MONASH University
Business and Economics

Centre for Health Economics

Research Paper 2011 (68)

Utility weights for the vision related Assessment of Quality of Life (AQoL)-7D instrument

Professor Jeff Richardson
Centre for Health Economics, Monash University

Angelo Iezzi
Centre for Health Economics, Monash University

A/Prof Stuart Peacock
Canadian Centre for Applied Research in Cancer Control (ARCC)
School of Population and Public Health, University of British Columbia

Dr Kompal Sinha
Centre for Health Economics, Monash University

Dr Munir A Khan
Centre for Health Economics, Monash University

Dr RoseAnne Misajon
School of Political and Social Inquiry, Monash University

Professor Jill Keeffe
Centre for Eye Research Australia, University of Melbourne
Vision CRC

August 2011

Centre for Health Economics
ISSN 1833-1173
ISBN 1 921187 67 0

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

Purpose: To obtain utility weights consistent with the needs of economic evaluation for the Assessment of Quality of Life (AQoL)-7D, a generic instrument created to increase the sensitivity of the measurement of quality of life amongst people with impaired vision.

Methods: Two extant instruments were combined, the VisQoL (Vision Quality of Life Index) and the Assessment of Quality of Life (AQoL)-6D. Utilities were obtained from patients with visual impairment and from the general population using the time trade-off (TTO) methodology. Dimensions were combined and an econometric adjustment used to eliminate the effects of instrument redundancy. Bias was tested by comparison of holistic TTO values with utility scores predicted from the AQoL-7D scoring formula.

Results: The AQoL-7D instrument consists of 26 items and 7 dimensions each with good psychometric properties. Their combination into a single instrument resulted in significant redundancy which was successfully eliminated. Utility formulae for both the public and patients produced bias free estimates of the utility of holistic health states describing visual impairment. Results imply differing valuations of health states by the public and by people with impaired vision.

Conclusions: The AQoL-7D can detect changes in health states affecting people with impaired vision which are likely to be overlooked by other generic instruments due to content insensitivity. The utilities it produces are generated using a ‘mainstream’ methodology, the TTO. QALY values based upon the AQoL-7D may therefore be used for economic evaluation of programs.

TABLE OF CONTENTS

List of Tables

List of Figures

Utility weights for the vision related Assessment of Quality of Life (AQoL)-7D instrument

1 Introduction

Over 82 percent of those who are blind worldwide are from the 950 million people aged 50 or over. By 2050 the number aged over 60 will exceed 2 billion. This combination of an ageing population and a high correlation between age and vision impairment indicates that visual impairment is likely to increase over time (Australian Institute of Health and Welfare 2007).

Vision impairment results in an increased risk of falls, hip fractures, depression, social isolation, need for community services and greater risk of admission to a nursing home (West, Munoz et al. 1997; Taylor, McCarty et al. 2000; Wang, Mitchell et al. 2001; McCarty, Fu et al. 2002). The consequences of these for an individual’s quality of life (QoL) need to be included in the economic evaluation of health programs: their exclusion would result in a systematic bias against the funding of services to prevent or cure visual impairment.

Presently, economic evaluations attempt to take QoL into account using the Quality Adjusted Life Year (QALY) as the unit of outcome, where QALYs are defined as life years multiplied by an index of ‘utility’ – that is, QoL is measured by ‘utility’, the strength of preference for a particular health state. Cost Utility Analysis (CUA) compares the extra costs with the extra QALYs gained when one program is compared with another. To obtain valid measures of QALYs requires a valid measurement of utility.

This is carried out one of two ways. With the holistic approach, the relevant health states are described in a series of scenarios. The index of utility is obtained by rating the scenarios using a ‘scaling instrument’, the most popular of which are the time trade-off (TTO) and standard gamble (SG).

With the alternative, ‘decomposed’ or multi attribute (MAU) approach, a generic instrument is constructed consisting of a standard questionnaire, with response categories similar to many disease specific instruments. The distinguishing feature of the MAU instrument is that it provides a method, based upon independent research, to attach utility weights to every possible combination of item responses. As the number of combinations is typically very large, the utilities are modelled and estimated from a smaller number of direct observations of individual utilities. The construction and use of MAU instruments including the Assessment of Quality of Life (AQoL) instruments have been published elsewhere (Brazier, Ratcliffe et al. 2007; Richardson, McKie et al. 2011 forthcoming).

The two largest studies comparing the most widely used MAU instruments found that an average of only 54 and 47 percent respectively of measured variation in scores could be ‘statistically explained’ by other instruments (Hawthorne, Richardson et al. 2001; Fryback, Palta et al. 2010). This means that purportedly generic instruments yield results which are quite different from each other implying significant differences in sensitivity. For example, in the earlier of the two studies, one individual who had hearing and visual impairment recorded a low utility score of 0.14 on the Health Utilities Index 3 (HUI 3). However, using the EQ-5D, (originally known as the EuroQol) the same individual’s score was 0.8. Unlike the HUI, the EQ-5D does not have a dimension for vision or hearing (Brazier, Ratcliffe et al. 2007). A program which returned the individual to full health (U = 1.00) would have recorded an increase in utility 4.3 times greater using the HUI than with the EQ-5D (1.00-0.14)/(1.00-0.80). Conversely, using the EQ-5D to measure utility would have had the same effect on the cost/QALY ratio as a 4.3 fold increase in costs. The example is extreme but indicates the potentially lethal effects of the choice of instrument upon the likelihood of a program being funded.

Except for the HUI and 15D, the major instruments make no explicit reference to the benefits associated with vision per se and their indirect descriptive items are likely to measure the effects of visual impairment imperfectly. The purpose of the present project was to develop an MAU instrument with increased sensitivity to visual acuity and vision related handicap.

