MONASH University
Business and Economics

Centre for Health Economics

Research Paper 2011 (63)

Modelling the Utility of Health States with the Assessment of Quality of Life (AQoL) 8D Instrument: Overview and Utility Scoring Algorithm

Professor Jeff Richardson
Foundation Director, Centre for Health Economics
Monash University

Dr Kompal Sinha
Research Fellow, Centre for Health Economics
Monash University

Angelo Iezzi
Research Fellow, Centre for Health Economics
Monash University

Dr Munir A Khan
Research Fellow, Centre for Health Economics
Monash University

June 2011
Centre for Health Economics

ISSN 1833-1173
ISBN 1 921187 62 X

Correspondence:

Jeff Richardson
Centre for Health Economics
Faculty of Business and Economics
Monash University Vic 3800
Australia

Ph: +61 3 9905 0754 Fax: +61 3 9905 8344
Email: Jeffrey.Richardson@monash.edu

Abstract

The Assessment of Quality of Life (AQoL) 8D is a 35 item, 8 dimensional, multi-attribute utility instrument (MAUI). It was derived using psychometric methods for achieving content validity, ie sensitivity to variation in the states that are of interest. The present paper outlines the methods, data collection and analyses for deriving the utility weights formula. The results are used to describe the distribution of scores for the instrument and its dimensions, to obtain reliability (test-retest and Cronbach's alpha) coefficients and to conduct preliminary tests of instrument validity. The instrument performs well with respect to these criteria.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Boxes

Modelling the Utility of Health States with the Assessment of Quality of Life (AQoL) 8D Instrument Overview and Utility Scoring Algorithm

1 Introduction

Economic evaluation of health related activities increasingly uses the Quality Adjusted Life Year (QALY) as a unit of outcome where QALYs are defined as the product of life years and health state utilities. Calculation of utilities, in turn, requires two tasks; first, describing the relevant health state and secondly, the use of a preference measurement technique such as the time trade-off (TTO) or standard gamble (SG) to estimate the utility of the health state.

Two approaches to this latter two stage procedure have been used, namely holistic (or 'composite') and multi-attribute utility (MAU) measurement (Torrance 1986). With the first of these, a scenario or vignette is constructed which describes the health state (Step 1). This is subsequently 'scaled' (Step 2) to elicit a utility score for the whole scenario. With the second, a generic multi-attribute utility instrument (MAUI) is created which is capable of describing and scoring a wide range of health states. The MAU questionnaire is distributed to the individuals whose health states are to be assessed. The questionnaire consists of a series of items, generally a ‘stem’ (‘you are energetic’) plus multiple response categories (‘all of the time...’; ‘none of the time...’) Answers are scored using the MAU algorithm, a predetermined formula for converting the description into a numerical value.

To date, only a handful of generic instruments have attempted to measure utility; viz, the UK Rosser-Kind Index (Rosser 1993), the US QWB (Kaplan and Anderson 1996), the Canadian HUI instruments (Feeny, Torrance et al. 1996), the Finnish 15D (Sintonen and Pekurinen 1993) and the UK SF-6D and SF12D (Brazier, Harper et al. 1998; Brazier, Roberts et al. 2002) and the EQ-5D (EuroQol) (Kind 1996). The construction methods employed have varied and the resulting instruments differ significantly in terms of the type and scope of questions, the scaling techniques and the models used to derive the scoring formula (Brazier, Ratcliffe et al. 2007; Richardson, McKie et al. 2011).

For many purposes the existence of differences between QoL instruments is unproblematical. These purposes include classification, description and, with caveats, ordinal ranking. Differences would be of relatively little concern if instruments were for specialised purposes which did not involve comparison between them. However the *raison d'être* of generic MAUI’s is that they allow comparison between all health states and a person’s preference for these states relative to life extension. If all the extant instruments were successful in this latter task they would produce

identical scores for identical health states. However multi instrument comparisons have not found this. In the largest studies Hawthorne et al. (2001) found that in paired comparisons of five instrument comparison (EQ-5D, HUI 3, SF-6D, 15D, AQoL-4D) on average only 54 percent of variation in one instrument was explained by another instrument. More recently Fryback et al. (2010) found a lower average association of 0.47 in a comparison of four instruments (EQ-5D, QWB, HUI 3, SF-6D)

There are potentially several reasons for this. First, different MAUI have been scaled using different methods (TTO, SG, RS, rank order). Secondly, the models used to interpolate and extrapolate from observed values to the full set of health states defined by the MAU descriptive systems have varied, compressing, combining or attenuating values in different ways over different domains of health. Third, the descriptive system itself may lack content validity. This is defined as an instrument's coverage or inclusion of a representative sample of elements from the construct (Streiner and Norman 2003). For example a test of numerical ability would lack content validity if it included items relating to addition and multiplication but none relating to subtraction and division. The former might correlate with the latter but unless the correlation was perfect the instrument would produce variable predictions of performance depending upon the type of results it was predicting. In a health related context, MAU preference weights for omitted elements might be zero, but this must be determined empirically.

There has been relatively little investigation of the reasons for the low correlation between MAUI and particularly of the hypotheses that it is attributable in large part to the instrument's descriptive system. Construction of the AQoL-8D was motivated by this possibility and particularly by the minimal descriptions of mental health in the major instruments. Items explicitly referring to mental health are summarised in Box 1. Only the 15D and AQoL-6D have more than a single dedicated item. In principle it is possible that the general terms in other MAUI items capture the content of mental health indirectly; that as any of the multiple facets of mental health change the number ticking different item response categories moves up and down in exact proportion to the changed QoL. But this possibility has been assumed and never demonstrated. If it were empirically true then the numerous multi item instruments measuring mental health QoL have contained largely redundant questions.

The AQoL-8D project commenced with the hypothesis that generic properties assumed in the existing instruments have been exaggerated and that content invalidity is a contributory factor to the low correspondence between instrument scores. If this hypothesis is true then the construction of MAUI encounters a dilemma. The Decision Analytic version of MAU theory assumes the items are structurally orthogonal (von Winterfeldt and Edwards 1986). Correlation between items results in 'redundancy' or double counting of content and therefore instruments cannot use correlated items or dimensions. However, language is imprecise and single items seldom 'capture' a full concept. For this reason in psychometric theory it is assumed that a minimum of 3-4 items will be needed to satisfactorily triangulate a concept. But this implies content redundancy.

Box 1 Explicit mental health components of the major MAUI

EQ-5D	
'Anxiety Depression'	(1) Not anxious or depressed; *to* (3) Extremely anxious or depressed (Dolan, Gudex et al. 1995; Brooks, Rabin et al. 2005)
HUI 2	
'Emotion'	(1) Generally happy and free from worry; *to* (5) Extremely fretful angry irritable anxious or depressed, usually requiring hospitalisation or psychiatric institutional care (Torrance, Feeny et al. 1996)
HUI 3	
'Emotion'	(1) Happy and interested in life; *to* (5) Unhappy that life is not worthwhile (Feeny, Furlong et al. 2002)
SF-6D	(version 1: reduced from SF36)
'Mental Health'	(1) You feel tense or downhearted and low *none of the time; to* (5) You feel tense or downhearted *most of the time* (Brazier, Roberts et al. 2002)
SF-6D	(reduced from SF-12)
'Mental Health'	(1) You feel downhearted and low *none of the time; to* (5) You feel tense or downhearted *most of the time* (Brazier and Roberts 2004)
QWB	
Symptom list includes:	Spells of feeling upset, being depressed or of crying; *and* Excessive worry or anxiety (Kaplan and Anderson 1996)
15D	
'Depression'	(1) I do not feel at all sad, melancholy or depressed; *to* (5) I feel extremely sad, melancholy or depressed
'Distress'	(1) I do not feel at all anxious stressed or nervous; *to* (5) I feel extremely anxious stressed or nervous (Sintonen and Pekurinen 1989)
AQoL-4D	
'Feeling'	(1) I do not feel anxious, worried or depressed; *to* (5) I am extremely anxious, worried or depressed (Hawthorne, Richardson et al. 1999)
AQoL-6D	
'Despair'	How often did you feel despair over the last 7 days (1) Never; *to* (5) All the time
'Worry'	How often did you feel worried (1) Never; *to* (5) All the time
'Sadness'	How often did you feel sad (1) Never; *to* (5) Nearly all the time
'Tranquillity'	Were you calm and tranquil or agitated (1) Always calm and tranquil; *to* (5) Always agitated (Richardson, Day et al. 2004)

In principle, this dilemma may be resolved using econometric procedures. Holistic health states created from an instrument's descriptive system are evaluated and regressed upon item scores or item response levels (as with EQ-5D and SF-6D). The regression methodology attributes variation in the holistic utility scores in a way which minimises the square of the error term and ensures that predicted utility is equal to actual utility at its mean value; that is, the effect of double counting is mitigated as the scoring equation 'passes through the data'. In practice the econometric solution has limitations. Large numbers of correlated items result in multi colinearity and this limits the size of instruments which, in turn, risks content invalidity.

A series of AQoL instruments have grappled with this dilemma (Box 2). The first, AQoL-4D (initially called 'AQoL') sought content validity within dimensions by allowing redundancy, but orthogonality between dimensions. These were combined, as with the HUI instruments, with a multiplicative model. AQoL-6D increased content (sensitivity) by adding 2 dimensions and 8 items. Redundancy resulted, and a second stage econometric correction was introduced as described in Section 3 below. AQoL-7D and 8D increased redundancy by added dimensions relating to vision and mental health respectively and adopted the same strategy. The evidence from these instruments supports the view that changing content alters utility values (Richardson, Elsworth et al. 2011)

Box 2 AQoL instruments

AQoL-4D Originally called 'AQoL' (Hawthorne, Richardson et al. 1999)**:** Initially a 5 dimension 15 item instrument. Dimensions were illness, independent living, social relationships, physical senses, psychological wellbeing. Illness was subsequently deleted. Utilities were combined with a multi-level model using multiplicative models for dimensions and an overall multiplicative model to combine them.

AQoL 8 (Hawthorne 2009) An 8 item (Brief) instrument which removes one item per dimension from AQoL-4D and imputes their values from remaining items.

AQoL-6D (Richardson, Day et al. 2004): A 6 dimensional 20 item instrument. Pain and coping were added to AqoL-4Das separate dimensions. Mental health and independent living items were increased from 3 to 4. Utility weights were constructed as for AQoL-4D but with an econometric adjustment for the final algorithm.

AQoL-7D (Misajon, Hawthorne et al. 2005): A 7 dimension 26 item instrument which adds an explicit dimension for vision (VisQoL). Scaling was carried out as for AQoL-6D. (Richardson, Iezzi et al. 2011)

AQoL-8D (Richardson, Elsworth et al. 2011): The 8 dimensional 35 item instrument shown in Figure 1 with utility weights explained in this paper.

The present paper proceeds as follows. In Section 2 below the methods used for constructing the AQoL-8D descriptive system are summarised and the instrument presented along with its psychometric properties. They are explained in greater detail in Richardson, Elsworth et al. (2011). Methods used for deriving utility weights are outlined in Section 3 and the results relating to the utility formula are presented in the following section. In Section 5 the reliability and validity of the model are addressed and tests of these properties presented.

