MONASH University
Business and Economics

Centre for Health Economics

Research Paper 2009 (39)

The AQoL-8D (PsyQoL) MAU Instrument: Overview September 2009

Jeff Richardson[1]

Munir Khan[1]

Angelo Iezzi[1]

Kompal Sinha[1]

Cathy Mihalopoulos[2]

Helen Herrman[3]

Graeme Hawthorne[3]

Issy Schweitzer[4]

[1] Centre for Health Economics, Monash University
[2] Health Economics Unit, Deakin University
[3] Department of Psychiatry, The University of Melbourne
[4] The Melbourne Clinic, The University of Melbourne

September, 2009

Centre for Health Economics
ISSN 1833 – 1173
ISBN:

ACKNOWLEDGEMENT

The research described in this paper is made possible through National Health and Medical Research Council (NHMRC) Project Grant ID: 284283 and Professorial Fellowship ID: 284363. It received the approval of the Monash University Human Research Ethics Committee (SCERH).

ABSTRACT

The AQoL-8D project was undertaken because of the need for a generic instrument which is sensitive to variations in the quality of life of people with abnormal, psycho-social conditions. Initially the instrument was designed to assist in the context of mental health disease. Early experience with the instrument suggests that it may have much wider application as numerous situations impact (positively or negatively) upon a person's psycho-social state.

The present paper firstly presents the rationale for the AQoL-8D – the need for a psycho-social quality of life, utility instrument. Secondly, it summarises the methods adopted in the project and the data that have been gathered. Finally it reports the structure of the descriptive instrument and its psychometric properties.

An appendix includes an algorithm for the use of the AQoL-8D as a 'psychometric instrument' (ie where the weights are obtained by the simple addition of item response order).

It is also available on the Monash AQoL website: http://www.buseco.monash.edu.au/centres/che/aqol/ . An algorithm for utility weights will be completed and on the AQoL website by December 2009.

Results from the associated surveys are presented in (Khan, Richardson et al. 2009). The measurement methods employed are described in Iezzi (2009). The psychometric analysis and item selection for the descriptive system are in Hawthorne (forthcoming) and the final algorithms in Sinha (forthcoming). Validation studies in different disease areas will follow.