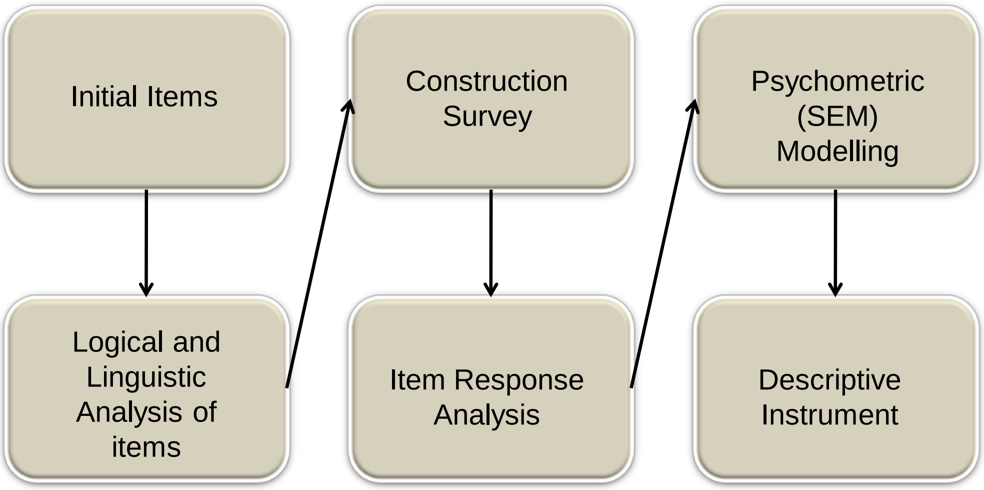
Constructing an MAU instrument involves two broad steps. First, it is necessary to derive a descriptive system with content validity – a set of questions and responses which incorporate the different dimensions of the concept to be measured. Secondly, a scoring system or formula must be developed which converts responses into an overall utility score. Psychometric analyses which resulted in the AQoL-7D descriptive system has been published elsewhere (Misajon, Hawthorne et al. 2005; Peacock, Misajon et al. 2008; Peacock, Richardson et al. 2010). It is summarised in Section 2 below. The aim of the present paper is to describe the second step, in which a scoring formula is derived to produce utility scores from the descriptive system. Section 2 outlines the data collection and statistical analyses for the utility formula. Section 3 presents results, and Section 4 presents an initial validation test from the use of the formulae with patients and the public. The formulae is available online in a program suitable for downloading and free use (http://www.aqol.com.au/).

2 Methods

The AQoL-7D instrument

The descriptive system for the AQoL-7D was created by combining the descriptive systems of two extant generic instruments, the VisQoL and the AQoL-6D. The former was constructed specifically to measure visual impairment related QoL on a scale from best to worst vision (Misajon, Hawthorne et al. 2005). The latter is a 20 item generic instrument whose 6 dimensions are independent living, coping, pain, relationships, mental health and senses. It is scored on a scale from best health to death (Richardson, Day et al. 2004). Combination of the scales results in redundancy (structural dependence or ‘double counting’) between overlapping dimensions as the items customised to measure visual impairment are similar to some items in the AQoL-6D, *albeit* nuanced for people with visual impairment. The problem is overcome using a correction to the scoring formula which is described below.

Figure 1 Construction of the descriptive system



Both the VisQoL and the AQoL-6D derived their descriptive systems using the steps prescribed by psychometric instrument construction theory (Figure 1). Initial items were obtained from previous research results and from focus groups. For the construction phase of VisQoL, participants who were vision impaired (n = 70) with visual acuity (VA) worse than 20/30 in their better eye were recruited from Vision Australia to reflect the profile of adults with impaired vision together with a representative sample of the unimpaired adult population (n = 86) with VA of 20/20 or better in both eyes. In both studies grounded theory techniques were used to analyse transcripts and observer notes from the focus groups. The resulting item banks were initially subject to ‘item analysis’ to refine and standardise items and expression. The items were then administered to samples of the population which, in the case of VisQoL, included patients with impaired vision. The final item selection was based upon psychometric properties of the items, exploratory factor analysis (EFA), structural equation modelling (SEM) and logical considerations.

The structure of the full AQoL-7D descriptive system is shown in Figure 2. The final instrument consisted of 26 items: 20 from AQoL-6D and 6 from VisQoL, grouped into 7 dimensions (Box 1). Items in dimension 7 – VisQoL – are structured in the same way as items in the other 6 dimensions with an initial stem followed by 4-6 response categories (Box 2).

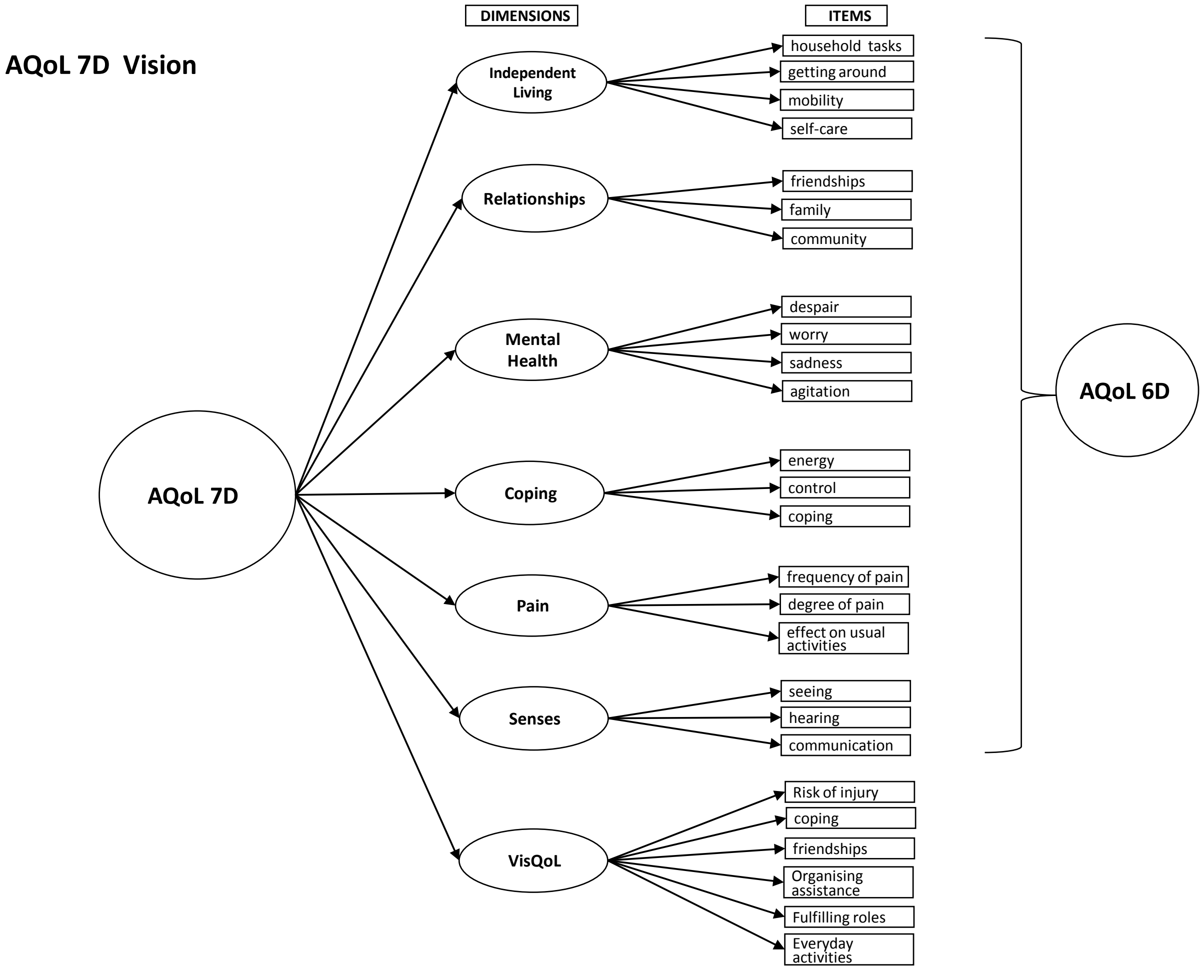
Box 1. AQoL-7D dimensions and item numbers