2 The AQoL-8D Descriptive System

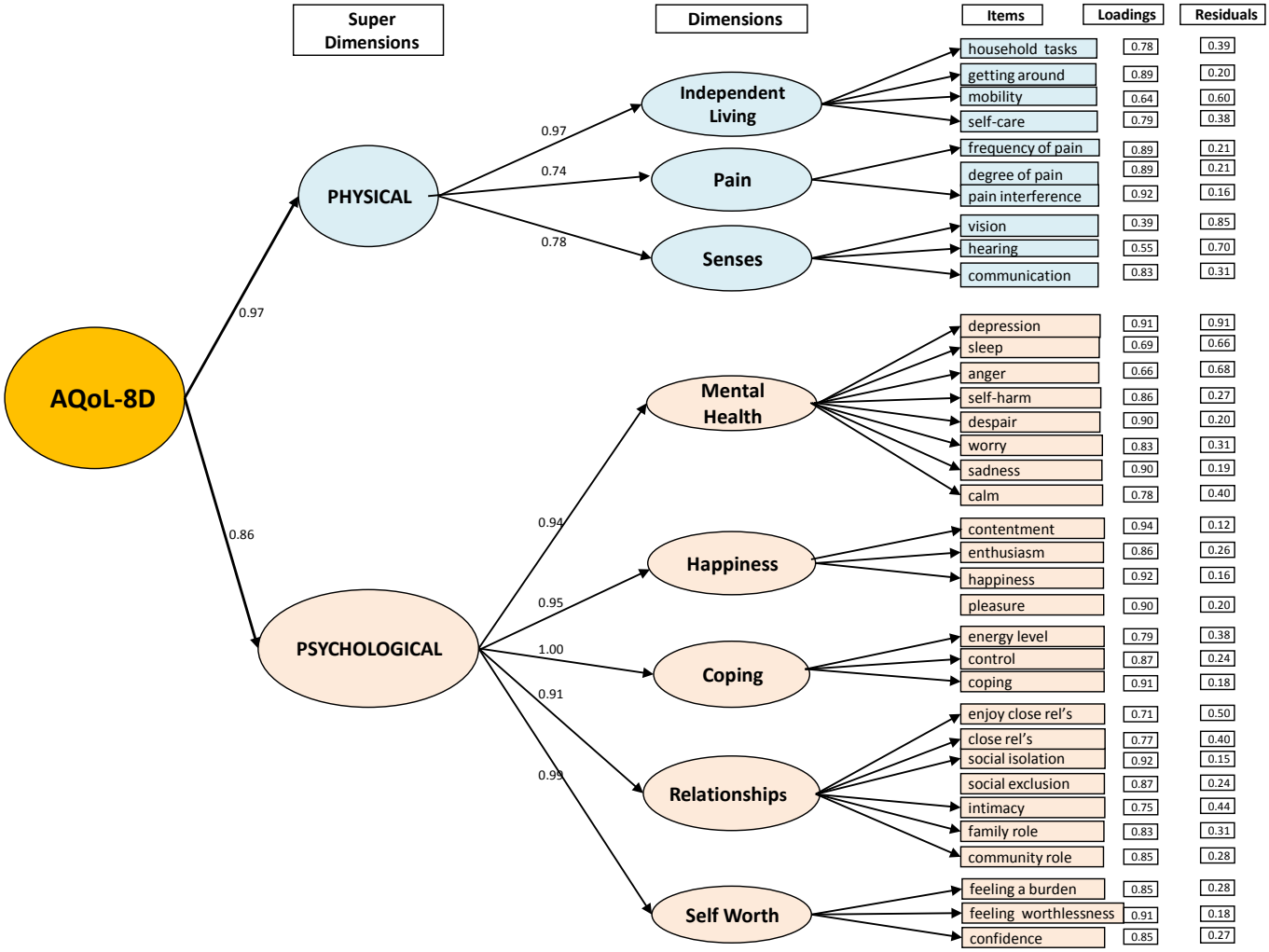
The psychometric theory of instrument construction is discussed by a number of authors (Fayers and Machin 2000; Streiner and Norman 2003; O'Connor 2004). First, an overall concept (or theory) of health must be selected. This is operationalised by determining the dimensions of the concept and describing the postulated elements of each dimension with multiple items. These are initially analysed using expert input to ensure an adequate sampling of key elements and then edited for grammatical clarity and consistency. The surviving elements in the 'item bank' are administered to a representative sample of the relevant population (the construction survey) to obtain the relationship between item responses. These are examined statistically to achieve item reduction while retaining the integrity of the dimension content and the overall construct. In practice, this implies the interplay of statistical and substantive arguments. Items are excluded if the resulting scale is theoretically inconsistent irrespective of the statistical properties. Conversely, when there are strong *a priori* grounds for including items they must also meet minimum statistical standards or be rejected.

AQoL-8D adopted the same concept of health – handicap – as the previous AQoL instruments. This basic conceptualisation was supplemented, where necessary, with elements of disability and impairment. The concept was operationalised by postulating dimensions and elements of QoL additional to those included in AQoL-6D, and locating or creating items which described them. Specifically, a literature search related to mental health QoL was undertaken which identified 8 commonly used instruments. Key items and elements from each of these were collated and this provided the starting point for the item bank. Four focus groups were conducted with 29 participants consisting of mental health patients and carers. Transcripts were examined to identify new issues and to guide the grammatical construction of new items. These were subject to linguistic, logical and content analyses by the research team in a series of meetings.

The initial data bank of 250 items was reduced to 135 items and the construction survey successfully administered by mail and by personal interview to 716 individuals (514 mental health patients by interview and 202 members of the public by mail). The larger number of patients was included as the survey's primary purpose was to observe the relationship between individual item responses and, for the new module, it was necessary for the majority of individuals to have experienced the health states of interest.

Item selection and scale validation were conducted in the tradition of classical test theory using unrestricted and restricted factor analyses. (These correspond with the less descriptively accurate terms 'exploratory' and 'confirmatory' factor analyses). Restricted analysis was required as the overall structure of the instrument was determined on substantive grounds. However as argued by McDonald (2005) it is desirable to accompany this with unrestricted analysis to check whether anything has been missed in this analysis and, if necessary, to revise hypotheses. Three models were constructed, tested and modified with data from the construction survey. First, the 32 item 7 dimensional PsyQoL was a relatively unrestricted best fit model of mental health. Secondly, the PsyQoL-Brief was a 22 item reduced form of PsyQoL. This was combined with the 20 item AQoL-6D and further reduced to form the 35 item AQoL-8D. This is shown with its psychometric properties in Figure 1. Psychometric analysis also demonstrated that the dimensions combine to form two 'super dimensions' 'physical' (independent living, senses, pain) and 'mental health' (self-worth, coping, relationships, happiness). Details of the construction are given in Richardson, Elsworth et el. (2011).

Figure 1 The AQoL-8D model (35 items)



Numbers attached to arrows refer to factor loadings
CFI= 0.974; FLI = 0.972; RMSEA = 0.073; WRMR = 1.64
Source: Richardson , Elsworth et al (2011)

The final AQoL-8D is a hierarchical causal model. The diagnostic statistics indicate a satisfactory statistical fit; that is, it is possible to confidently reject the hypothesis that variation in items cannot be explained by variation in dimension scores and dimension scores by a single, latent variable – HRQoL conceptualised in terms of handicap.

The AQoL-8D questionnaire is reproduced in Appendix 5.

3 Methods

Estimation of formulae for converting item responses into dimension values and a final utility score requires a pre-specified model and a population survey to obtain the data needed for 'scaling' the model. For reasons discussed earlier, AQoL-6D adopted two stage modelling. Stage 1 used the multiplicative model recommended by MAU theory to combine the large number of items (dimensions) into Stage 1 estimates. The construction of the descriptive system outlined above ensured (in fact, required) non-orthogonality. Consequently, to remove the effects of this (double counting) and to test for structural dependence between items, a second stage econometric correction was carried out. This two stage procedure was used to obtain estimates for each of the 8 dimensions from the items and to combine these to obtain a single utility score for the overall construct.

Modelling: The stage 1 multiplicative model is similar to equation 1.

$$V = \prod_{i=1}^{n} V_i \qquad \text{... (1)}$$

where V_i are the values of the items (dimensions) to be combined and V is the multiplicative score. The actual model is somewhat more flexible. It is calculated using disvalues rather than values and these are adjusted for the relative importance of each of the n items (dimensions) using an importance weight. This results in equation 2 in which DV,(x_{ij}) is the disvalue of level $_j$ in item (dimension) i, w_i is the corresponding item (dimension) importance weights and k is the overall scaling constant which is similar to the requirement in an additive model that the dimension weights sum to unity. It is obtained by solving equation 3.

$$DV = \frac{1}{k}\left[\prod_{i=1}^{n}\left[1 + kw_i DV_i\left(x_{ij}\right)\right] - 1\right] \qquad \text{... (2)}$$

$$k = \prod_{i=1}^{n}(1 + kw_i) - 1 \qquad \text{... (3)}$$

The relationship between value and disvalue is given in equation 4.

$$V^* = 1 - DV^* \qquad \text{... (4)}$$

As noted, this multiplicative model was applied at two levels; first, to combine items into dimensions and, secondly, to combine dimensions into the overall AQoL score.

The second stage adjustment drew upon the econometric approach to scaling which could not be used directly because of the size of the instrument. A limited number of MA health states were directly assessed using the TTO and equation 5 which was the most successful functional form identified using the analysis for the AQoL-6D.

$$TTO = MULT^{x}$$

$$x = a_0 + \sum_j b_i I_i + \sum_i \sum_j c_{ij} I_i I_j \qquad \text{... (5)}$$

where MULT = multiplicative model

a_i b_i c_{ij} = parameters

I_i = item dimension score for item (dimension) i

$I_i I_j$ = dimension I_i score times Dimension I_j score

Weights Survey: The relationship between the data collected and the stages of analysis is shown in Figure 2. Commencing on the left of this figure, A data were used to attach weights to each of the response levels of each of the 35 items in the instrument. The scale was from item best (=1) to item worst (=0). Following the recommendation of decision analytic theory item worst disutilities were used as item weights (B data).

These were evaluated using a VAS rating of the item worst outcome on a scale calibrated from dimension best (all item bests) to dimension worst (all item worsts). Multi-attribute health states were evaluated for each dimension using both the VAS and TTO ('ED data'). These were used, first, to construct a transformation between VAS and TTO and secondly as the dependent variable in the econometric adjustment of dimension scores. Dimension weights were obtained from TTO valuations of the dimension worst health state (all item worsts) on an AQoL best-death scale ('C data'). Multi-attribute health states spanning all combinations of dimensions were similarly obtained from TTO valuations on an AQoL best-death scale ('E data') for stage 2 adjustment. Details of the collection evaluation methods, instruments and props are reported in Iezzi and Richardson et al. (2009).

Multiplicative dimension scores were obtained by inserting values in equation 2: A data equate with $DV_i(x_{ij})$; B data with w_i and k is solved by inserting w_i in equation 3. Final dimension scores were obtained inserting the dimension multiplicative score, $MULT_i$, in equation 5 along with items scores (I_i data).

A multiplicative AQoL score was obtained by inserting these dimension scores and dimension weights – C data – in equation 2 and C data in equation 3 to solve for k. The final AQoL-8D algorithm was estimated by inserting this latter multiplicative score as MULT in equation 5 and setting dimension scores equal to I_i.

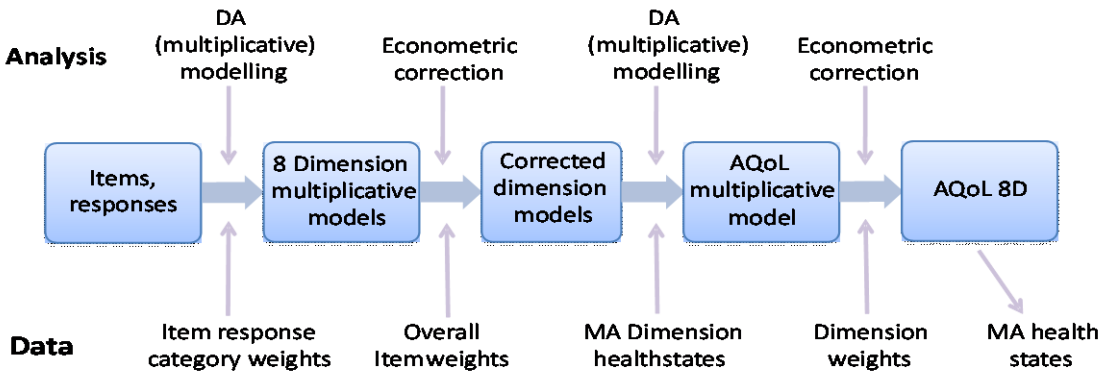
Data Collection: The population sample was drawn from a computer readable phone directory, using a stratified, clustered two-stage design, similar to Hawthorne et al.'s (1999) procedures in the AQoL-4D validation study. Postcodes were used as the primary sampling units with the selection based on the Australian Bureau of Statistics Socio Economic Index for Areas (SEIFA) and the probability of selection proportionate to the population size (to reduce the effect of socio-economic confounding). Telephone subscribers (above 18) were sampled, contacted by letter and subsequently by telephone. The use of post-codes as the primary sampling unit meant that informants were fairly tightly clustered, minimizing travel costs.

People with mental disorders ('neuroses' and psychotic disorders) were accessed via a number of channels including St Vincent's Hospital Mental Health Service, The Melbourne Clinic (private hospital) and centres for post-traumatic stress disorder. Treatment providers were approached to assist in the recruitment of people with non-psychotic disorders. Case managers and treating clinicians were approached to ensure people were well enough to participate in the study. Informed consent was obtained from all subjects.

Past experience indicated that to obtain a satisfactory response rate it would be necessary to make a payment in compensation for out of pocket costs. Before interview, respondents completed the AQoL-8D questionnaire, VAS questions and personal details. These were checked at the beginning of interviews. For the public these took place at the AQoL research office. Patients were all interviewed in their treatment service site by one of a team of interviewers experienced in the application of HRQoL instruments, and the use of TTO. The order of items was randomly varied to check for framing effects.