TABLE OF CONTENTS

List of Tables

List of Figures

The AQoL-8D (PsyQoL) MAU Instrument: Overview September 2009

1 Introduction

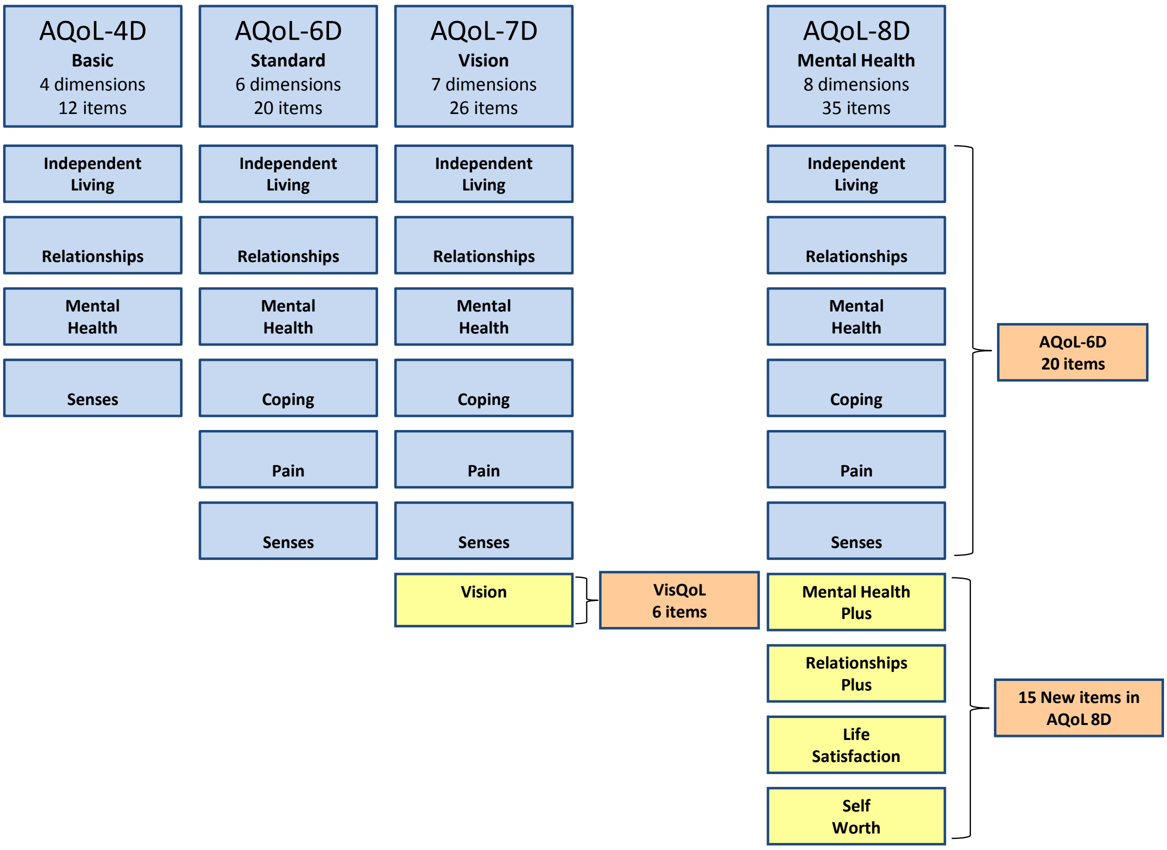
The overall aim of the AQoL-8D project was to develop a mental health module for the Assessment of Quality of Life (AQoL-6D) instrument for the measurement of particular mental health states and to develop a fully integrated and unified scoring system for it. The instrument is the fifth in the AQoL suite of instruments (see Figure 1).

The Burden of Disease (BOD) associated with mental health is likely to increase significantly in the next few decades and depression alone will create a BoD which will be second only to cardiovascular disease (1996). This will lead to an escalating demand for mental health services. With the development of costly new generation pharmaceuticals, as well as other treatment modalities such as cognitive behavioural therapy and social interventions, there will be a significant strain upon the health budget. For this reason the assessment of mental health interventions, including the measurement of health states before and after treatment has become an important focus of health research and policy. Tension between the alternative uses of health resources will inevitably escalate. At present there is no satisfactory instrument for determining the appropriate allocation of resources between interventions for mental and other health problems. This makes public health policy vulnerable to ad hoc decision making and the wastage which that implies.

In the measurement literature there are two streams of work related to this problem. First, a large number of psychometric, disease-specific instruments have been developed to describe (and crudely quantify) the disutility associated with mental health problems. Secondly, economists have developed a small number of generic Multi Attribute Utility (MAU) instruments. These seek to measure the strength of preference (utility) for different health states and, using as a common metric the change in the Quality Adjusted Life Years (QALYs), they seek to establish a ‘level playing field’ between disparate interventions.

While ‘generic’ instruments purport to measure all health states, in practice they do not (and cannot) do this. First, ‘health’ states may be conceptualised in terms of ‘disability’ and ‘impairment’, which are a ‘within the skin’ description of the body. However, the disutility or wellbeing of a single health state described in this way might vary significantly with the context (eg the availability of wheelchairs and ramps for paraplegics). Second, while a ‘handicap based’ descriptive system describes health in terms of the effect upon functioning in a social context, these contexts may also vary significantly and the descriptive system may be more or less accurate in a particular context (eg the stigma of mental disorders can mean a person in recovery cannot find a job).

Figure 1. Structure of Assessment of Quality of Life (AQoL) instruments



Note: For a description of AQoL-Bréf (a reduced form of AQoL) see Hawthorne 2009.

Finally, the sensitivity of an instrument is limited by the number of measured health states, which in the literature have varied from 27 (Rosser-Kind) to over a billion (AQoL-6D, HUI III). In the largest instrument comparison to date, Hawthorne, Richardson and Day (2001) found enormous discrepancies attributable to the differences in descriptive systems of the instruments used. The general conclusion from this is that while 'generic instruments' cover many health states and place utility scores on these, they minimally require 'validation' in different contexts. This implies they should be augmented or reconstructed for specialist purposes. The AQoL-6D was specifically designed to permit such augmentation and a second instrument, the AQoL-7D (Vis-QoL), has been completed which is a specialist vision related version of AQoL-6D.