1. Independent living	…	4 items
2. Relationships	…	3 items
3. Mental health	…	4 items
4. Coping	…	3 items
5. Pain	…	3 items
6. Senses	…	3 items
7. VisQoL	…	6 items

Box 2. Dimension 7: VisQoL

1. Does my vision make it likely I will injure myself (ie when moving around the house, yard, neighbourhood or workplace)? … 5 response levels
2. Does my vision make it difficult to cope with the demands in my life? … 6 response levels
3. Does my vision affect my ability to have friendship? …7 response levels
4. Do I have difficulty organizing any assistance I may need? … 6 response levels
5. Does my vision make it difficult to fulfil the roles I would like to fulfil in my life (eg family roles, work roles, community roles)? … 6 response levels
6. Does my vision affect my confidence to join in everyday activities? … 6 response levels

Figure 2 Structure of the AQoL-7D



Deriving the utility formula

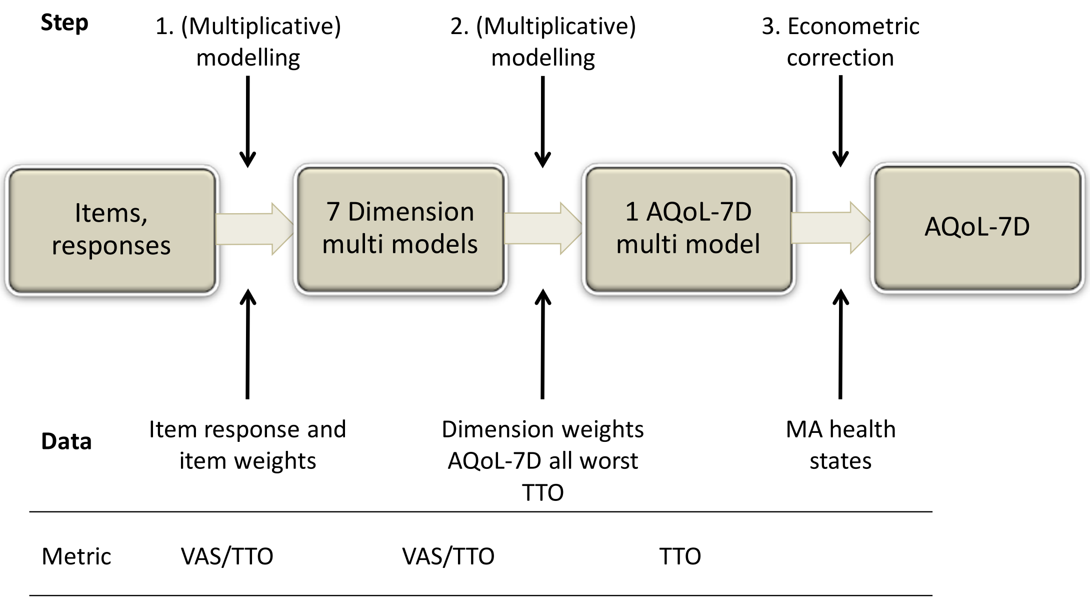
The number of health states described by AQoL-6D is too large to derive utility weights for each state individually. Consequently, values were modelled, i.e. inferred from survey responses to a limited number of questions. Modelling may be carried out two basic ways. First, drawing upon Decision Analytic (DA) theory, some form of weighted averaging of item scores is employed (as for the 15D, Health Utility Index (HUI) 1, 2, and 3, and AQoL-4D). For example, in a simple model where the ‘physical’ and ‘mental’ health dimension scores were 0.9 and 0.7 and the importance weights for physical and mental health were 0.6 and 0.4 respectively, the weighted average score using an additive model would be $0.6*(0.9) + 0.4*(0.7) = 0.82$. (The importance weights must sum to 1.00.) Secondly, regression analysis can be used to predict or ‘explain’ the utilities of a limited number of multi attribute states which are assessed holistically using the TTO or another scaling method. The final regression equation may take the form:

$$\text{Predicted Utility} = 0.1 + 0.4\ (\text{Physical score}) + 0.5\ (\text{Mental score})$$

where the parameters 0.1, 0.4, 0.5 are derived from the regression analysis. While the regression is estimated from a limited number of cases, the equation may be used with any combination of physical and mental scores. The former method has the advantage of simplicity when there are a large number of items. However, if there is structural or preference dependency, any form of simple averaging may cause ‘double counting’. The second method overcomes this difficulty but the statistical procedure is problematical when there are large numbers of correlated items.

AQoL-7D used a combination of these approaches (Figure 3). Initially, item scores were combined into dimension scores using a multiplicative averaging of scores (Step 1). In Step 2, these dimension scores were combined into a single, overall score using the same multiplicative averaging procedure. In Step 3 both the overall and the dimension scores were included in a regression analysis to ‘explain’ independently estimated utilities for MA (multi attribute) health states. The best fitting regression was selected to ‘correct’ estimates for redundancy (double counting of elements) in the final formula which converts item responses into utility scores.