The survey comprised two strata: (a) a sample of the Australian population; and (b) a sample of those with mental health disorders. The ideal target sample size was based on the minimum acceptable sampling errors — in this case 5% was set, resulting in a target of 400 cases in each strata (Hoinville, Jowell et al. 1977).

Figure 2 Summary of data and analysis for the scaling of AQoL-8D



Label	A data	B data	ED data	C data	E data
Metric	VAS	VAS	VAS/TTO	TTO	TTO

4 Results

Respondents to the weights (scaling) survey are described in Tables 1 and 2. In sum, 670 individuals participated, 323 patients and 347 members of the public (Table 1). Those above 65 were underrepresented relative to the Australian population, but in other age cohorts satisfactory numbers were obtained. From Table 2 lower SEIFA groups are overrepresented but subsequent analysis did not indicate differences between the top two and bottom three groups.

Table 1 Public and patient respondents to weights survey by age

Age Group	Public			Patient			Total	Percent	
	Male	Female	Total	Male	Female	Total		Total	Aust
18 to 24 years	22	31	53	22	23	45	98	14.7	12.5
25 to 34 years	38	36	74	35	30	65	139	20.2	17.7
35 to 44 years	32	35	67	35	43	78	145	21.8	19.5
45 to 54 years	26	43	69	41	29	70	139	20.3	18.3
55 to 64 years	30	42	72	44	15	59	131	20.1	14.5
65 years +	7	5	12	3	3	6	18	2.9	17.5
Total	155	192	347	180	143	323	670	100	100

Table 2 Public and patient respondents to weights survey by SEIFA location

SEIFA Group	Respondent		Total	Percent	
	Public	Patient		Total	Aust
1	18	30	48	8.4	12.6
2	1	14	15	2.6	20.8
3	9	38	47	8.2	25.1
4	95	108	203	35.4	20.5
5	136	124	260	45.4	21.0
Total	259	314	573	100	100

Missing = 97 (patients in hospital with no fixed address)

Data obtained from respondents are described and analysed in detail in Iezzi and Richardson (2009). With few exceptions all item response categories were used by respondents and resulting frequency distributions indicated the absence of content 'gaps'. VAS ratings of item responses were very similar by the public and patient respondents. There were no significant differences in mean scores for 14 of the 35 items and for another 10 only one response level differed significantly. Where differences occurred they were small and, contrary to the conventional wisdom that adaptation results in higher QoL assessments by those affected than by the general public, patient scores were always lower when differences were significant. These differences occurred in the mental health, self-worth and happiness dimensions.

Dimension models: Item worst scores and their use in the calculation of item weights are reported in Table 3. Inserting the weights for each dimension in equation 2, in conjunction with the dimension scaling constant calculated from equation 3, creates the multiplicative dimension formulae reported in Box 3.

The 174 'within dimension' multi attribute health states created for the econometric correction of dimensions resulted in 2787 observations (ED data), an average of 348 per dimension.

Box 3 Multiplicative equations for dimensions

General formula	$MULT_d = \frac{1}{k}\left[1 - \prod_{i=1}^{n}(1 - kw_i DV_i)\right]; k_d > 0$
Ind Living	$MULT(IL) = 1.02\left[1 - (1 - 0.54dv_1)(1. - 0.59dv_2)(1 - 0.87dv_3)(1 - 0.81dv_4)\right]$
Happiness	$MULT(HAP) = 1.01\left[1 - (1 - 0.66dv_5)(1. - 0.60dv_6)(1 - 0.77dv_7)(1 - 0.77dv_8)\right]$
Mental Health	$MULT(MH) = 1.00\left[\begin{matrix}1 - (1. - 0.74dv_9)(1 - 0.63dv_{10})(1 - 0.70dv_{11})(1. - 0.86dv_{12}) \\ (1 - 0.74dv_{13})(1 - 0.66dv_{14})(1. - 0.69dv_{15})(1 - 0.76dv_{16})\end{matrix}\right]$
Coping	$MULT(COP) = 1.03\left[1 - (1 - 0.60dv_{17})(1. - 0.69dv_{18})(1 - 0.78dv_{19})\right]$
Relationships	$MULT(REL) = 1.00\left[\begin{matrix}1 - (1 - 0.67dv_{20})(1 - 0.69dv_{21})(1 - 0.69dv_{22})(1 - 0.74dv_{23}) \\ (1 - 0.65dv_{24})(1 - 0.67dv_{25})(1 - 0.66dv_{26})\end{matrix}\right]$
Self Worth	$MULT(SW) = 1.02\left[1 - (1 - 0.69dv_{27})(1 - 0.74dv_{28})(1 - 0.70dv_{29})\right]$
Pain	$MULT(PAIN)1.04\left[1 - (1 - 0.69dv_{30})(1 - 0.67dv_{31})(1 - 0.68dv_{32})\right]$
Senses	$MULT(SEN) = 1.04\left[1 - (1 - 0.63dv_{33})(1 - 0.61dv_{34})(1 - 0.69dv_{35})\right]$

Notes

W = the conversion factor between the 0-1 (death, full health) model.

Box 4 Econometric correction to multiplicative scores

Ind Living	$DV(IL) = 0.123dv_4 + 0.525DMULT(IL)$
Happiness	$DV(HAP) = 0.176dv_5 + 0.175dv_6 + 0.198dv_7 + 0.106dv_8 + 0.137MULT(HAP)$
Mental health	$DV(MH) = 0.103dv_9 + 0.092dv_{11} + 0.169dv_{12} + 0.443MULT(MH)$
Coping	$DV(COP) = 0.204dv_{17} + 0.090dv_{18} + 0.198dv_{19} + 0.232MULT(COP)$
Relationships	$DV(REL) = 0.176dv_{20} - 0.103dv_{22} + 0.107dv_{23} - 0.079dv_{25} - 0.109dv_{26} + 0.533MULT(REL)$
Self Worth	$DV(SW) = 0.184dv_{27} + 0.115dv_{28} + 0.258dv_{29} + 0.190MULT(SW)$
Pain	$DV(PAIN) = 0.205dv_{30} + 0.277dv_{31} + 0.205dv_{32} + 0.205.MULT(PAIN)$
Senses	$DV(SEN) = 0.084dv_{33} + 0.113dv_{35} + 0.582.MULT(SEN)$

AQoL disutilities may be transformed into utility scores using the equation $U_i = 1 - DU$ where U_i and DU_i are utility and disutility respectively.

As reported in Richardson and Iezzi (2009) , the frequency distributions of these data span the full range of scores for each dimension, ie from 1.00 to 0.00 (dimension best to dimension worst). For each dimension these data were used as dependent variables and explained by the predicted multiplicative score and by the average item scores for the health states. Various models were tested but simple linear regressions proved to be the most satisfactory. They are reported in Box 4.

AQoL-8D model: TTO scores for dimension worst health states and their use in calculating dimension weights are reported in Table 3. For 7 of the 8 dimensions public and patient respondents gave significantly different scores and for the last dimension, Senses, the difference was significant at the 10 percent level. As previously, in every case, it was the public which assigned the higher score, suggesting that the public is limited in its capacity to imagine social and mental ill health which they have not experienced. Inserting the dimension weights in equations 2 and 3 results in the multiplicative models for the two super dimensions and for the overall AQoL-8D. These are reported in Box 5.

Table 3 Item weights for use in dimension models

Item / Dimension	(-) k_d (1) w_i(2)	= wt_i(3)	Item / Dimension	(-) k_i(1) w_i(2)	= wt_i(3)
Ind Living			Coping		
1	(0.995) * (.55)	= 0.54	1	(0.971) * (.62)	= 0.60
2	(0.995) * (.60)	= 0.59	2	(0.971) * (.71)	= 0.69
3	(0.995) * (.87)	= 0.87	3	(0.971) * (.80)	= 0.78
4	(0.995) * (.81)	= 0.81			
Happiness			Pain		
1	(0.992) * (.66)	= 0.65	1	(0.966) * (.71)	= 0.69
2	(0.992) * (.60)	= 0.60	2	(0.966) * (.69)	= 0.67
3	(0.992) * (.77)	= 0.76	3	(0.966) * (.70)	= 0.68
4	(0.992) * (.78)	= 0.77			
Mental Health			Relationships		
1	(0.999) * (.74)	= 0.74	1	(0.999) * (.67)	= 0.67`
2	(0.999) * (.63)	= 0.63	2	(0.999) * (.69)	= 0.69
3	(0.999) * (.70)	= 0.69	3	(0.999) * (.69)	= 0.69
4	(0.999) * (.86)	= 0.86	4	(0.999) * (.74)	= 0.74
5	(0.999) * (.74)	= 0.74	5	(0.999) * (.65)	= 0.65
6	(0.999) * (.66)	= 0.66	6	(0.999) * (.67)	= 0.67
7	(0.999) * (.69)	= 0.69	7	(0.999) * (.66)	= 0.66
8	(0.999) * (.76)	= 0.76			
Self Worth			Senses		
1	(0.976) * (.71)	= 0.69	1	(0.955) * (.66)	= 0.63
2	(0.976) * (.76)	= 0.74	2	(0.955) * (.64)	= 0.61
3	(0.976) * (.72)	= 0.70	3	(0.955) * (.72)	= 0.69

Notes

(1) Scaling constant, k, derived from equation (3).

(2) Item worst disvalue, measured on a VAS scale where 0= no disvalue; 1 = disvalue when all of the dimension items are at their worst level.

(3) Item weight w_i.

For the final econometric correction 3178 TTO observations were obtained from 370 MA health states (E-data) which spanned all dimensions and combinations of dimensions. Details of the construction, definitions and descriptive statistics are reported in Richardson, Iezzi et al. (2009). The frequency distribution of the TTO scores is shown in Figure 3. The scores were used in the final econometric correction, equation 5.

Table 4 Regression results for equation 5[(1)]

TTO (E-health states) on multiplicative (stage 1) estimates (MEAN DATA): (exponential regressions) (n=370)

Independent variables[(1)]	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Constant	0.145	0.25	0.19	0.33	0.29	0.15
IL D1				ns	Ns	
HAP D2				-0.04**	ns	0.14*
MH (D3)				-0.06	ns	0.11
COP (D4)				-0.06		-0.05
REL (D5)				0.05*	0.41*	0.20
SW (D6)				-0.07	-0.35*	-
PAIN (D7)				-0.04	-0.21*	-0.04
SEN (D8)				-0.06		
SUPER (Phys)		-0.08				
SUPER (MH)		-0.18				
SUPER (P)			-0.29			
IL * REL					-0.24***	
IL * SW					0.20***	
IL * PAIN					0.10***	
* HAP * REL					-0.31	-0.29*
MH * PAIN					0.17**	
MH * SEN					-0.10**	-0.11
COP * SEN					-0.05	
MH * SW						-13
R^2 (adj)	0.84	0.87	0.87	0.88	0.89	0.89
F	1944	800	1264	299	166	329
ROOT MSE	0.24	0.21	0.21	0.20	0.20	0.19
Regression: TTO on AQoL-8D						
a	-0.21	0.005 ns	0.05	0.05 ns	0.01 ns	0.02 ns
b	1.41	1.03	0.93	1.03	1.00	0.97
R^2	0.62	0.66	0.76	0.68	0.72	0.71
F	606	717	739	802	928	901
ROOT MSE	0.12	0.11	0.05	0.11	0.10	0.10

Notes

All coefficients are significant at 0.1 percent level unless shown

* Significant at 1 percent level

** Significant at 5 percent level

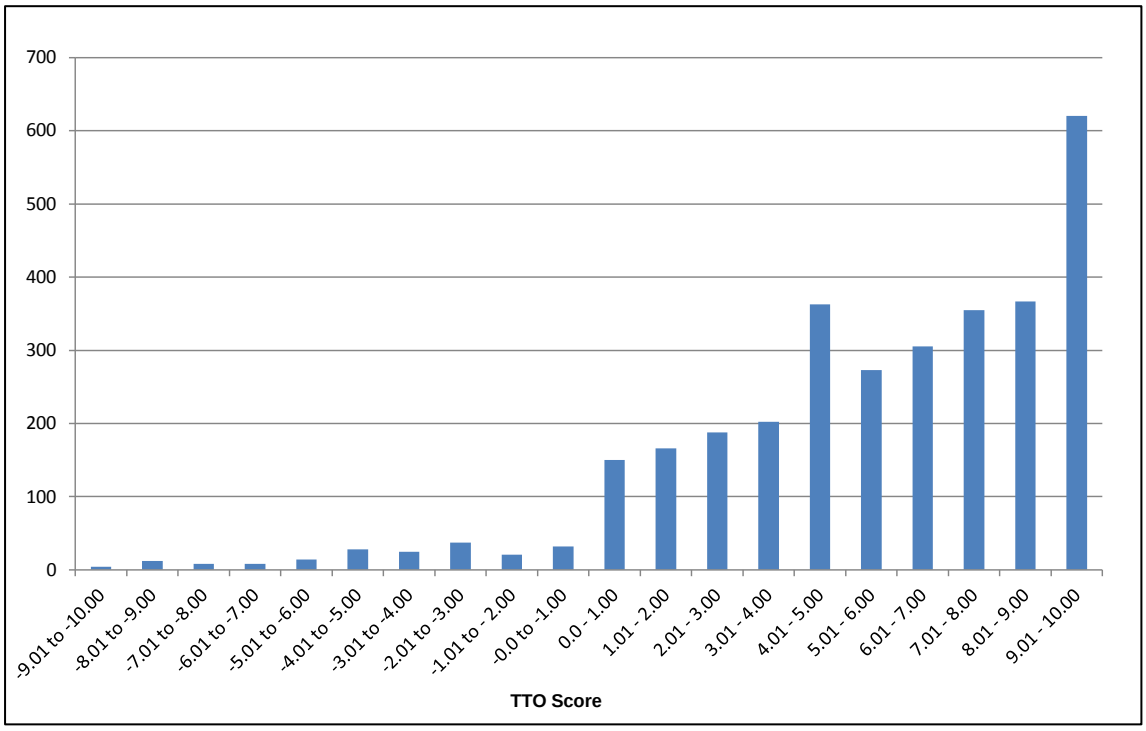
*** Significant at 10 percent level.