The specific aims of the AQoL-8D instrument were as follows:

i. To develop a 'descriptive system' which was simultaneously suitable for use as a generic instrument but also which improved the accuracy (sensitivity) of measurement of the major categories of mental illness including neuroses (depression, anxiety disorders, and Post Traumatic Stress Disorder), psychotic disorders and substance abuse problems but also for the measurement of good (ie above average) mental health states;

ii. To develop 'scaling systems'— sets of importance weights—which will indicate the utility value and dimension scores of different health states for the general Australian population;

iii. To contrast the relationship between utility scores obtained from the general population and from patients who had experienced the health state and, where significant differences were found, to provide alternative scaling systems specific to particular patient groups;

iv. To validate the AQoL-8D using two disease-specific instruments, namely Lehman Quality of Life Interview (Lehman 1988), and the Quality of Life in Depression Scale (QLDS) (McKenna and Hunt 1992); and two subjective wellbeing scales, the Life Satisfaction and the Personal Wellbeing Index (Cummins et al. 2003); and

v. To provide population norms for psychiatric patients and for the general population.

2 Background

Measurement of Quality of Life in Mental Disorders

The measurement of HRQoL in mental health disorders has lagged measurement for physical conditions. In general medicine there has been f a narrow focus upon disease related functional capacity (eg mobility) and exclusion of more social measures of quality of life (QoL) such as relationships and independent living (Guyatt, Feeny et al. 1993; Katschnig 1997). However in mental health both the social and 'subjective' features of HRQoL are important as they are intrinsically related to psychopathology and the treatment of mental disorders (Katschnig 1997; Awad and Vorunganti 2000). The subjective evaluation of HRQoL for physical conditions has been widely accepted and used, but it has been commonly believed that people with mental disorders lacked the necessary insight to accurately evaluate their QoL (due to their altered mental states). This perception is changing and recent studies have demonstrated such people can validly evaluate their HRQoL (Herrman, Hawthorne et al. 2002).

Several QoL measures have been developed for use with people with mental disorders. The majority of these are based on a general QoL framework which includes functional status, access to resources and opportunities and a sense of well being across different life domains affected by health (eg housing and income) (Lehman 1997). However the item content, focus and psychometric properties of these measures varies ((Lehman 1997; Gladis, Gosch et al. 1999)). The majority of studies which measure broad QoL in mental disorders use such scales. In contrast, the HRQoL framework explicitly excludes non-health QoL indices (or social measures of QoL) (Guyatt, Feeny et al. 1993; Lehman 1997). Most utility-based HRQoL indices fit this framework (see next section). Smith, Avis et al (1999) argue that such measures do not adequately capture true QoL for people with psychiatric conditions due to their excessive focus on functional capacity.

A number of disease specific quality of life measures have also been developed specifically for use in schizophrenia, depression and more recently anxiety and personality disorders (Mendlowicz and Stein 2000; Narud and Dahl 2002). These measures tend to contain more symptomatic measures than general scales and are problematic due to lack of construct discrimination between symptomatic, functional and broad QoL indices (Gladis, Gosch et al. 1999).

Recently researchers have investigated the use of utility based HRQoL measures in people with mental disorders. Examples include a validation study of the AQoL-4D and a generic HRQoL measure (the WHOQOL-Brèf) in people with a long-standing psychotic disorder (Herrman, Hawthorne et al. 2002). A utility measure for use in major depression has also been developed

(McSAD), though the disadvantage of this scale is that it cannot be used with other mental disorders (Bennett, Torrance et al. 2000). Three studies have attempted to derive utilities from health state descriptions pursuant to depression and schizophrenia. One used the SF-12 as the basis for depression health state descriptions (Lenert, Sherbourne et al. 2000), another used a symptomatic measure (the PANSS) as the basis for health state descriptions in schizophrenia (Chouinard and Albright 1997) and the third asked people with schizophrenia to value different health states pursuant to this condition using five different valuation techniques (Voruganti and Awad 2000). This work is promising in that such approaches appear to be feasible in mental disorders.