Figure 3 Derivation of the AQoL-7D utility formula



In principle, the multiplicative model is very similar to the simple multiplication of scores (when each score is on a 1.00-0.00 scale) as shown in equation 1. For example, if every dimension weight, $U_i = 0.95$, then the multiplicative score would be $(0.95)^7 = 0.70$. The mathematics is simpler using disutilities, DU, which are related to utilities by equation 2. The decision analytic equation increases flexibility by including the importance weight of each dimension, w_i, which is obtained from survey data (equation 3). The overall multiplicative score is constrained to the range (0.00-1.00) by the scaling constant, k, calculated in equation 4. This has a similar effect to the requirement in an additive model that the importance weights must sum to unity. The model is explained in detail elsewhere (Von Winterfeldt and Edwards 1986).

$$U = U_1 * U_2 * U_3 * \ldots * U_7 \qquad \ldots (1)$$

$$U = 1 - DU \qquad \ldots (2)$$

$$DU = \frac{1}{k}\left[\prod_{i=1}^{n}\left[1 + kw_i DU_i\left(x_{ij}\right)\right] - 1\right] \qquad \ldots (3)$$

$$k = \prod_{i=1}^{n}(1 + kw_i) - 1 \qquad \ldots (4)$$

Key U = Utility (0.00 = death; 1.00 = instrument best)
U_i = utility of dimension (items) i
w_i = dimension (item) importance weight for dimension (item) i
DU = disutility
$DU(x_{ij})$ = disutility of response level j, dimension (item) i
k = scaling constant

In Step 3, a variety of econometric models were used: double log, linear, quadratic and higher order. Two criteria were used to select the best fitting model; namely, the size of the conventional R^2 coefficient and, secondly, the existence of bias (a systematic discrepancy between estimated and observed TTO scores).

Data

The data required for equation (3) consisted of: (i) item response weights, $DU(x_{ij})$; (ii) item importance weights, w_i; and (iii) dimension importance weight w_i. The econometric analysis in Step 3 additionally required TTO values for multi attribute (MA) health states.

Population sampling: Both visually impaired patients and a representative cross section of the Australian population were interviewed. The reason for including the two groups was to determine whether or not people with impaired vision evaluated vision related health states differently from the general population. The population sample was drawn from a computer readable telephone directory, using a stratified, clustered two-stage design, similar to Hawthorne et al. procedure in the AQoL-4D validation study (1999). Based on the Australian Bureau of Statistics Socio-Economic Index for Areas (SEIFA 2001) scores, postcodes were the primary sampling unit, with sampling probabilities proportionate to population size (to reduce the effect of socio-economic confounding). From these postcode areas, telephone subscribers (18 years+) were randomly sampled (n=184) and contacted initially by letter and subsequently by telephone. The use of postcodes as the primary sampling unit meant that respondents would be fairly tightly clustered, minimising the travel costs.

Patient sampling: 180 participants with impaired vision were recruited from the Royal Victorian Eye and Ear Hospital (RVEEH), Vision Australia and Retina Australia. Three levels of impairment were sampled: those with mild (less than 20/20-20/60); moderate (less than 20/60 to 20/200) and severe (worse than 20/200) impairment. Eligible participants from the RVEEH special eye clinics were invited by letter to participate in the study. Peer workers and staff who were vision impaired were similarly invited to participate. Clients from Vision Australia day centres, training classes and support groups were recruited directly. Retina Australia included a letter of invitation to their members with their regular newsletter.

Prior to their interview participants completed rating scales (VAS) for the item responses for the six VisQoL items and these were subsequently transformed into TTO equivalent scores using the transformation algorithm produced for AQoL-6D (Richardson, Day et al. 2004). During the interview TTO data were obtained using the props and protocols described in Iezzi and Richardson (2009). Ethics approval was granted by the Royal Victorian Eye and Ear Hospital Human Research and Ethics Committee and protocols adhered to the tenets of the Declaration of Helsinki.

3 Results

Data: Table 1 reports the number of respondents by age and gender and Table 2 the distribution according to the SEIFA group of their postcodes. The age distribution of patients is skewed towards the elderly reflecting increasing macular degeneration with age. Public respondents were also skewed towards the elderly but subsequent analysis found no significant response difference by age.

Table 1 Age and gender distribution of survey respondents

Age Group	Public			Patients			Total
	Male	Female	Total	Male	Female	Total	
18 to 34 Years	6	7	13	9	9	18	31
35 to 44 Years	8	10	18	3	6	9	27
45 to 54 years	13	21	34	12	9	21	55
55 to 65 Years	15	39	54	17	16	33	87
66 Years +	22	37	59	35	63	98	157
Total	64	114	178	76	103	179	357

Missing n = 7

Step 1: The first step of the modelling required item responses and item importance weights to create multiplicative dimension models. This step created scores for each dimension which are of independent interest but also reduced the 26 items of the AQoL-7D to a smaller, more manageable number of variables for the Step 3 econometric analysis. Item response scores for the vision dimension (VisQoL) are reported in Table 3. Other item scores are reported in Richardson et al.(2004). From Table 3 patient and public respondents gave very similar responses. In half of the cases the difference was significant at the 5 percent level but the absolute differences were small – the largest three differences being 8, 6 and 5 points on a 100 point scale.

Table 2 Distribution of respondents by SEIFA group

SEIFA Group*	Public	Patients	Total
1	42	25	67
2	27	19	46
3	42	34	76
4	35	38	73
5	35	66	101
Total	181	180	363

Missing n = 3

* Poorest socio-economic status is 1 and richest is 5

Table 3 Response category VAS scores for the 6 VisQoL items: Best = 1.00; Worst = 0.00

	Mean VAS score		
	Public	Patient	Total
1: It is most unlikely I will injure myself because of my vision	100.0	100.0	100.0
There is a small chance	85.99	83.99	85.00
There is a good chance	55.07	55.26	55.16
It is very likely	29.70	30.26	29.98
Almost certainly my vision will cause me to injure myself	0.00	0.00	0.00
2: Has no effect on my ability to cope with the demands in my life	100.0	100.0	100.0
Does not make it difficult at all to cope with the demands in my life	96.25	96.46	96.35
Makes it a little difficult to cope	79.89	77.29	78.59
Makes it moderately difficult to cope	57.77	56.02	56.90
Makes it very difficult to cope	26.26	23.05	24.67
Makes me unable to cope at all	0.00	0.00	0.00
3: My vision makes having friendships easier	100.0	100.0	100.0
Has no effect on my friendships	96.45	97.59	97.02
Makes friendships more difficult	72.45	67.63	70.05
Makes friendships a lot more difficult	46.83	42.69	44.77
Makes friendships extremely difficult	26.08	20.82	23.46
Makes me unable to have friendships	0.00	0.00	0.00
4: I have no difficulty organising any assistance	100.0	100.0	100.0
I have a little difficulty organising assistance	87.24	83.03	85.14
I have moderate difficulty organising assistance	65.91	60.21	63.08
I have a lot of difficulty organising assistance	34.32	26.13	30.25
I unable to organising assistance at all	0.00	0.00	0.00
5: My vision has no effect on my ability to fulfil these roles	100.0	100.0	100.0
Does not make it difficult to fulfil these roles	96.66	98.21	97.43
Make it a little difficult to fulfil these roles	79.53	77.54	78.54
Make it moderately difficult to fulfil these roles	56.24	54.59	55.42
Make it very difficult to fulfil these roles	28.85	23.28	26.08
My vision means I am unable to fulfil these roles	0.00	0.00	0.00
6: My vision makes me more confident to join in everyday activities	100.0	100.0	100.0
Has no effect on my confidence to join in everyday activities	96.82	98.64	97.72
Makes me feel a little less confident	81.63	78.70	80.17
Makes me feel moderately less confident	60.69	56.40	58.56
Makes me feel a lot less confident	35.53	27.07	31.32
Makes me not confident at all	0.00	0.00	0.00