(1) The dependent variable is log TTO; the independent variables are all multiplied by log (Mult.AQoL). The equations are therefore of the form TTO = (Multi AQoL)$^{\text{constant} + \Sigma \text{ bi Variable i}}$ See Appendix 1.

Box 5 AQoL-8D and super dimension (SD) multiplicative scores

SD (Physical)	$SD\,(\text{PHYSICAL}) = 1-(1-0.65dud1_{EcCor})(1-0.72dud7_{EcCor})(1-0.69dud8_{EcCor})$
SD (Mental)	$SD\,(\text{MENTAL}) = 1-(1-0.71dud2_{EcCor})(1-0.74dud3_{EcCor})(1-0.64dud4_{EcCor})$ $(1-0.72dud5_{EcCor})(1-0.66dud6_{EcCor})$
AQoL	$\text{MULT AQoL} = 0.999[1-(1-0.65DU(IL))(1-0.71DU(HAP))(1-0.74DU(MH))(1-0.64DU(COP))$ $(1-0.72DU(REL))(1-0.66DU(SW))(1-0.72DU(PAIN))(1-0.69DU(SEN))]$

Figure 3 Frequency distribution of multi-attribute (MA) TTO scores, n = 3178

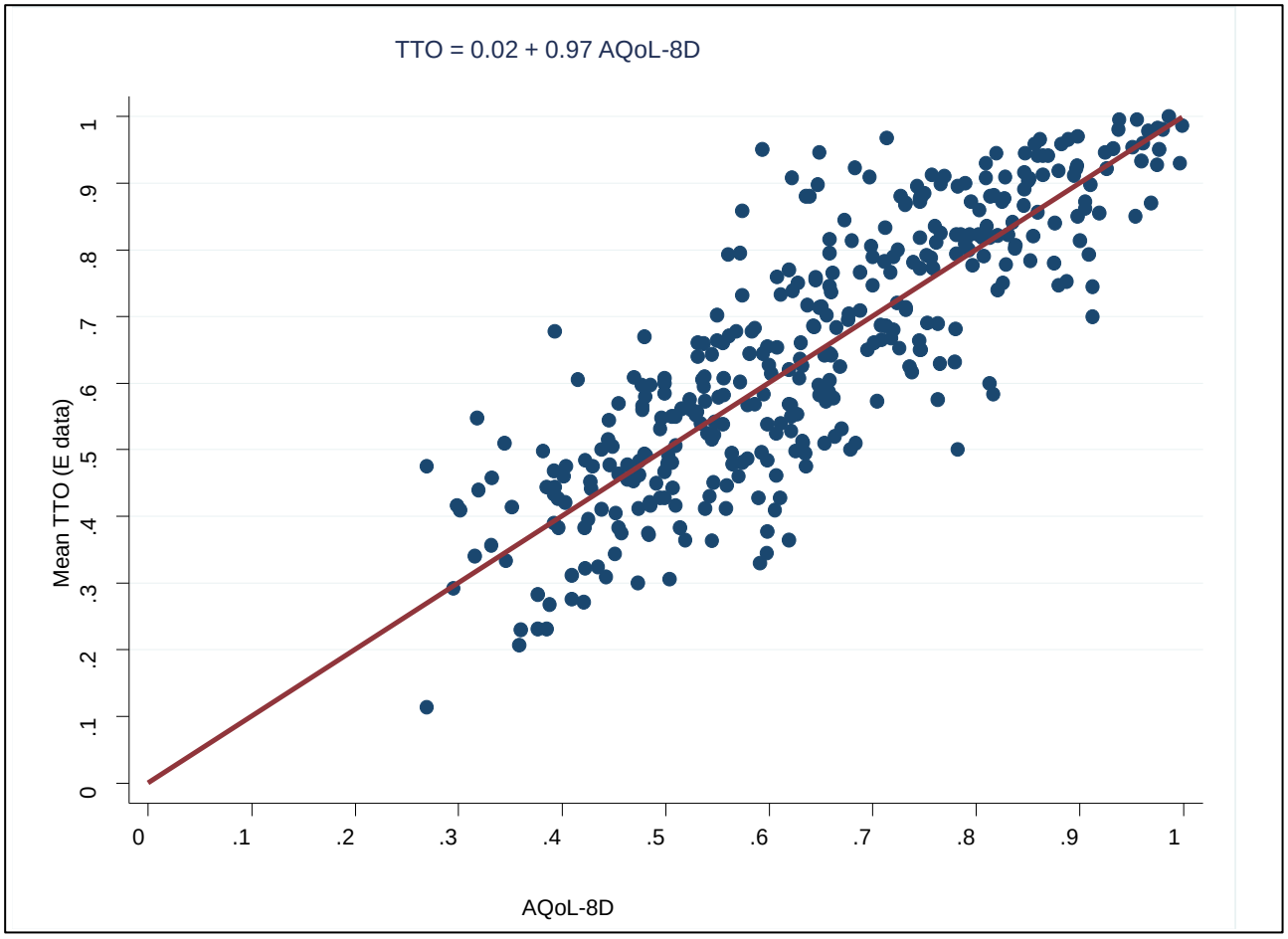


This was estimated with both individual data and with mean scores for the 370 MA states (E data). Different models were created by the inclusion and exclusion of variables. With individual data, model R^2 values varied from 0.43 to 0.48 (F values from 209 to 3037). With mean data, R^2 were between 0.84 and 0.89 (F values from 299 to 1944).

As a test of model bias predicted AQoL-8D utilities on the 370 directly elicited mean TTO scores were regressed upon the predicted scores using linear OLS. For an unbiased estimate of the TTO the resulting equation should be of the form: TTO = 0.00 + 1.00 AQoL-8D. Models estimated from mean data consistently performed better on this test. The constant term from individual data regressions were all statistically and quantitatively significant, varying from 0.15 to 0.33. In contrast the constant terms from mean data were insignificant in all but one of the reported equations with the b coefficient varying from 0.93 to 1.03 with the same single exception. For this reason the final model was chosen from the latter group.

Results using mean data to estimate equation 5 and the second stage linear regressions are reported in Table 4. Results using individual data are reported in Appendix 1. Model 1, (no dimension or interaction terms) produced a good overall fit (R^2 = 0.84) but resulted in poor second stage prediction (TTO = -0.21 + 1.41 AQoL-8D). Using super dimensions as independent variables resulted in a good fit (models 2, 3), with model 2 having the greatest overall predictive power. However the quantitatively large negative coefficient in the interaction term (super Phys* super MH) results in a non-trivial chance of perverse prediction above the upper limit (U = 1.00). Including all dimensions (model 4) also resulted in a good fit and nearly perfect second stage prediction but included insignificant variables. Various combinations of dimensions and dimension interactions generally resulted in good overall fit, the best of which are shown as equations 5 and 6. While these are similar with respect to their predictive power in stage 2 the large number of marginally significant interaction terms in model 5 resulted in the selection of equation 6 as the final algorithm. Results of the second stage predictive model are shown in Figure 4. When AQoL-8D = 1.00 the predicted TTO in this model is 0.9985.

Figure 4 Mean analysis: actual TTO against predicted TTO



5 Properties of the Model

Summary statistics for the AQoL-8D and its dimensions are reported in Table 5. The AQoL-8D is a measure of utility and numerical values have meaning in relation to best health (1.00) and death (0.00). In contrast dimension specific scales, relative scores and changes in scores are meaningful but simple, numerical comparisons across scales are not. The AQoL-8D frequency distribution for 1149 members of the general population, selected to be representative of the Australian age, gender, socio economic profile is presented in Figure 5. Dimension frequencies are shown in Appendix 2. By the standards of HR-MAU instruments the distributions are very satisfactory, with an absence of significant ceiling or floor effects which are negligible except for Independent Living, Pain and Senses. The numbers achieving maximum scores for these dimensions (33, 36 and 21 percent respectively) are unsurprising in the general, non-institutional population.

Table 5 AQoL-8D Summary statistics for general public

Dimension and Instrument	Mean[(1)]	SE	Min	Max	N
AQoL-8D Utility	.82	.004	.17	1.00	1149
Dimensions					
Happiness	.78	.004	.23	1.00	1149
Mental Health	.62	.004	.21	1.00	1149
Coping	.80	.004	.28	1.00	1149
Relationships	.74	.005	.31	1.00	1149
Self-Worth	.84	.004	.25	1.00	1149
Mental Health (Super Dim)	.42	.005	.03	1.00	1149
Independent Living	.90	.004	.37	1.00	1149
Pain	.87	.004	.31	1.00	1149
Senses	.88	.003	.32	1.00	1149
Physical (Super Dim)	.79	.005	.22	1.00	1149

Notes

(1) Dimension scores are each on a separate scale and comparison between them is therefore involved.

Figure 5 AQoL-8D and dimension score, general population

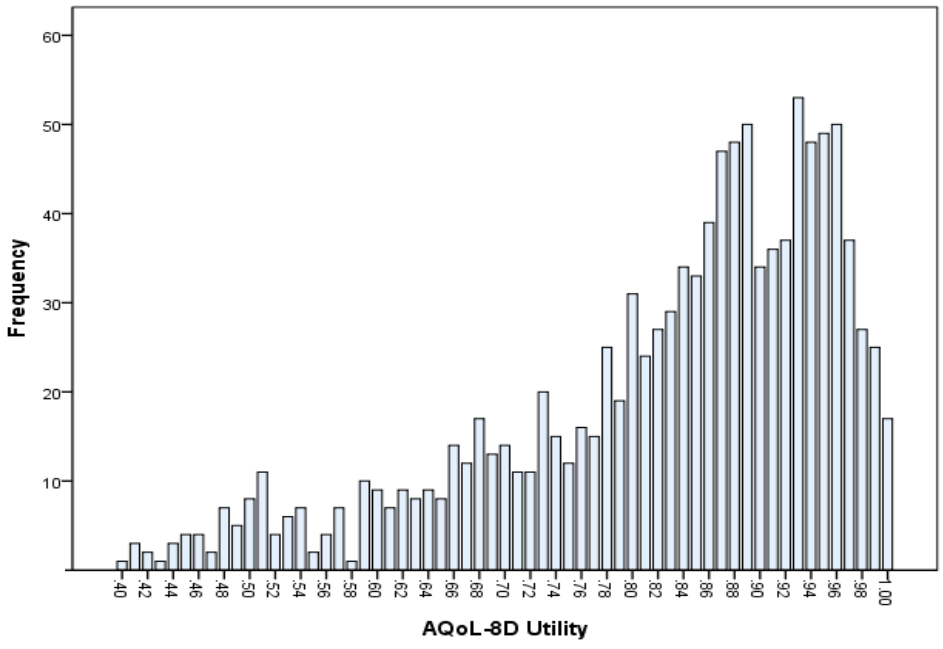


Figure 6 and 7 display scores by gender and by age respectively. Data are reproduced in Appendix 3 and 4. While there is no significant difference in overall AQoL-8D utilities between men and women, dimension scores indicate that women score more highly on independent living and the physical super dimension but significantly worse on the mental health dimension, self worth and the mental health super dimension.