Mendlowicz and Stein (2000) suggest the ideal way to measure QoL, particularly for mental disorders, is to use a generic and disease specific instrument. These would be combined through integrating the descriptive systems and providing two alternate scoring methods. This would allow a generic instrument to have modules, which investigate the 'special' characteristics of the disease in question. This is similar to the approach adopted by the AQoL-8D reported here.

Utility Measurement and MAU Instruments

Broadly, HRQoL may be measured one of two ways: by scaling specific health state scenarios or through the use of a HRQoL *instrument* (which fits a health state into a pre constructed 'descriptive system') (Torrance 1986). Two types of HRQoL *instrument* exist, viz (1) disease-specific instruments which only provide a profile; and (2) those providing a single index. A very large number of disease-specific and a smaller number of generic profile instruments now exist (Bowling 1995). These cannot, however, be used for economic evaluation as the different dimensions or items cannot be combined to provide a single valid index of HRQoL or utility which has the properties required for an economic evaluation and particularly a ('strong') interval property (Richardson 1994).

For the widespread use of Cost Utility Analysis (CUA) in the health field there must be a reliable, sensitive and valid multi-attribute utility (MAU) instrument capable of measuring a wide range of health states. Six have been commonly used; viz, Rosser, Kind (1978); the QWB (Kaplan, Bush et al. 1976), the 15D (Sintonen 2001) SF6D, the HUI Mark I, II and III (Feeny, Torrance et al. 1996) and the EQ5D (originally the EuroQoL (Kind 1996).

Early instruments have serious defects (Froberg and Kane 1989); (Nord, Richardson et al. 1993). Except for the SF6D and AQoL none derived their 'descriptive system' using accepted psychometric techniques and consequently *inter alia* structural independence is unknown (Hawthorne, McNeil et al. 1996). Instruments also have questionable sensitivity (Hawthorne, Richardson et al. 2001). The HUI Mark III is explicitly a 'within the skin' instrument, ie it does not purport to measure the importance of role function or social interaction. The EuroQoL instrument descriptive system describes only 147 health states and remains insensitive (Kind 1996; Hawthorne and Richardson 2001). In the largest comparative study to date Hawthorne, Richardson and Day (Hawthorne, Richardson et al. 2001) found variation in the utility scores provided by 5 instruments, viz, the EQ5D, HUI III, AQoL-4D, SF 36 (Brazier 1 weights (Brazier, Roberts et al. 2002)) and the 15D. Using results from 878 respondents the correlation between instruments varied between 0.66 and 0.80. Two main reasons for this low correlation were identified, viz, (i) differences in the models and scaling techniques employed; and (ii) omissions or differences in the descriptive systems; that is, instruments' sensitivity varied enormously between different health states. This implies the need for health state specific validation of instruments, a requirement widely acknowledged in the psychometric but not the economics literature.

In sum, the state of MAU instruments is highly imperfect and remains in its developmental and experimental phase.

Instrument Construction: There are well defined steps for the correct construction of instruments. These steps are as follows:

(i) as there are alternate ways of conceptualising the quality of life an overall structure must be postulated (the latent construct or universe). For example, QoL may be described in terms of an individual's impairment, disability or handicap (ie functional status) and, within each of these conceptual frameworks, different dimensions of health may be postulated;

(ii) a series of items are collected which describe the dimensions and sub-dimensions postulated;

(iii) items are analysed initially for content, coherence and grammatical consistency and subsequently statistically analysed to obtain the descriptive system – the final set of questions which encapsulate the dimensions and sub-dimensions;

(iv) the instrument must be tested to ensure there is structural independence – a particular element or aspect of health should not be included in multiple items (redundancy) – and preference independence – (simplifying) the importance of an element or aspect of health should not depend upon the level of health in another dimension;

(v) the final descriptive system (list of items) must be 'scaled' – utility scores must be attached. As the number of health states is very large this must be achieved by 'modelling' health states; that is, health states are inferred by extrapolation or interpolation from direct measurement;

(vi) the resulting model must be tested to ensure that the implied life-death 'exchange rate' is valid; and

(vii) scores from the final instrument (the manifest model) are then used in a series of 'validation' studies to verify that the relationship between the manifest and latent models is isomorphic. This is usually achieved through tests of concurrent and predictive validity. Concurrent validity is where instrument scores predict a series of health states which may be independently evaluated by an instrument or instruments of known validity. Predictive validity is where instrument scores predict future health states, or the consequences of current health states.