Following DA theory, worst response item scores were used as importance weights. They were valued on a scale whose endpoints were the dimension-best and dimension-worst health states. Thus, respondents were asked to consider a health state consisting of an item at its worst level with other items in the same dimension at their best level. Patient and public scores shown in Table 4 are again similar, differing by an average of 0.2 and a maximum of 0.37 points on a 10 point scale. Despite the statistical insignificance of individual differences, however, every patient's score is greater than the corresponding public score. The probability that this would occur by chance is less than 5 percent, which suggests that patients assign less disutility to problems which are similar to those they have experienced.

Table 4 Item worst TTO scores on a 10 point Best Health-Death scale

VisQoL items	Mean TTO			SE			Sig
	Public (N=184)	Patients (N=182)	Total (N=366)	Public	Patients	Total	
Does my vision make it likely I will injure myself?	7.12	7.49	7.30	.17	.17	.12	.120
Does my vision make it difficult to cope with the demands in my life?	6.17	6.43	6.30	.20	.18	.14	.335
Does my vision affect my ability to have friendships?	6.84	7.00	6.92	.21	.17	.14	.552
Do I have difficulty organising any assistance I may need?	6.97	7.09	7.03	.19	.17	.13	.629
Does my vision make it difficult to fulfil the roles I would like to fulfil in life?	6.57	6.86	6.71	.20	.18	.13	.276
Does my vision affect my confidence to join in everyday activities?	7.18	7.20	7.19	.19	.17	.13	.946

Combining VisQoL item responses and importance weights with the multiplicative model (equations 3 and 4) resulted in the multiplicative dimension, equation 5 below.

Dimension 7 (VisQoL)

$$\begin{aligned} DU_7 = 1.20[1 - &(1-0.22du_{21})(1-0.31du_{22})(1-0.26du_{23}) \\ &(1-0.25du_{24})(1-0.27du_{25})(1-0.23du_{26}) \qquad \ldots(5) \end{aligned}$$

where DU_7 is the multiplicative dimension disutility and du_{21} … du_{26} are the average disutility scores obtained for the VisQoL items (AQoL-7D items 21 to 26).

Error! Not a valid link.**Step 2**: The procedures described for Step 1 were repeated in Step 2 to combine dimensions into a multiplicative AQoL-7D model. As with Step 1, this collapsed the number of variables, in this case from 7 to 1, which maximised the chance of a simple three stage econometric model. In Step 2, dimension scores replaced item scores – DU (x_{ij}) – in equations 3 and 4. The AQoL-6D importance weights were used for the 6 non vision dimensions. TTO valuation of the vision dimension (VisQoL) worst health state on a full health-death scale yielded an importance weight of 0.747. The resulting multiplicative model for AQoL-7D (AQoL-6D plus VisQoL) is shown in equation 6 below.

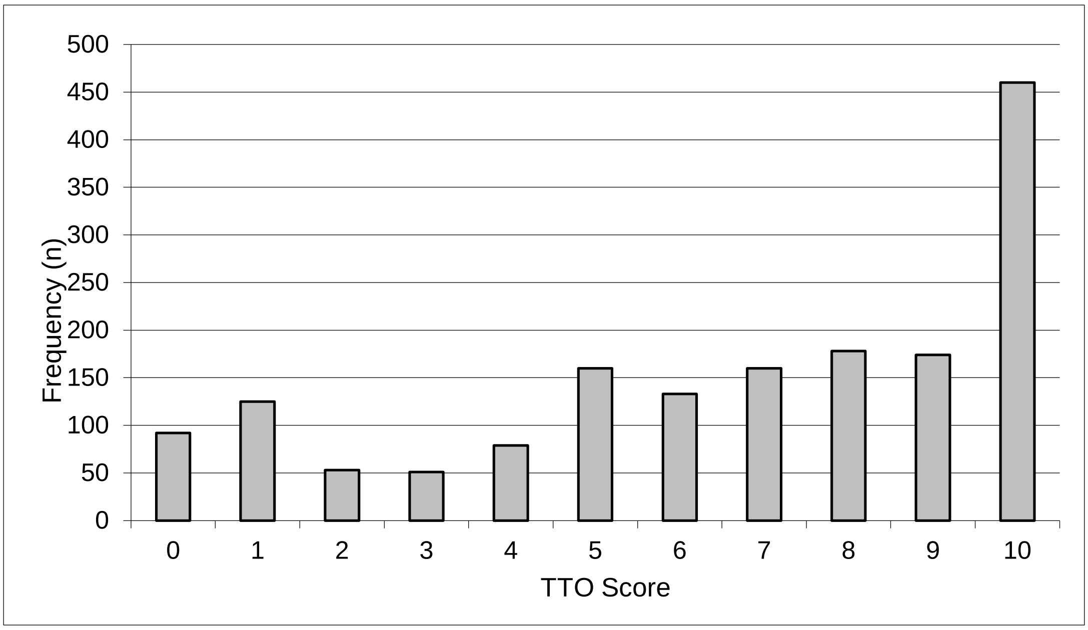
AQoL-7D (Multiplicative)

$$DU_{AQoL-7D} = 1.02[1-(1-0.407DU_1)(1-0.387DU_2)(1-0.413DU_3)$$
$$(1-0.298DU_4)(1-0.511DU_5)(1-0.550DU_6)(1-0.747DU_7)] \qquad \ldots(6)$$

Step 3: In Step 3 the multiplicative score derived in Step 2 was 'corrected' to allow for the double counting of disutilities. Independently assessed TTO (utility) scores of MA health states were regressed upon the multiplicative estimate and upon dimension scores to allow for dimension specific effects. The MA health states were constructed logically using criteria of (i) coherence; (ii) multi dimensionality; and (iii) simplicity (to minimise the cognitive load on respondents). To obtain interaction effects health states were constructed in which all combinations of dimensions had imperfect health. This resulted in the use of 28 health states. An average of 45 'observations' (respondent valuations) were obtained for each of these giving a total of 1,251 independent observations. The frequency distribution of these is shown in Figure 4. The states are defined in an appendix to Richardson et al. (2010) and were evaluated during the interview using the TTO technique.