The pattern of scores by respondent's age differs by dimension (Figure 7). The first three dimensions which constitute the physical super dimension decline with age and particularly sense perceptions. The physical super dimension score similarly declines with age. In contrast the mental health super dimension rises. Happiness remains unchanged which is consistent with psychological observation and the theory of homeostasis based upon a genetically determined 'set point'. Other dimensions improve with age and in particular Mental Health and Coping. Scores for Relationships dip in the middle age group and Self Worth increases only marginally with age. These results are also consistent with independent evidence.

Reliability: Two standard measures of reliability are reported below. These are the Cronbach alpha measure of the internal consistency of a scale and the intra-class correlation between repeated applications of the instrument, ie the test – retest reliability. To obtain results for these, the AQoL-8D was administered on three occasions to an additional 224 members of the public selected to be representative of the Australian age-gender-education profile. The second and third completions occurred two and four weeks after the baseline data were collected.

Table 6 Reliability: Cronbach's alpha

Observation	Hap	MH	Cop	Rel	SW	Pain	Sen	IL	MSD	PDS	AQoL-8D
1	0.902	0.890	0.908	0.852	0.899	0.710	0.676	0.729	0.939	0.829	0.954
2	0.919	0.892	0.919	0.860	0.921	0.732	0.545	0.721	0.935	0.806	0.952
3	0.932	0.900	0.927	0.874	0.944	0.717	0.615	0.719	0.934	0.811	0.974
Ave	0.918	0.894	0.918	0.879	0.921	0.720	0.612	0.723	0.936	0.815	0.960

Table 7 Test-Retest reliability: intra class correlation[a] coefficients (ICC)

	Hap	MH	Cop	Rel	SW	IL	Pain	Sen	MSD	PDS	AQoL-8D
Base – 2 weeks	.858	.870	.816	.783	.863	.861	.851	.644	.902	.842	.907
Base 1 month	.846	.844	.795	.733	.848	.856	.851	.691	.863	.874	.894

Table 5 indicates a very high Cronbach alpha coefficient for AQoL-8D. This is, in part, attributable to the size of the instrument as the alpha rises with the number of items. It may also reflect redundancy, although this is improbable given the psychometric properties obtained during the instrument construction. Coefficients for the individual scales are also high with the exception of pain and senses which are complex dimensions with possibly too few items to be good stand-alone scales.

Test-retest correlation coefficients are also high for all of the dimensions with the exception of senses. The ICC for super dimensions and the overall AQoL are very high indicating good instrument reliability.

Validity: The first test of instrument validity was reported above, viz, that there is 'internal predictive power'. The final AQoL-8D algorithm has unbiased prediction of TTO values and explains a significant proportion of the variance. As reported above the correlation of 0.84 (R^2 equals 0.71) is imperfect but satisfactory for an instrument with the breadth of the AQoL-8D.

A second test is a comparison of scores for the general and patient respondents. The latter consists of the mental health patients recruited for this study and a group of morbidly obese patients awaiting bariatric surgery. Results of the comparison are presented in Figure 8. Differences between the general and mental health respondents are statistically and quantitatively significant for every dimension and for the overall AQoL-8D. The magnitude of the difference is largest for the Mental Health dimensions and particularly Self Worth. It is smallest for Senses. The obese patients also had lower AQoL and dimension scores than the general population. Differences were smaller than for mental health patients but statistically and quantitatively significant with the sole exception of Senses. An important result highlighted in Figure 8 is that the psycho-social dimensions included in the AQoL-8D were of considerable importance for this group despite the *prima facie* expectation that they might be of peripheral relevance. In fact the percentage reduction in the mental health super dimension score was greater than the percentage reduction in the physical super dimension score.

Overall, these results demonstrate strong discriminative validity by the AQoL-8D and its dimensions.

Figure 6 AQoL-8D and dimension score by gender, general population

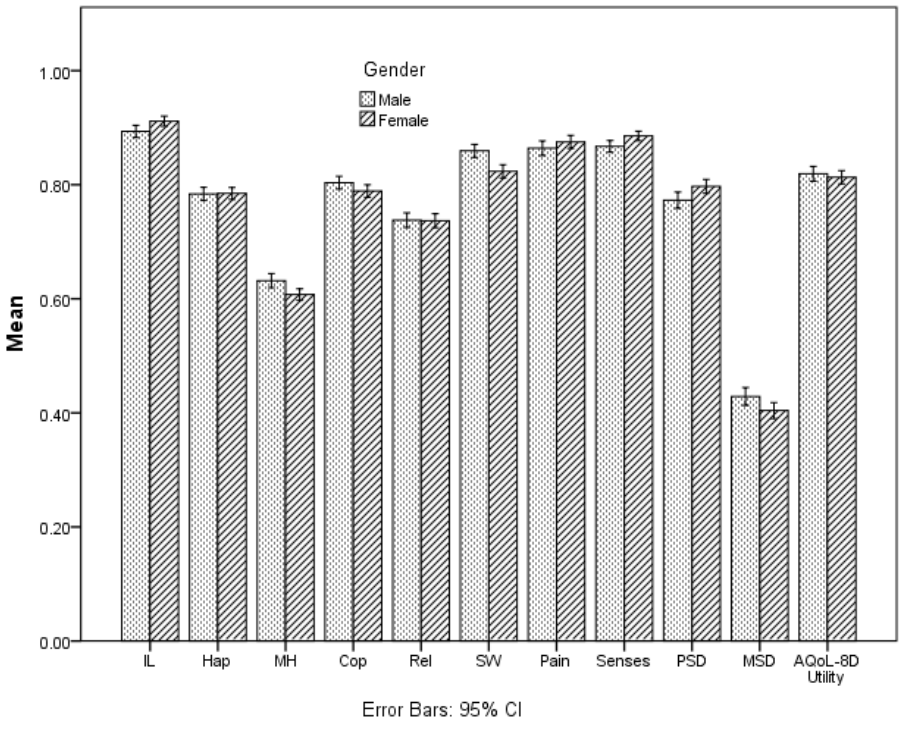


Figure 7 AQoL-8D and dimension scores by age group, general population n = 884

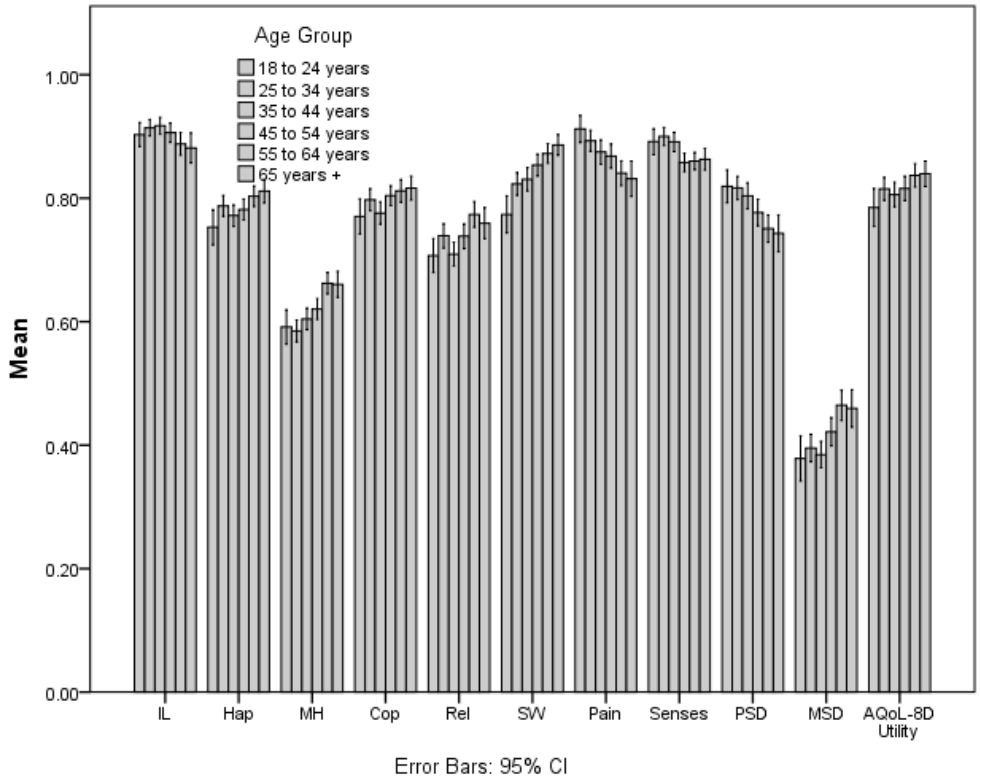
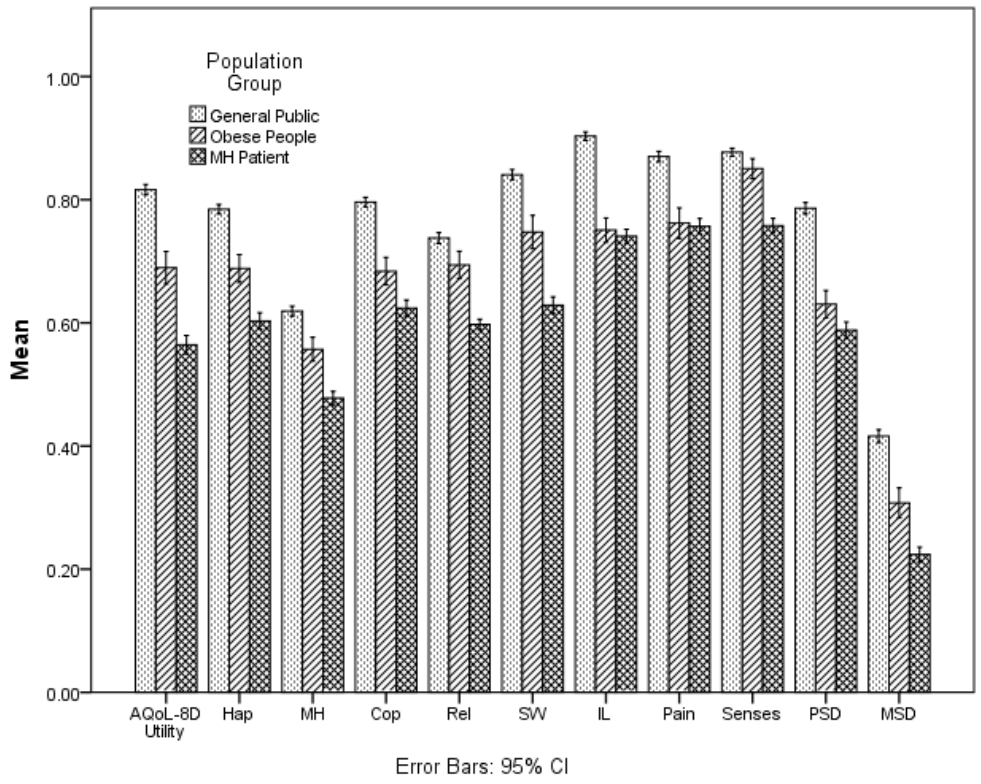


Figure 8 Validation: Comparison of AQoL-8D and dimension scores, general population, mental health and obese patients



Notes

n (general population) = 884; n (obese patients) = 196; n (mental health patients) = 832.

6 Conclusions

By the standards of MAUI the AQoL-8D is large. Nevertheless, records from its on-line application indicate an average completion time of 5.4 minutes. The size of the instrument is a reflection of the breadth of the construct and the goal of achieving content validity in the psychometric sense. The final descriptive instrument was a compromise driven by theoretical and statistical concerns and the pressure for instrument brevity. Despite this, AQoL-8D has good psychometric properties: it describes an underlying latent variable which explains a significant part of the variation in dimensions each of which in turn has content validity. From its construction the latent AQoL-8D variable is a measure of HRQoL conceptualised as handicap.

To calibrate such a large instrument and to derive a single valid utility score it was necessary to derive a methodology which overcame the problem of non-orthogonality which arose from the achievement of content validity. AQoL-8D followed the AQoL-6D in using a two stage multiplicative-econometric scaling procedure and extended the method to the constituent dimensions.

Prima facie the resulting instrument has greater content validity in the areas of social and psychological health than other MAU instruments. However this must be independently demonstrated in other studies. Preliminary results reported elsewhere are encouraging (Richardson and Khan 2009; Richardson, Elsworth et al. 2011).

AQoL-8D was constructed for use in economic evaluation studies and had to be easy to administer and score. As noted, its web version takes, on average, 5.4 minutes to complete. The questionnaire and scoring algorithm are web based and experience with its use to date indicates that, despite the complexity of its construction, its administration and use are simple.