AQoL-4D was constructed to address the various problems with existing instruments. More specifically, its objectives were (i) to create an instrument with 'construct validity' by the use of the procedures summarised above (Hawthorne and Richardson 1995); (ii) to increase the sensitivity of the descriptive system by employing a nested or hierarchical structure which permitted item overlap – redundancy – within the dimensions in order to achieve sensitivity, but maintaining orthogonality between dimensions; (iii) to use a flexible multiplicative model for combining attributes (as used with the HUI instruments); and (iv) to obtain Australian utility weights. Confirmatory factor analysis on AQoL-4D indicated excellent psychometric properties. A comparative fit index of 0.90 was obtained. By December 2002 60 research teams had adopted AQoL-4D. Results have been good with AQoL generally performing as well as or better than other instruments (Hawthorne, Richardson et al. 1999; Hawthorne, Richardson et al. 2001). Where the AQoL has been used in specific studies, the results have demonstrated its appropriateness, reliability, validity and sensitivity (Osborne, Hawthorne et al. 2000; Herrman, Hawthorne et al. 2002; Sturm, Osborne et al. 2002).

AQoL-6D was designed to improve AQoL-4D in several methodological respects. These were (i) to increase instrument sensitivity in the range of full health; (ii) to include items appropriate for evaluating health promotional programs; (iii) to adopt 'deliberative weights' – scores obtained after deliberation; (iv) to employ a 2-stage modelling procedure in which an econometric correction is made to the initial multiplicative model scores; (v) to include person trade-off (PTO) weights to obtain a QALY-DALY exchange rate; (vi) to include a ratings scale in the deliberative system. AQoL-6D was also designed to facilitate the addition of 'modules' for specified purposes. As noted earlier it has been extended and validated for vision related disease AQoL-7D (VisQoL). The comparative fit index of the AQoL-6D descriptive system is 0.94. In January 2002 interviews for the scaling of AQoL-6D were completed. This instrument was completed in July 2002.

3 Methodology

Construction of the AQoL-8D involved four broad steps:

(i) construction of the descriptive system, conceptualisation, construction survey and statistical analysis;

(ii) construction of stage 1 weights for instrument dimensions and the overall instrument;

(iii) construction of stage 2 weights to obtain the final dimension and instrument models; and

(iv) instrument validation: exploration of properties in comparison with other instruments

At the time of writing (August 2009) the first two stages are completed and data for the remainder of the analysis has been collected. A preliminary psychometric weighting system has been constructed and placed upon the Centre for Health Economics website: http://www.buseco.monash.edu.au/centres/che/aqol-website/aqol-page-1.html

3.1 Construction of the Descriptive System

The relationship between the stages of the analyses and the data collection for the descriptive system are summarised in Figure 2 and described below.

(a) Conceptual basis: For the reasons described in Richardson et al. (2007) the AQoL-8D adopted the same concept of health – handicap – as the previous AQoL instruments. In sum, it is postulated that quality of life (QoL) is best conceptualised and measured in a social context: that is, in terms of how health related problems impact upon a person's life. This basic conceptualisation is supplemented, when necessary, with elements of disability and impairment. The concept was operationalised by postulating dimensions of QoL – life satisfaction, activities of daily living, etc and identifying or creating items which encompassed these.

