A	5	5	5	5	3	1	2	3	3	4	3	1	2
50	B	5	4	1	1	3	1	4	3	4	3	3	2	1
91	A	4	2	1	4	5	1	2	1	4	1	2	4	3
91	B	4	1	5	4	1	4	2	2	2	1	4	4	1
146	A	1	2	3	2	4	2	3	4	2	4	1	3	1
146	B	1	4	1	3	1	2	3	1	1	4	4	1	1
173	A	2	2	5	5	5	3	4	2	2	4	3	4	1
173	B	1	3	5	5	4	4	1	3	4	4	3	2	1
184	A	1	3	1	4	5	2	2	3	1	1	3	1	4
184	B	1	1	3	4	1	4	2	4	2	2	2	1	4
201	A	1	3	2	3	2	3	4	2	2	2	2	2	3
201	B	5	3	1	5	2	4	3	4	2	3	2	2	1
206	A	1	3	5	5	1	1	1	3	3	3	2	1	2
206	B	3	1	1	5	1	1	3	1	1	3	4	3	2
210	A	1	1	1	5	4	3	3	2	3	3	2	4	2
210	B	1	2	2	2	4	2	4	2	2	3	4	1	2
221	A	1	5	2	2	3	2	1	2	2	3	3	1	3
221	B	3	4	2	5	2	2	1	1	4	4	3	3	3
263	A	5	1	2	2	3	2	3	2	3	2	2	1	1
263	B	4	2	3	4	2	2	3	2	1	4	2	3	1
264	A	3	5	1	2	2	4	3	4	3	3	2	2	1
264	B	1	5	2	2	4	4	2	3	1	3	2	3	2

268	A	2	4	3	5	3	4	3	4	3	2	1	3	2
268	B	2	1	4	5	3	3	3	1	2	1	1	2	1
272	A	2	4	4	4	4	2	3	4	4	4	2	3	2
272	B	2	4	3	4	4	4	2	4	1	3	1	1	1

Notes and level descriptions:

EQ 1: Mobility

1: You have no problems in walking about
2: You have slight problems in walking about
3: You have moderate problems in walking about
4: You have severe problems in walking about
5: You are unable to walk about

EQ 2: Self-care

1: You have no problems washing and dressing yourself
2: You have slight problems washing and dressing yourself
3: You have moderate problems washing and dressing yourself
4: You have severe problems washing and dressing yourself
5: You are unable to wash and dress yourself

EQ 3: Usual activities

1: You have no problems doing your usual activities
2: You have slight problems doing your usual activities
3: You have moderate problems doing your usual activities
4: You have severe problems doing your usual activities
5: You are unable to do your usual activities

EQ 4: Pain/discomfort

1: You have no pain or discomfort
2: You have slight pain or discomfort

3: You have moderate pain or discomfort
4: You have severe pain or discomfort
5: You have extreme pain or discomfort

EQ 5: Anxiety/depression
1: You are not anxious or depressed
2: You are slightly anxious or depressed
3: You are moderately anxious or depressed
4: You are severely anxious or depressed
5: You are extremely anxious or depressed

ASCOT 1: Control
1: You have as much control over your daily life as you want
2: You have adequate control over your daily life
3: You have some control over your daily life, but not enough
4: You have no control over your daily life

ASCOT 2: Personal cleanliness
1: You feel clean and are able to present yourself the way you like
2: You feel adequately clean and presentable
3: You feel less than adequately clean or presentable
4: You don't feel at all clean or presentable

ASCOT 3: Food and Drink
1: You get all the food and drink you like when you want
2: You get adequate food and drink at okay times
3: You don't always get adequate or timely food and drink
4: You don't always get adequate or timely food and drink, and think there is a risk to your health

ASCOT 4: Personal safety
1: You feel as safe as you want
2: You feel adequately safe, but not as safe as you would like
3: You feel less than adequately safe
4: You don't at all feel safe

ASCOT 5: Social participation
1: You have as much social contact as you want with people you like
2: You have adequate social contact with people
3: You have some social contact with people, but not enough
4: You have little social contact with people and feel socially isolated

ASCOT 6: Occupation
1: You are able to spend time as you want, doing things you value or enjoy
2: You are able to do enough of the things you value or enjoy with your time
3: You do some of the things you value or enjoy with your time, but not enough
4: You don't do anything you value or enjoy with your time

ASCOT 7: Accommodation cleanliness and comfort
1: Your home is as clean and comfortable as you want
2: Your home is adequately clean and comfortable
3: Your home is not quite clean or comfortable enough
4: Your home is not at all clean or comfortable

ASCOT 8: Dignity
1: The way you are helped and treated makes you think and feel better about yourself
2: The way you are helped and treated does not affect the way you think or feel about yourself
3: The way you are helped and treated sometimes undermines the way you think and feel about yourself

4: The way you are helped and treated completely undermines the way you think and feel about yourself