The instrument and its algorithm are available for use on the AQoL website: http://www.aqol.com.au/

Appendix 1 Regression results for Equation 5: Individual data

TTO (E- Health States) on multiplicative (Stage 1) (individual data)

(Double log regression)

N = 4023	**M1**	**M2**	**M3**	**M4**	**M5**
AQoL (α)	0.25	0.35	0.35	0.78	0.89
IL (D1)				ns	
HAP (D2)				-0.09*	1.47
MH (D3)				-0.20	-1.17
COP (D4)				-0.09	
REL (D5)				0.05**	0.81
SW (D6)				-0.13	-1.08
PAIN (D7)				-0.18	-0.81
SENSES (D8)				-0.14	
SUPER Phys		ns			
SUPER MH		ns			
Super P * Super MH		0.73	-0.68		
IL * LS					-1.33
IL * MH					0.57**
IL* REL					-0.60*
IL * SW					0.61
IL* PAIN					0.41
HAP * MH					0.25**
HAP * REL					-0.66
MH * PAIN					0.55
MH * SP					-0.30
COP * SP					-0.08
REL * SW					0.24**
SW * SP					0.19*
MH * SW					-
R^2 ADJ	0.43	0.46	0.46	0.47	0.48
F	3037	865	1729	393	209
ROOT MSE	1.05	1.02	1.02	1.01	1.00
Regression: TTO on AQoL-8D					
a	0.15	0.33	0.15	0.30	0.33
b	1.14	0.60	1.13	0.70	0.65
R^2 (adj)	0.61	0.65	0.61	0.66	0.34
F	580	695	580	724	180
ROOT MSE	0.12	0.12	0.12	0.11	0.22

Notes

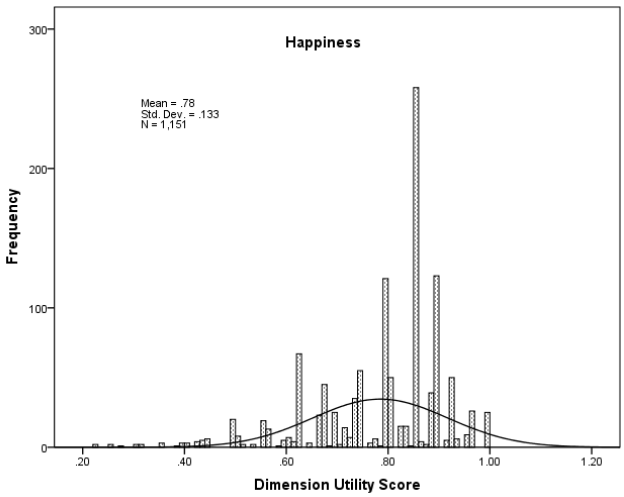
All coefficients are significant at 0.1 percent level unless shown

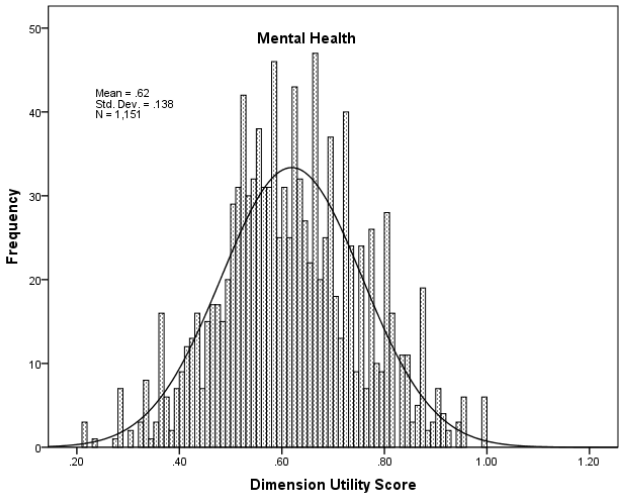
* Significant at 1 percent level

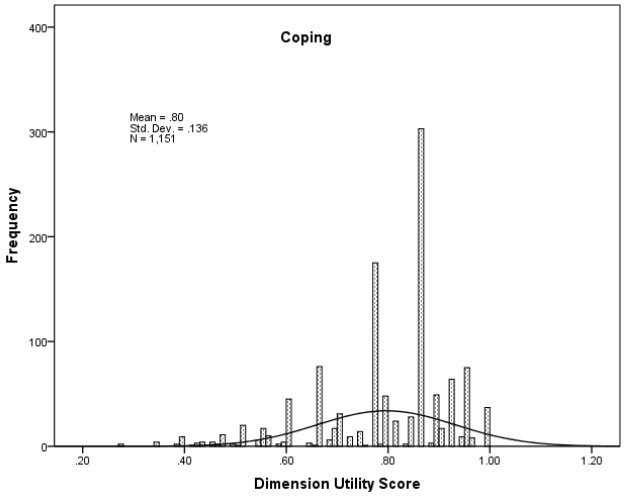
** Significant at 5 percent level

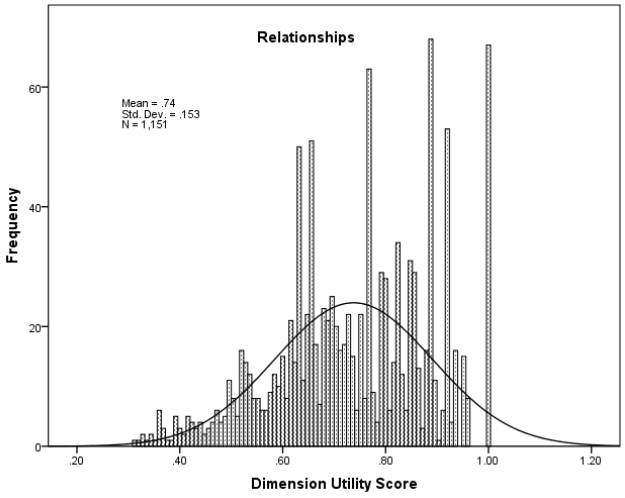
*** Significant at 10 percent level.

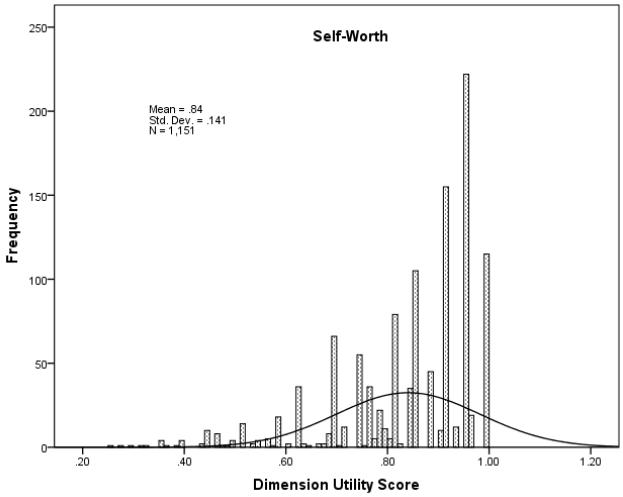
Appendix 2 Dimension frequency distributions for the general population

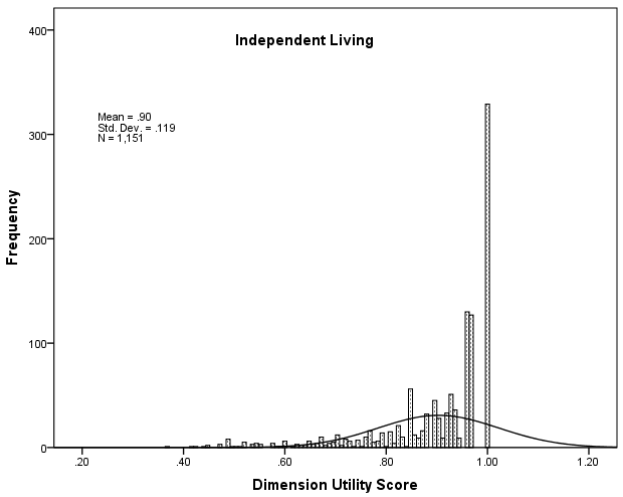


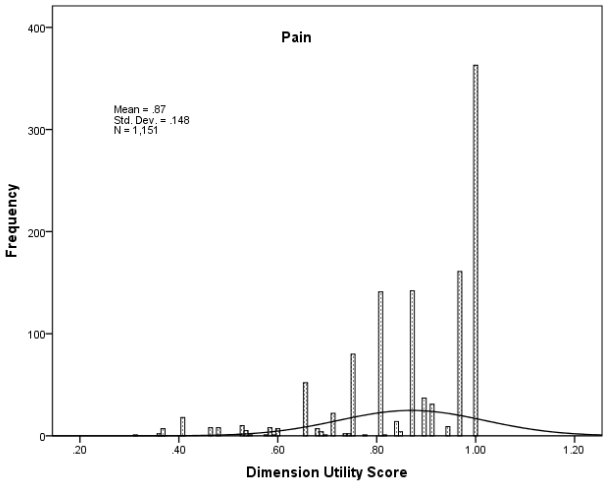


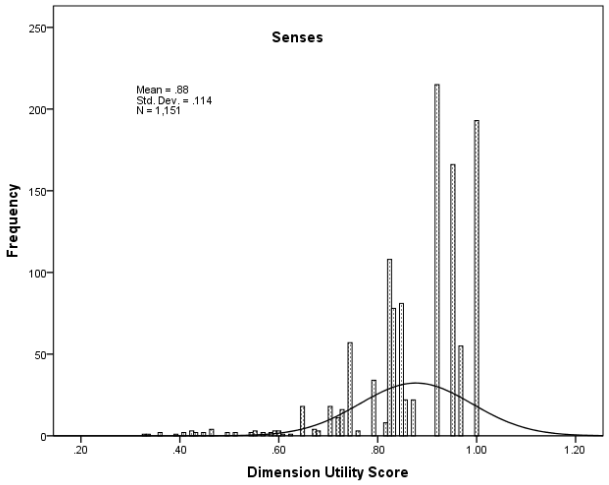


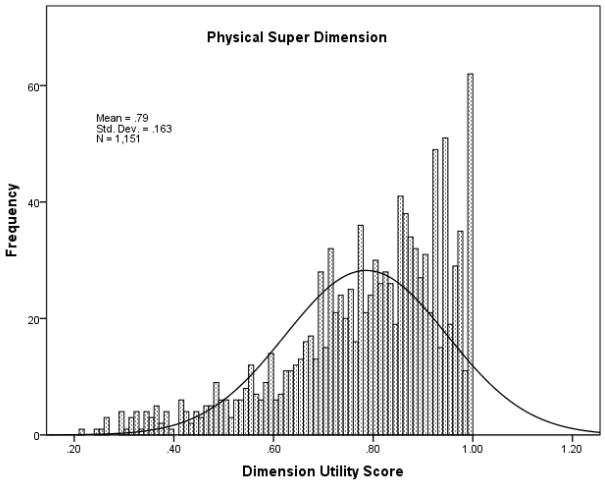


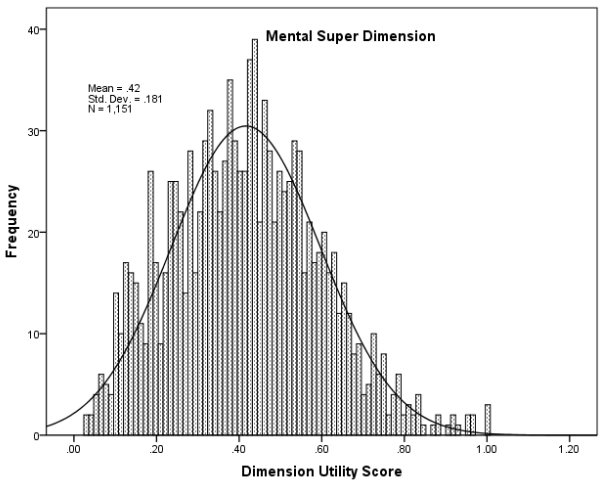





Pain
Mean = .87
Std. Dev. = .148
N = 1,151
Frequency
Dimension Utility Score


Senses
Mean = .88
Std. Dev. = .114
N = 1,151
Frequency
Dimension Utility Score


Physical Super Dimension
Mean = .79
Std. Dev. = .163
N = 1,151
Frequency
Dimension Utility Score


Mental Super Dimension
Mean = .42
Std. Dev. = .181
N = 1,151
Frequency
Dimension Utility Score

Appendix 3 Dimension scores and AQoL-8D utility score of general public by gender