Few of the values were for relatively good health states. Consequently, a procedure was adopted from the AQoL-6D to estimate a selection of these. Some MA health state descriptions were split into two parts. For example, an MA state with item response levels (a_1, 1, b_1, 1, a_2, b_2…) might be divided into the two states (a_1, 1, 1, 1, a_2, 1…) and (1, 1, b_1, 1, 1, b_2…). The disutility of the original MA state, assessed from the TTO interview, was apportioned between the two new 'pseudo health states' according to the ratio of the multiplicative scores of the pseudo states . An additional 37 pseudo health states were constructed in this way, giving a total of 65 health states and 1,665 individual observations.

Figure 4 Frequency distribution of TTO scores (n = 1,665)



The best fitting linear models are reported in Table 5 for the public, patients and the total sample. In all of these, observed disutilities, DU, (measured by TTO) are ‘explained’ by the multiplicative AQoL-7D score and the dimension scores. Mean values of health states were used as observations to overcome the common problem of extreme non normality of the data and particularly the number of observations where respondents would not trade (see TTO score of 10 in Figure 4). Each observation was therefore the result of an average of 45 individual responses. None of the average utility scores for the health states was below 0.5. Consequently, extrapolating the present results and drawing conclusions for health states with utility scores below 0.5 would yield unreliable conclusions. The pertinence of this, however, may not be great as there are probably few illnesses so severe that individuals would be willing to sacrifice half the remaining years of their life in order to avoid them.

A requirement of the final model is that it predicts a disutility of 0.00 when the disutility of every dimension is 0.00. Consequently, the constant term in the regression analyses was set equal to zero. Non linear models were tested but linear equations fitted the data better. Four dimension variables were statistically significant in at least one regression. These were retained in the equations for the separate public and patient populations.

4 Validation and discussion

The most striking implication of the stage 2 adjustment reported in Table 5 is the 63-66 percent reduction in the magnitude of the first stage estimate of the disutility. This is similar to the reduction found in the analysis for the AQoL-6D. As expected, the multiplicative model ‘double counts’ elements of disutility because of the correlation of items in the descriptive system. In addition to the overall reduction in the stage 1 multiplicative score, there is a further reduction in the patient regression associated with dimension 6, senses (coefficient of -0.1). This is consistent with the earlier, tabulated result.

Bias: The scoring formula is designed to predict a valid ‘utility’. The formula must therefore predict scores which are equal to holistically derived (MA) TTO scores, as the TTO score is (in this analysis) the gold standard. Perfect prediction would result in an equation of the form TTO = 0.00 + 1.00*(prediction). In contrast, an equation of the form TTO = 0.2 + 0.7*(prediction), would imply bias. A predicted score of 0.00 would correspond with a true value of 0.2; an increase in the predicted value of 0.1 would correspond with a true increase of 0.7*0.1 = 0.07.

Bias was tested by regressing the average TTO values on the scores predicted by the formulae in Table 5. Results are shown in Table 6. This indicates that, despite a better overall fit the result for the total population, regression 3, produces a biased estimate: the ‘b’ coefficient of 0.75 implies that an increase in the predicted TTO of 0.1 corresponds with a true increase of 0.078. By contrast, the separate public and patient equations produce unbiased estimates. Results are plotted in Figures 5, 6 and 7.

Table 5 Regression of mean TTO disutility scores on predicted (multiplicative) AQoL disutilities (OLS Linear Models)

Dependent mean TTO / Independent	Public		Patient		Total	
	b	t	b	t	b	t
DU$_0$ (Mult) AQoL-7D	0.68*	(10.8)	.65*	(11.9)	.71*	(21.2)
DU$_2$ Dim 2 Relationships	-0.16*	(-3.07)	-0.08	(-1.5)	-0.11*	(-4.4)
DU$_4$ Dim 4 Coping	-0.13*	(-2.3)	-0.02	ns	-0.07*	(-2.9)
DU$_5$ Dim 5 Pain	-0.01	ns	-0.12	(-2.1)	-0.09*	(-2.9)
DU$_6$ Dim 6 Senses	-0.04	ns	-0.10	(-1.76)	-0.07*	(-2.6)
R^2 (Pseudo)	0.95		0.93		0.95	
RMSE	0.06		0.05		0.06	
n	65		65		65	
F-statistic	116		98		266	
Prob > F	0.00		0.00		0.00	
If DU = 0, U =	1.0		1.0		1.0	
If DU = 1 U =	.66		.64		.63	

*Statistically significant at 5% level of significance

Table 6 Regression of mean TTO on predicted AQoL utility

Dependent mean TTO / Independent	Regression*		
	Public (1)	Patient (2)	Total (3)
n	30	35	65
TTO Pred*	1.00	1.00	0.78 (13.1)
Constant	0.00	0.00 3	0.05 (3.0)
R^2	0.73	0.64	0.73
RMSE	0.06	0.07	0.05
F	79	58	171

*The ‘independent’ variable is the value predicted from the corresponding regression in Table 5