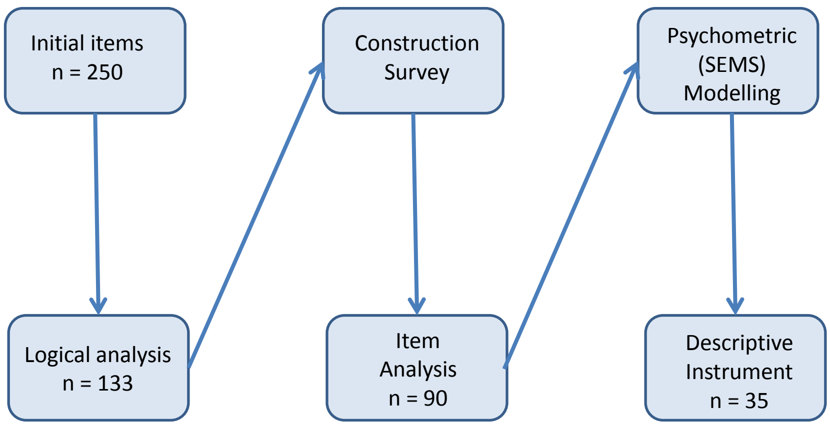
(b) Item bank and survey 1: An item bank of 250 items was constructed which included items from the AQoL-8D item bank and items from other generic and *disease-specific* instruments (such as the Lehman Quality of Life Interview (Lehman 1988). There were a number of focus groups with patients (Survey 1) and interviews with mental health professionals to generate additional items and to review existing items. This step was to 'validate' the items selected and to suggest new items; it was not to generate the entire databank. Focus groups continued to be convened until no new information could be elicited (ie saturation). There were 29 participants in four groups.

(c) Item critical analysis: New items in the item bank were subject to linguistic, logical and content analysis to ensure suitability for the final structure of the instrument. A reduced number of items (90) were selected. Response categories were reviewed to ensure sensitivity to mental health in the domain of good health.

(d) Construction survey: Survey 2 administered the selected items to a stratified population including the Australian general public and patients in the target groups. Sampling procedures are described under Survey 3 (below).

(e) Data analysis and question selection: Data obtained in Survey 2 were subject to principle component, exploratory factor and structural equation modelling. The objective was to validate the dimension structure of AQoL-6D in the present context and to identify one or more dimensions relevant for psychiatric health states. Results of this stage of the analyses are to be reported in Hawthorne et al. (forthcoming).

Figure 2. Construction of the descriptive system



3.2 Construction of Stage 1 and Stage 2 weights

Two methods have been used by economists to obtain utility scores (to ‘scale’ or ‘calibrate’) multi attribute instruments. These involve the use of decision analytic (DA) theory to construct an additive or multiplicative model (such as the 15D and HUI 1, 2, 3 instruments) or the econometric ‘prediction’ of independently observed multi attribute health state utilities from the single attribute (item) scores. The resulting econometric equation is adapted as the scaling algorithm. The advantage over the DA approach is that the prediction must produce scores which are the correct order of magnitude if the utilities are correctly measured and the econometrics is valid. (Regressions must pass through the observed utility points.) However the method limits the size of the instrument which may be scaled as the feasible number of observations limits the number of variables in the analysis (AQoL-8D, for example, would require 176 independent variables).

Following the former, DA, approach all of the AQoL instruments commence with a multiplicative model. This is similar to equation 1 below.

$$U(AQoL) = U_1 * U_2 * U_3 * U_4 * U_5 * U_6 \ldots (1)$$

The actual model is somewhat more flexible. It is calculated using disutilities rather than utilities and these are adjusted for the relative importance of each of the dimensions (1)-(6). This results in equation 2 in which w_i are the dimension (or item) weights and k is the overall scaling constant and is similar to the requirement in an additive model that the dimension weights sum to unity. It is obtained by solving equation 3. .

$$DU = \frac{1}{k}\left[\prod_{i=1}^{n}\left[1 + kw_i DU_i\left(x_{ij}\right)\right] - 1\right] \quad \ldots (2)$$

$$k = \prod_{i=1}^{n}(1 + kw_i) - 1 \quad \ldots (3)$$

The relationship between utility and disutility is given in equation 4.

$$U^* = 1 - DU^* \quad \ldots (4)$$

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

All of the AQoL instruments departed from other instruments in having a multi level structure. Dimension scores are calculated using a multiplicative model and these are then used to calculate the multiplicative score for the entire health state. The resulting equations for AQoL-6D are shown in Appendix 2 which also illustrates their use.