Dimension and Instrument	Gender	N	Mean	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
IL	Male	542	.89	.005	.88	.90	.37	1.00
	Female	607	.91*	.005	.90	.92	.47	1.00
	Total	1149	.90	.004	.90	.91	.37	1.00
Hap	Male	542	.78	.006	.77	.80	.26	1.00
	Female	607	.78	.005	.77	.80	.23	1.00
	Total	1149	.78	.004	.78	.79	.23	1.00
MH	Male	542	.63	.006	.62	.64	.21	1.00
	Female	607	.61*	.005	.60	.62	.22	1.00
	Total	1149	.62	.004	.61	.63	.21	1.00
Cop	Male	542	.80	.006	.79	.81	.28	1.00
	Female	607	.79	.006	.78	.80	.28	1.00
	Total	1149	.80	.004	.79	.80	.28	1.00
Rel	Male	542	.74	.007	.72	.75	.33	1.00
	Female	607	.74	.006	.72	.75	.31	1.00
	Total	1149	.74	.005	.73	.75	.31	1.00
SW	Male	542	.86	.006	.85	.87	.28	1.00
	Female	607	.82*	.006	.81	.84	.25	1.00
	Total	1149	.84	.004	.83	.85	.25	1.00
Pain	Male	542	.86	.007	.85	.88	.36	1.00
	Female	607	.88	.006	.86	.89	.31	1.00
	Total	1149	.87	.004	.86	.88	.31	1.00
Senses	Male	542	.87	.005	.86	.88	.32	1.00
	Female	607	.89	.004	.88	.89	.36	1.00
	Total	1149	.88	.003	.87	.88	.32	1.00
PSD	Male	542	.77	.007	.76	.79	.22	1.00
	Female	607	.80	.006	.78	.81	.25	1.00
	Total	1149	.79	.005	.78	.79	.22	1.00
MSD	Male	542	.43	.008	.41	.44	.05	1.00
	Female	607	.40	.007	.39	.42	.03	1.00
	Total	1149	.42	.005	.41	.43	.03	1.00
AQoL-8D Utility	Male	542	.82	.007	.81	.83	.17	1.00
	Female	607	.81	.006	.80	.82	.17	1.00
	Total	1149	.82	.004	.81	.82	.17	1.00

Notes

*Significant difference (from 18-24) at 5% level.

Appendix 4 Dimension scores and AQoL-8D utility score of general public by age

Dimension and Instrument	Age Group	N	Mean	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
IL	18 to 24 years	133	.90	.010	.88	.92	.48	1.00
	25 to 34 years	221	.91	.007	.90	.93	.51	1.00
	35 to 44 years	225	.92	.007	.90	.93	.48	1.00
	45 to 54 years	240	.91	.008	.89	.92	.42	1.00
	55 to 64 years	208	.89	.009	.87	.91	.37	1.00
	65 years +	122	.88	.012	.86	.91	.42	1.00
	Total	1149	.90	.003	.90	.91	.37	1.00
Hap	18 to 24 years	133	.75	.014	.72	.78	.23	1.00
	25 to 34 years	221	.79	.009	.77	.80	.39	1.00
	35 to 44 years	225	.77	.009	.75	.79	.23	1.00
	45 to 54 years	240	.78	.009	.76	.80	.26	1.00
	55 to 64 years	208	.80*	.008	.79	.82	.39	.97
	65 years +	122	.81*	.010	.79	.83	.44	1.00
	Total	1149	.78	.004	.78	.79	.23	1.00
MH	18 to 24 years	133	.59	.014	.56	.62	.21	1.00
	25 to 34 years	221	.59	.009	.57	.60	.29	1.00
	35 to 44 years	225	.60	.009	.59	.62	.21	1.00
	45 to 54 years	240	.62	.009	.60	.64	.29	1.00
	55 to 64 years	208	.66	.009	.64*	.68	.36	.96
	65 years +	122	.66	.011	.64*	.68	.37	.96
	Total	1149	.62	.004	.61	.63	.21	1.00
Cop	18 to 24 years	133	.77	.014	.74	.80	.34	1.00
	25 to 34 years	221	.80	.009	.78	.82	.28	1.00
	35 to 44 years	225	.78	.009	.76	.79	.28	1.00
	45 to 54 years	240	.80	.008	.79	.82	.39	1.00
	55 to 64 years	208	.81	.009	.79	.83	.38	1.00
	65 years +	122	.82*	.010	.80	.84	.46	1.00
	Total	1149	.80	.004	.79	.80	.28	1.00
Rel	18 to 24 years	133	.71	.014	.68	.73	.34	1.00
	25 to 34 years	221	.74	.010	.72	.76	.31	1.00
	35 to 44 years	225	.71	.010	.69	.73	.32	1.00
	45 to 54 years	240	.74	.010	.72	.76	.36	1.00
	55 to 64 years	208	.77*	.011	.75*	.79	.33	1.00
	65 years +	122	.76*	.013	.73*	.78	.37	1.00
	Total	1149	.74	.005	.73	.75	.31	1.00
SW	18 to 24 years	133	.77	.015	.74	.80	.25	1.00
	25 to 34 years	221	.82*	.009	.80	.84	.32	1.00
	35 to 44 years	225	.83*	.010	.81	.85	.35	1.00
	45 to 54 years	240	.85*	.009	.84	.87	.28	1.00
	55 to 64 years	208	.87*	.008	.86	.89	.48	1.00
	65 years +	122	.89*	.009	.87	.90	.46	1.00

Dimension and Instrument	Age Group	N	Mean	SE	95% Confidence Interval for Mean		Min	Max
					Lower Bound	Upper Bound		
	Total	1149	.84	.004	.83	.85	.25	1.00
Pain	18 to 24 years	133	.91	.011	.89	.93	.41	1.00
	25 to 34 years	221	.89	.009	.88	.91	.41	1.00
	35 to 44 years	225	.87	.010	.86	.89	.36	1.00
	45 to 54 years	240	.87	.010	.85	.89	.31	1.00
	55 to 64 years	208	.84	.010	.82	.86	.36	1.00
	65 years +	122	.83	.014	.80	.86	.37	1.00
	Total	1149	.87	.004	.86	.88	.31	1.00
Senses	18 to 24 years	133	.89	.011	.87	.91	.44	1.00
	25 to 34 years	221	.90	.007	.89	.91	.32	1.00
	35 to 44 years	225	.89	.008	.88	.91	.36	1.00
	45 to 54 years	240	.86*	.007	.84	.87	.39	1.00
	55 to 64 years	208	.86*	.007	.85	.87	.36	1.00
	65 years +	122	.86*	.009	.85	.88	.43	1.00
	Total	1149	.88	.003	.87	.88	.32	1.00
PSD	18 to 24 years	133	.82	.014	.79	.85	.25	1.00
	25 to 34 years	221	.82	.010	.80	.84	.22	1.00
	35 to 44 years	225	.80	.011	.78	.83	.30	1.00
	45 to 54 years	240	.78	.011	.76	.80	.25	1.00
	55 to 64 years	208	.75	.011	.73	.77*	.26	1.00
	65 years +	122	.74	.015	.71	.77*	.34	1.00
	Total	1149	.79	.005	.78	.80	.22	1.00
MSD	18 to 24 years	133	.38	.019	.34	.42	.03	1.00
	25 to 34 years	221	.40	.011	.37	.42	.06	.96
	35 to 44 years	225	.38	.011	.36	.41	.03	.93
	45 to 54 years	240	.42	.012	.40	.44	.05	1.00
	55 to 64 years	208	.46*	.013	.44	.49	.09	.91
	65 years +	122	.46*	.015	.43	.49	.10	.97
	Total	1149	.42	.005	.41	.43	.03	1.00
AQoL-8D Utility	18 to 24 years	133	.78	.016	.75	.82	.17	1.00
	25 to 34 years	221	.82	.009	.80	.83	.30	1.00
	35 to 44 years	225	.81	.010	.79	.83	.19	1.00
	45 to 54 years	240	.82	.010	.80	.84	.17	1.00
	55 to 64 years	208	.84*	.010	.82	.86	.37	1.00
	65 years +	122	.84*	.010	.82	.86	.39	1.00
	Total	1149	.82	.004	.81	.83	.17	1.00

Notes

*Significant difference (from 18-24) at 5% level.

Appendix 5 Pearson correlation matrix: AQoL-8D and dimensions at baseline, 2 weeks and 1 month

Baseline, T2 and T3 AQoL-8D Dimens and Instrument	Base IL	Base Hap	Base MH	Base Cop	Base Rel	Base SW	Base Pain	Base Sen	Base MSD	Base PSD	Baseline AQoL-8D
Base IL	1										
Base Hap	.361**	1									
Base MH	.371**	.731**	1								
Base Cop	.383**	.772**	.690**	1							
Base Rel	.326**	.720**	.618**	.617**	1						
Base SW	.351**	.666**	.712**	.699**	.647**	1					
Base Pain	.625**	.280**	.336**	.349**	.210**	.208**	1				
Base Sen	.323**	.285**	.234**	.372**	.337**	.330**	.275**	1			
Base MSD	.406**	.852**	.870**	.814**	.830**	.797**	.322**	.315**	1		
Base PSD	.795**	.392**	.402**	.474**	.364**	.361**	.852**	.663**	.444**	1	
Base AQoL-8D	.598**	.830**	.823**	.844**	.754**	.822**	.551**	.532**	.887**	.708**	1
T2 IL	.863**	.355**	.334**	.342**	.331**	.287**	.622**	.351**	.386**	.748**	.562**
T2 Hap	.361**	.858**	.690**	.707**	.669**	.633**	.300**	.218**	.758**	.369**	.772**
T2 MH	.307**	.700**	.878**	.652**	.575**	.690**	.281**	.199**	.806**	.333**	.759**
T2 Cop	.404**	.780**	.693**	.817**	.634**	.690**	.351**	.333**	.772**	.455**	.817**
T2 Rel	.330**	.726**	.672**	.587**	.788**	.702**	.275**	.236**	.794**	.361**	.742**
T2 SW	.356**	.670**	.717**	.676**	.598**	.865**	.296**	.286**	.746**	.389**	.796**
T2 Pain	.572**	.309**	.323**	.318**	.211**	.160*	.852**	.292**	.309**	.754**	.507**
T2 Sen	.222**	.168*	.157*	.229**	.151*	.197**	.173**	.653**	.152*	.440**	.337**
T2 MSD	.381**	.810**	.822**	.741**	.728**	.763**	.319**	.263**	.911**	.409**	.833**
T2 PSD	.688**	.359**	.354**	.381**	.292**	.257**	.752**	.532**	.365**	.848**	.597**
T2 AQoL-8D	.541**	.789**	.779**	.749**	.672**	.747**	.524**	.412**	.805**	.621**	.912**
T3 IL	.858**	.345**	.345**	.336**	.314**	.286**	.622**	.407**	.373**	.774**	.570**
T3 Hap	.400**	.849**	.695**	.714**	.669**	.678**	.310**	.273**	.770**	.407**	.795**
T3 MH	.341**	.693**	.861**	.652**	.559**	.715**	.326**	.235**	.802**	.384**	.772**
T3 Cop	.405**	.748**	.668**	.799**	.640**	.722**	.324**	.353**	.755**	.449**	.811**
T3 Rel	.356**	.687**	.615**	.598**	.749**	.663**	.243**	.293**	.754**	.369**	.713**
T3 SW	.408**	.686**	.722**	.647**	.595**	.856**	.317**	.335**	.743**	.435**	.813**
T3 Pain	.570**	.298**	.287**	.337**	.216**	.171*	.854**	.347**	.297**	.784**	.511**
T3 Sen	.285**	.243**	.247**	.336**	.248**	.334**	.198**	.732**	.265**	.506**	.443**
T3 MSD	.403**	.785**	.794**	.735**	.703**	.770**	.331**	.313**	.890**	.442**	.830**
T3 PSD	.707**	.368**	.359**	.421**	.318**	.309**	.757**	.603**	.391**	.890**	.632**
T3 AQoL-8D	.560**	.768**	.759**	.738**	.659**	.767**	.507**	.450**	.789**	.633**	.909**

Appendix 6 AQoL-8D Instrument: Researcher's copy

Tick the box which best describes your situation as it has been over the past week

Q1 Thinking about how much energy you have to do the things you want to do:
❒ I am
❒ always full of energy
❒ usually full of energy
❒ occasionally energetic
❒ usually tired and lacking energy
❒ always tired and lacking energy

Q2 How often do you feel socially excluded or left out?
❒ never
❒ rarely
❒ sometimes
❒ often
❒ always