Figure 5 Observed versus predicted disutility: Public

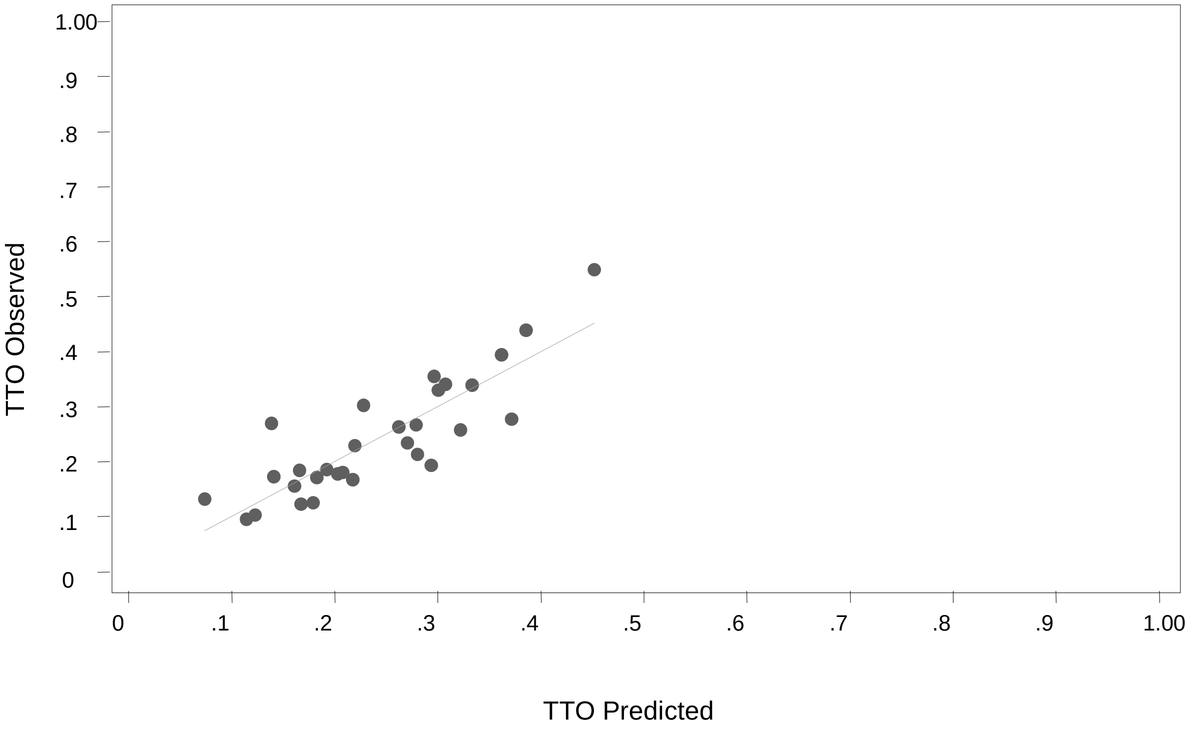


Figure 6 Observed versus predicted disutility: Patients

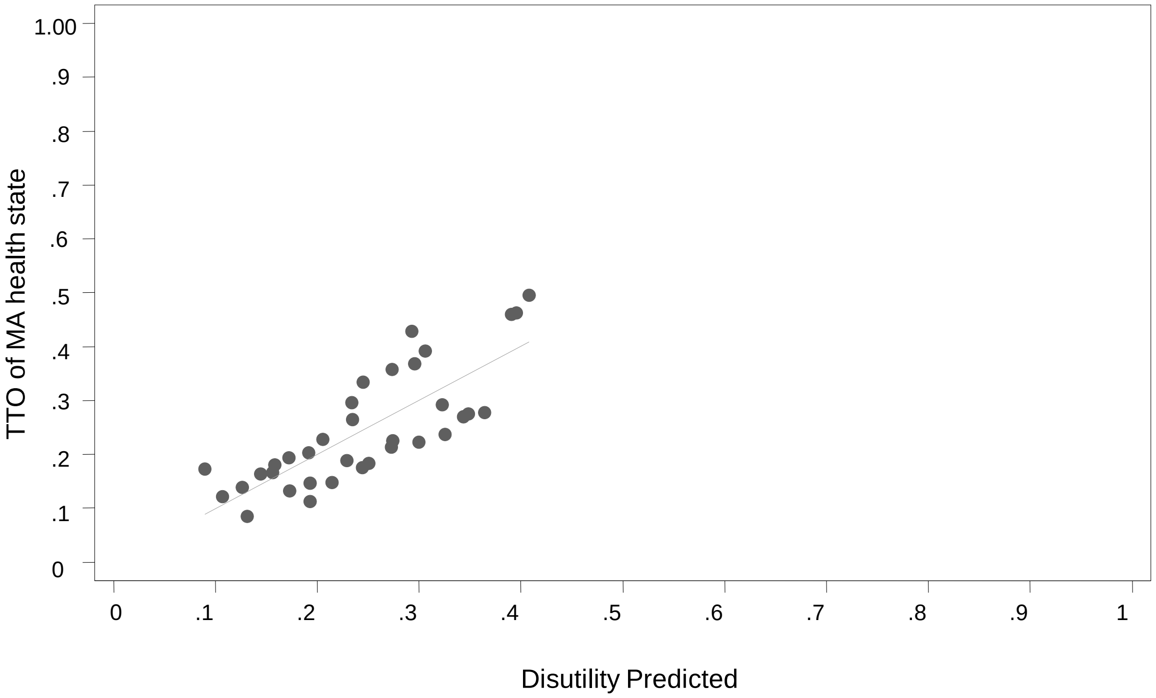
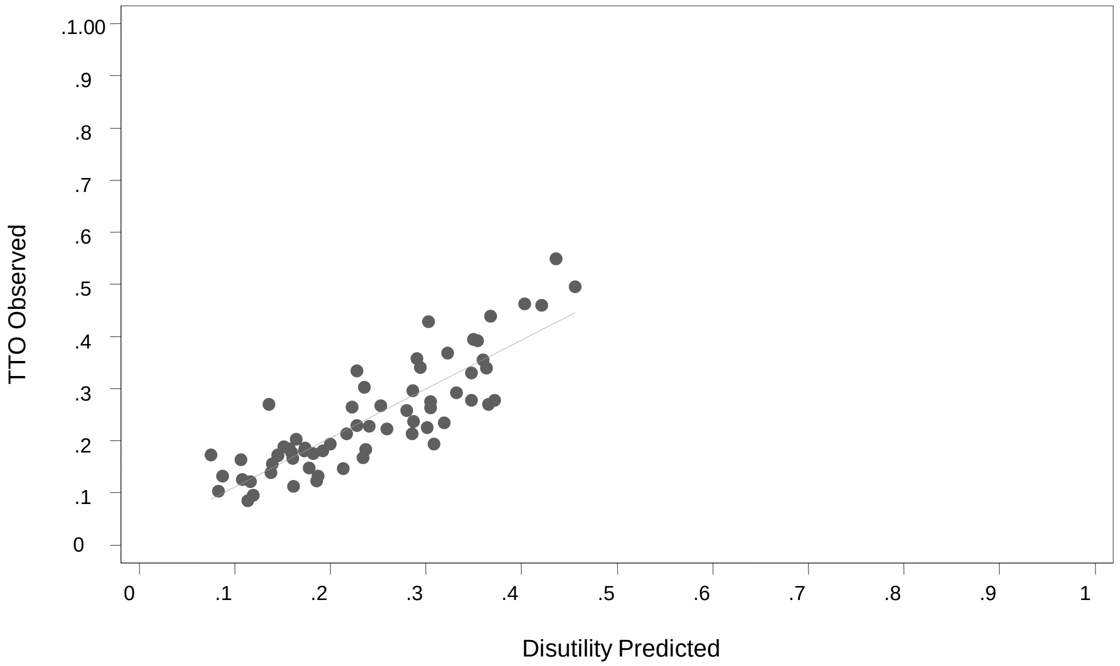