AQoL-4D (the original AQoL) made no further adjustment, and the multiplicative score was adopted as the estimate of utility. The validity of the procedure depended upon the assumption that dimensions were structurally independent (orthogonal) and that there would be no 'double counting' of elements of poor health. The assumption is difficult to demonstrate as it does not imply zero correlation as different elements of poor health tend to occur simultaneously. Nevertheless the correlations in AQoL-4D were low and within the range which is usually accepted as indicating orthogonality in psychometric studies.

With the increasing complexity of AQoL-6D this assumption became untenable and an additional 'stage 2' correction was employed. This 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 was fitted to this data.

$$TTO = AQoL^x$$

$$x = \alpha_0 + \sum_j \alpha_i D_i + \sum_i \sum_j B_{ij} D_i D_j + \sum_{i=1}^{4} shift_i \quad \ldots (5)$$

Where α_0 = constant

D_i = dimension score for dimension i

D_iD_j = dimension Di times Dimension D_j

$shift_i$ = dummy variables indicating that the TTO has a disutility score in excess of 0.2; 0.4; 0.6; 0.8.

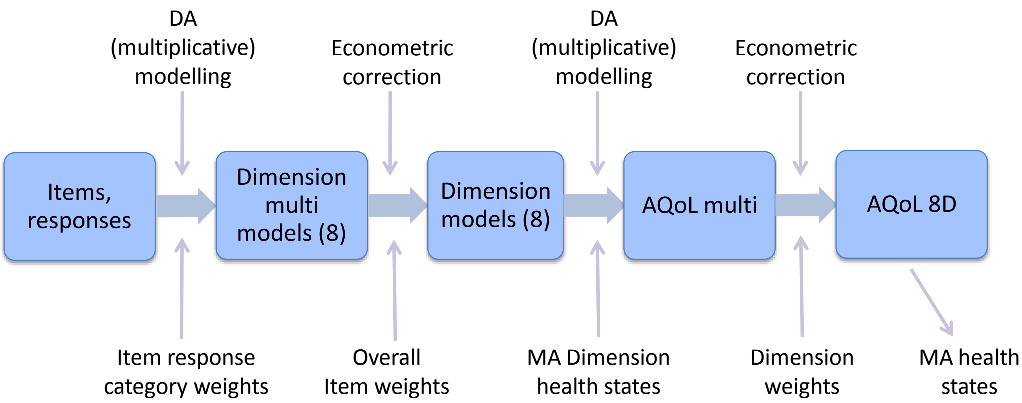
This led to a ‘correction’ to dimension and overall scores but left item response and importance weights unchanged.

The AQoL-8D is employing a similar second stage correction in the calculation of the final utility but is, additionally, carrying out a similar correction to each of the dimensions.

3.3 Weights Survey (Survey 3)

For reasons discussed elsewhere (Richardson 1994; Richardson 2002) the time trade-off (TTO) technique was used to measure utility (as with AQoL-4D and 6D). This element of the program is conventional as the TTO is possibly the most widely used method for measuring ‘utility’. The task was complex and required one-on-one interview with additional linked data collected by a prior VAS questionnaire. The VAS scale was used to obtain results for item response level weights (A data), overall item weights (B data), dimension weights (C data) and some multi attribute health states within dimensions (Ed data). TTO data were also collected for the latter category and for E data. Following the procedures used in the construction of AQoL-6D, a transformation between VAS and TTO was estimated and used to convert all VAS to TTO equivalent values.

Figure 3. 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

The relationship between the types of data and the stages of analysis is shown in Figure 3. 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. Following the recommendation of decision analytic theory item worst disutilities were used as item weights (B data). These were obtained by rating the worst outcome for each item on a 0-1 disutility scale where the endpoints were defined by the best and worst item outcomes which were assigned the weights 0.00 and 1.00 respectively. Item weights allowed the construction of 8 multiplicative models, 1 per dimension. This procedure only requires the calculation of the multiplicative scaling constant as described below.