Q3 Thinking about how easy or difficult it is for you to get around by yourself outside your house (eg shopping, visiting):
❒ getting around is enjoyable and easy
❒ I have no difficulty getting around outside my
❒ house
❒ a little difficulty
❒ a lot of difficulty
❒ I cannot get around unless somebody is there to help me

Q4 Thinking about your health and your role in your community (that is to say neighbourhood, sporting, work, church or cultural groups):
❒ my role in the community is unaffected by ❒ my health
❒ there are some parts of my community role ❒ I cannot carry out
❒ there are many parts of my community role I cannot carry out
❒ I cannot carry out any part of my community role

Q5 How often do you feel sad:
❒ never
❒ rarely
❒ some of the time
❒ usually
❒ nearly all the time

Q6 Thinking about how often you experience serious pain:
❒ I experience
❒ very rarely
❒ less than once a week
❒ three to four times a week
❒ most of the time

Q7 How much confidence do you have in yourself?
❒ complete confidence
❒ a lot
❒ a moderate amount
❒ a little
❒ none at all

Q8 When you think about whether you are calm and tranquil or agitated:
I am
❒ always calm and tranquil
❒ usually calm and tranquil
❒ sometimes calm and tranquil, sometimes agitated
❒ usually agitated
❒ always agitated

Q9 Thinking about your health and your relationship with your family:
❒ my role in the family is unaffected by my health
❒ there are some parts of my family role I cannot carry out
❒ there are many parts of my family role I cannot carry out
❒ I cannot carry out any part of my family role

Q10 Your close relationships (family and friends) are:
❒ very satisfying
❒ satisfying
❒ neither satisfying nor dissatisfying
❒ dissatisfying
❒ unpleasant
❒ very unpleasant

Q11 When you communicate with others (eg by talking, listening, writing or signing)

- ❒ I have no trouble speaking to them or understanding what they are saying
- ❒ I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me
- ❒ I am understood only be people who know me well. I have great trouble understanding what others are saying to me
- ❒ I cannot adequately communicate with others

Q12 How often do you have trouble sleeping?

- ❒ never
- ❒ almost never
- ❒ sometimes
- ❒ often
- ❒ all of the time

Q13 How often do you feel worthless?

- ❒ never
- ❒ almost never
- ❒ sometimes
- ❒ often
- ❒ all of the time

Q14 How often do you feel angry?

- ❒ never
- ❒ almost never
- ❒ sometimes
- ❒ often
- ❒ all of the time

Q15 Thinking about your mobility, including using any aids or equipment such as wheelchairs, frames, sticks:

- ❒ I am very mobile
- ❒ I have no difficulty with mobility (for example, going uphill)
- ❒ I have some difficulty with mobility. I can go short distances only
- ❒ I have a lot of difficulty with mobility. I need someone to help me
- ❒ I am bedridden

Q16 Do you ever feel like hurting yourself?

- ❒ never
- ❒ rarely
- ❒ sometimes
- ❒ often
- ❒ all of the time

Q17 How enthusiastic do you feel?

- ❒ extremely
- ❒ very
- ❒ somewhat
- ❒ not much
- ❒ not at all

Q18 And still thinking about the last seven days, how often did you feel worried?

- ❒ never
- ❒ occasionally
- ❒ sometimes
- ❒ often
- ❒ all the time

Q19 Thinking about washing yourself, toileting, dressing, eating or looking after your appearance:

- ❒ these tasks are very easy for me
- ❒ I have no real difficulty in carrying out these tasks
- ❒ I find some of these tasks difficult, but I manage to do them on my own
- ❒ many of these tasks are difficult, and I need help to do them
- ❒ I cannot do these tasks by myself at all

Q20 How often do you feel happy?

- ❒ all the time
- ❒ mostly
- ❒ sometimes
- ❒ almost never
- ❒ never

Q21 How much do you feel you can cope with life's problems?

- ❒ completely
- ❒ mostly
- ❒ partly
- ❒ very little
- ❒ not at all

Q22 How much pain or discomfort do you experience:

- ❒ none at all
- ❒ I have moderate pain
- ❒ I duffer from severe pain
- ❒ I suffer unbearable pain

Q23 How much do you enjoy your close relationships (family and friends)?

❒ immensely
❒ a lot
❒ a little
❒ not much
❒ I hate it

Q24 How often does pain interfere with your usual activities?

❒ never
❒ rarely
❒ sometimes
❒ often
❒ always

Q25 How often do you feel pleasure?

❒ always
❒ usually
❒ sometimes
❒ almost never
❒ never

Q26 How much of a burden do you feel you are to other people?

❒ not at all
❒ a little
❒ a moderate amount
❒ a lot
❒ totally

Q27 How content are you with your life?

❒ extremely
❒ mainly
❒ moderately
❒ slightly
❒ not at all

Q28 Thinking about your vision (using your glasses or contact lenses if needed):

❒ I have excellent sight
❒ I see normally
❒ I have some difficulty focusing on things, or I do not see them sharply. For example small print, a newspaper or seeing objects I the distance
❒ I have a lot of difficulty seeing things. My vision is blurred. I can see just enough to get by
❒ I only see general shapes. I need a guide to move around
❒ I am completely blind

Q29 How often do you feel in control of your life?

❒ always
❒ mostly
❒ sometimes
❒ only occasionally
❒ never

Q30 How much help do you need with jobs around the house (eg preparing food, cleaning the house or gardening):

❒ I can do all these tasks very quickly and efficiently without any help
❒ I can do these tasks relatively easily without help
❒ I can do these tasks only very slowly without help
❒ I cannot do most of these tasks unless I have help
❒ I can do none of these tasks by myself

Q31 How often do you feel socially isolated?

❒ never
❒ rarely
❒ sometimes
❒ often
❒ always

Q32 Thinking about your hearing (using your hearing aid if needed):

❒ I have excellent hearing
❒ I hear normally
❒ I have some difficulty hearing or I do not hear clearly. I have trouble hearing softly-spoken people or when there is background noise
❒ I have difficulty hearing things clearly. Often I do not understand what is said. I usually do not take part in conversations because I cannot hear what is said
❒ I hear very little indeed. I cannot fully understand loud voices speaking directly to me
❒ I am completely deaf

Q33 How often do you feel depressed?

❒ never
❒ almost never
❒ sometimes
❒ often
❒ very often
❒ all the time

Q34 Your close and intimate relationships (including any sexual relationships) make you:

❒ very happy
❒ generally happy
❒ neither happy nor unhappy
❒ generally unhappy
❒ very unhappy

Q35 How often did you feel in despair over the last seven days?

❒ never
❒ occasionally
❒ sometimes
❒ often
❒ all the time

References

Brazier, J., R. Harper, et al. (1998). "Deriving a preference based single index measure from the SF36." Journal of Clinical Epidemiology **51**(11): 1115-1129.

Brazier, J., J. Ratcliffe, et al. (2007). Measuring and Valuing Health Benefits for Economic Evaluation. Oxford, Oxford University Press.

Brazier, J. and J. Roberts (2004). "The estimation of a preference-based measure of health from the SF-12." Medical Care **42**(9): 851-859.

Brazier, J., J. Roberts, et al. (2002). "The estimation of a preference-based measure of health from the SF-36." Journal of Health Economics **21**: 271-292.

Brooks, R., R. Rabin, et al. (2005). The Measurement and Valuation of Health Status using EQ-5D: A European Perspective, Evidence from the EuroQoL BIOMED Research Programme. Dordrecht, Kluwer Academic Publishers.

Dolan, P., C. Gudex, et al. (1995). A social tariff for EuroQoL: Results from a UK general population survey, Discussion Paper No 138. York, Centre for Health Economics, University of York.

Fayers, P. M. and D. Machin (2000). Quality of Life: Assessment, Analysis and Interpretation. Chichester, John Wiley & Sons Ltd.

Feeny, D., W. Furlong, et al. (2002). "Multi attribute and single attribute utility functions for the Health Utilities Index Mark 3 System." Medical Care **40**(2): 113-128.

Feeny, D. H., G. W. Torrance, et al. (1996). Health Utilities Index. Quality of Life and Pharmacoeconomucs in Clinical Trials. B. Spilker. Philidelphia, Lippincott-Raven Publishers. **Second Edition:** 239-252.

Fryback, D. G., M. Palta, et al. (2010). "Comparison of 5 health related quality of life indexes using item response theory analysis." Medical Decision Making **30**(1): 5-15.

Hawthorne, G. (2009). "Assessing utility where short measures are required: development of the short Assessment of Quality of Life 8 (AQoL 8) instrument." Value in Health **12**(6): 948-957.

Hawthorne, G., J. Richardson, et al. (2001). "A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments." Annals of Medicine **33**: 358-370.

Hawthorne, G., J. Richardson, et al. (1999). "The Assessment of Quality of Life (AQoL) instrument: A psychometric measure of health related quality of life." Quality of Life Research **8**: 209-224.

Hawthorne, G., J. Richardson, et al. (1999). "The Australian Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life." Quality of Life Research **8**: 209-224.

Hoinville, G., R. Jowell, et al. (1977). Survey Research Practice. London, Heinemann Educational Books.

Iezzi, A., Non-existent, et al. (2009). Measuring Quality of Life at the CHE: Databases, their use and availability, Research Paper 42. Melbourne, Centre for Health Economics, Monash University.

Iezzi, A. and J. Richardson (2009). Measuring Quality of Life at the CHE: Description of instruments, interview props and their administration, Research Paper 41. Melbourne, Centre for Health Ecoomics, Monash University.

Kaplan, R. and J. Anderson (1996). "The quality of wellbeing scale." Medical Outcome Trust **Bulletin 4**: 2-3.

Kaplan, R. and J. P. Anderson (1996). The General Health Policy Model: An Integrated Approach. Quality of Life and Pharmacoeconomics in Clinical Trials. B. Spilker. Philadelphia, Lippinocott-Raven Publishers**:** 309-322.

Kind, P. (1996). The EuroQoL instrument: An index of health related quality of life. Quality of Life and Pharmacoeconomics in Clinical Trials. B. Spilker. New York, Lippincott-Raven Publishers.

McDonald, R. P. (2005). "Semiconfirmatory Factor Analysis: The Example of Anxiety and Depression." Structural Equation Modeling: A Multidisciplinary Journal **12**(1): 163-172.

Misajon, R., G. Hawthorne, et al. (2005). "Vision and quality of life: The development of a utility measure." Investigative Ophthalmology & Visual Science **46**(11): 4007-4015.

O'Connor, R. (2004). Measuring Quality of Life in Health. Edinburgh, Churchill Livingstone.

Richardson, J., N. A. Day, et al. (2004). "Measurement of the quality of life for economic evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument." Australian Economic Review **37**(1): 62-88.

Richardson, J., G. Elsworth, et al. (2011). Increasing the Sensitivity of a the AQoL Inventory for Evaluation of Interventions Affecting Mental Health, Research Paper 61. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., A. Iezzi, et al. (2009). Data used in the development of the AQoL-8D (PsyQoL) Quality of Life Instrument, Research Paper 40. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., A. Iezzi, et al. (2011). Utility weights for the Vision Related Assessment of Quality of Life (AQoL) 7D instrument, Research Paper 65. Melbourne, Centre for Health Economics.

Richardson, J. and M. A. Khan (2009). Preliminary results for the Validation of the Assessment of Quality of Life AQoL-8D Instrument, Research Paper 47. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., J. McKie, et al. (2011). Multi Attribute Utility Instruments and their use, Research Paper 63, Centre for Health Economics, Monash University.

Rosser, R. (1993). A health index and output measure. Quality of Life Assessment: Key Issues in the 1990s. S. Walker and R. Rosser. Dordrecht, Kluwer Academic Publishers

Sintonen, H. and M. Pekurinen (1989). "A generic 15 dimensional measure of health-related quality of life (15D)." Journal of Social Medicine **26**: 85-96.

Sintonen, H. and M. Pekurinen (1993). A fifteen-dimensional measure of health related quality of life (15D) and its applications. Quality of Life Assessment. S. Walker and R. Rosser. Dordrecht, Kluwer Academic Publishers.

Streiner, D. and G. R. Norman (2003). Health Measurement Scales: A Practical Guide to their Development and Use. Oxford, Oxford University Press.

Streiner, D. L. and G. R. Norman (2003). Health Measurement Scales: A practical guide to their development and use. Oxford, Oxford University Press.

Torrance, G. (1986). "Measurement of health state utilities for economic appraisal." *Journal of Health Economics* **5**: 1-30.

Torrance, G., D. Feeny, et al. (1996). "Multiattribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2." Medical Care **34**(7): 702-722.

von Winterfeldt, D. and W. Edwards (1986). Decision Analysis and Behavioral Research. Cambridge, Cambridge University Press.