Figure 7 Observed versus predicted disutility: Total



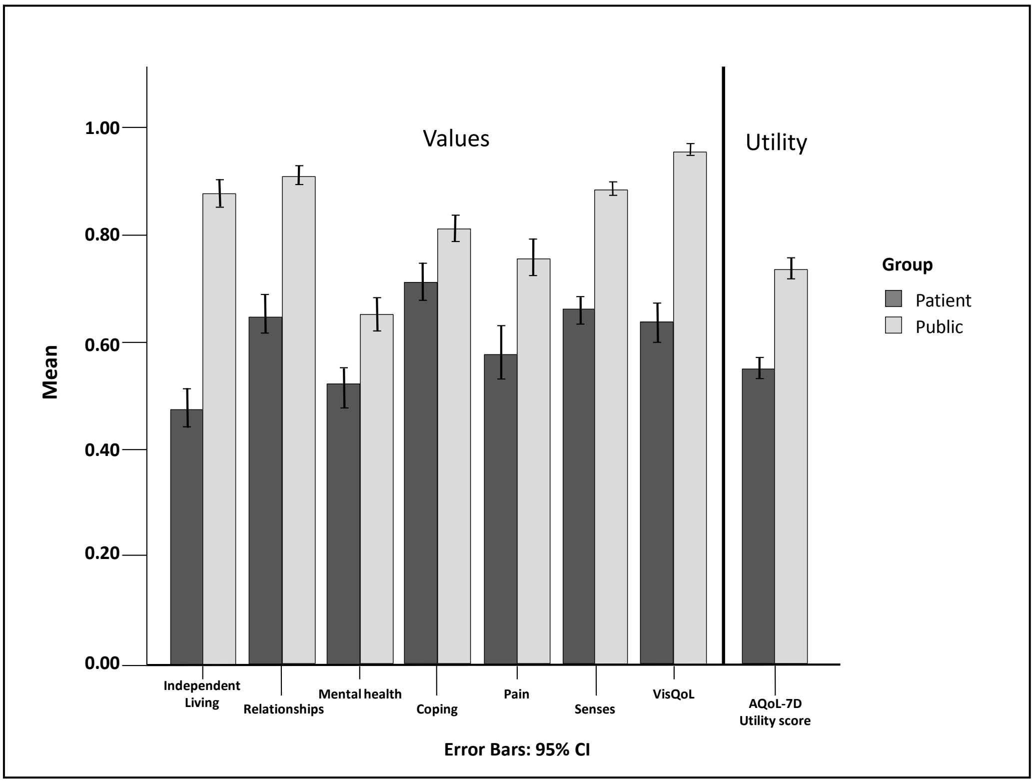
Validity

The validation of an instrument is an ongoing process in which tests are conducted to determine whether or not the instrument's performance is satisfactory and justifies confidence in the results it produces (Richardson, McKie et al. 2011 forthcoming). Validity may be context specific: an instrument which has performed well for one disease may or may not do so for another. The low correlation between instrument scores noted in the introduction suggest that this may be true for some widely accepted MAU instruments in the context of vision.

Confidence in the content of an instrument (content validity) depends, in large part, upon the process of instrument construction. As reported earlier, the AQoL-7D performance in this respect is very good (Misajon, Hawthorne et al. 2005; Peacock, Misajon et al. 2008). Nevertheless, ongoing testing of an instrument is needed. An initial and important test is the ability of an instrument to discriminate between populations which would be expected to yield different scores.

Before interview, all survey respondents completed the AQoL-7D questionnaire (describing their own health states). The results were used to predict dimension and utility scores using the separate equations for the two groups. The results do not demonstrate causation because the two groups have other potentially relevant differences. For example, people with impaired vision also had lower incomes. Rather, the comparison sought to determine the ability of the AQoL-7D to discriminate between the two groups. Results shown in Figure 8 are striking. The differences are not just confined to the vision specific dimension but are significant for all 7 dimensions. The result demonstrates the discriminant validity of AQoL-7D but also suggests that use of QoL instruments which are confined to vision related items and exclude the other dimensions may fail to measure effects that are important for a person's utility. Nevertheless, the utility scores obtained from the formulae in this paper should be treated with caution as the instrument has not, to date, been widely used.

Figure 8 AQoL-7D Mean utility score by dimension for Patients and Public



5 Conclusion

The evidence presented here indicates that the utility scores of the AQoL-7D are plausible and that the formulae used to derive them produce unbiased estimates of directly elicited TTO scores. The results suggest that to achieve content validity, the quality of life of people with impaired vision should be assessed using a generic multi attribute instrument: the difference in QoL scores between people with impaired vision and the general public is not properly measured using only the vision related items. Since the final scores of the AQoL-7D provide an unbiased estimate of TTO utilities, they may be used with some confidence in economic evaluation studies, although their usefulness is more general and may be used for description and the generation of dimension profiles.

References

Australian Institute of Health and Welfare (2007). A guide to Australian eye health data, Cat. no. PHE 86. Canberra, Australian Institute of Health and Welfare (AIHW).

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

Fryback, D. G., M. Palta, et al. (2010). "Comparison of 5 Health Related Qof Life indexes using Item Response Theory analysis." Medical Decision Making **30**(1): 5-15.

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.

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.

McCarty, C. A., C. L. Fu, et al. (2002). "Predictors of falls in the Melbourne visual impairment project." Aust N Z J Public Health **26**: 116-119.

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.

Peacock, S., R. Misajon, et al. (2008). "Vision and quality of life: Development of methods for the VisQoL vision related utility instrument." Ophthalmic Epidemiology **15**(4): 218-223.

Peacock, S., J. Richardson, et al. (2010). Construction of the Descriptive System for the Assessment of Quality of Life (AQoL-6D) Utility Instrument, Research Paper 49. Melbourne, Centre for Health Economics, Monash University.

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., J. McKie, et al. (2011 forthcoming). Multi attribute utility instruments and their use. Encyclopedia of Health Economics. T. Culyer. San Diego, Elsevier Science.

Richardson, J., S. Peacock, et al. (2004). The Assessment of Quality of Life (AQoL) 2 instrument: Derivation of the scaling weights using a multiplicative model and econometric second stage correction. Working Paper 142. Melbourne, Centre for Health Economics, Monash University: http://www.buseco.monash.edu.au/centres/che/pubs/wp142.pdf.

Taylor, H. R., C. A. McCarty, et al. (2000). "Vision impairment predicts five year mortality." Trans Am Ophthalmol Soc **98**: 91-96.

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

Wang, J. J., P. Mitchell, et al. (2001). "Incidence of nursing home placement in a defined community." Medical Journal of Australia **174**: 271-275.

West, S. K., B. Munoz, et al. (1997). "Function and visual impairment in a population-based study of older adults. The SEE project. Salisbury Eye Evaluation." Invest Ophthalmol Vis Sci **38**: 72-82.