Unlike AQoL-6D, the AQoL-8D dimension multiplicative models were subject to an econometric correction to allow for structural and preference dependence. The correction required the use of multi attribute dimension health states (Ed data), a procedure paralleling the final stage correction in AQoL-6D. The Ed data which are holistically evaluated are regressed upon the values calculated from the multiplicative models and from the items (the independent variables). The resulting econometric relationship is adopted as the dimension model. This procedure is repeated for each of the 8 dimensions.

The final AQoL model is derived in a way that is analogous to the dimension models. Initially a multiplicative model is calculated using the dimension weights (C data) ie the value of the dimension worst outcome assessed on a 0-1, AQoL best-death scale. Multi attribute health states (E data) are independently collected which span all items and dimensions. These are regressed on the calculated score from the AQoL multiplicative model and the 8 dimension scores from the final dimension models. The best fitting econometric relationship is adopted as the AQoL-8D model.

Upon completion these results will be entered into a program and available on the web for use as presently occurs with the AQoL-6D http://www.buseco.monash.edu.au/centres/che/aqol-website/instruments-and-algorithms.html

Population sampling: The 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. Based on the Australian Bureau of Statistics Socio-Economic Indicators for Areas (SEIFA) scores, postcodes were the primary sampling unit, with probability proportionate to population size (to reduce the effect of socio-economic confounding). From these postcode areas, telephone subscribers (18 years+) were sampled. Subscribers were contacted by letter and subsequently by telephone. The use of post-codes as the primary sampling unit meant that informants would be fairly tightly clustered, minimizing the travel costs. These procedures were also employed in AQoL-4D.

Patient sampling: People with mental disorders ('neuroses' and psychotic disorders) were accessed via a number of channels. Mental Health Services (such as St Vincent's Mental Health) were used to recruit subjects. Case-managers and treating clinicians were approached to ensure people were well enough to participate in the study. Other organisations such as The Melbourne Clinic and treatment providers for posttraumatic stress disorder were also approached to assist in the recruitment of people with non-psychotic disorders. Informed consent was obtained from all potential subjects.

Survey design and field procedure: Past experience indicated that to obtain a satisfactory response rate it would be necessary to make a payment in compensation for time and travel costs. Interviews for the public took place at the AQoL research office. Patients were all interviewed in their treatment service site.

Utility weights were obtained using the TTO technique (as with AQoL-4D and 6D). Depending upon the magnitude of the burden on individual patients, we included a limited number of Person Trade-Off (PTO) questions (as with the construction of AQoL-6D) and also a 'self-TTO' (a TTO based on respondent's current health). These data permit the construction of two alternative scaling systems. The Centre for Health Economics had a team of interviewers experienced in the application of HRQoL instruments, and in the use of TTO and PTO techniques. This reduced both interviewer training costs and data measurement error. The order of items was randomly varied to check for framing effects. The scaling survey included sufficient multi attribute states –

descriptions spanning more than one dimension – to allow 'internal validation' and a subsequent econometric analysis of the multiplicative model and its performance against these health states.

Sample size for survey 3: Survey 3 comprised two strata: (a) a representative sample of the Australian population; and (b) a representative sample of those with mental health disorders. The reason was to provide two sets of weights either or both of which could be selected depending upon the purpose of the study. Given that the veracity of the AQoL-8D depended upon the representativeness of the utility weights, the sample size should be based on the minimum acceptable sampling errors — in this case 5% was set, resulting in the need to randomly recruit 400 cases in each strata (Hoinville, Jowell et al. 1977) ie a total of 800 cases.

4 Results

Results from the construction survey and their analyses are shown in Tables 1 and 2 and Figure 4. In total 711 individuals completed the survey, 514 patients and 197 members of the public (Table 1). There were more respondents resident in SIEFA 5 postcodes, but the sample was otherwise evenly distributed by SES area (Table 2). After SEM analysis the final instrument had 8 dimensions and 35 items. It takes about 10 minutes to complete. Details of the data are given in Richardson and Iezzi (2009).

Table 1. Respondents to the construction survey by age and gender

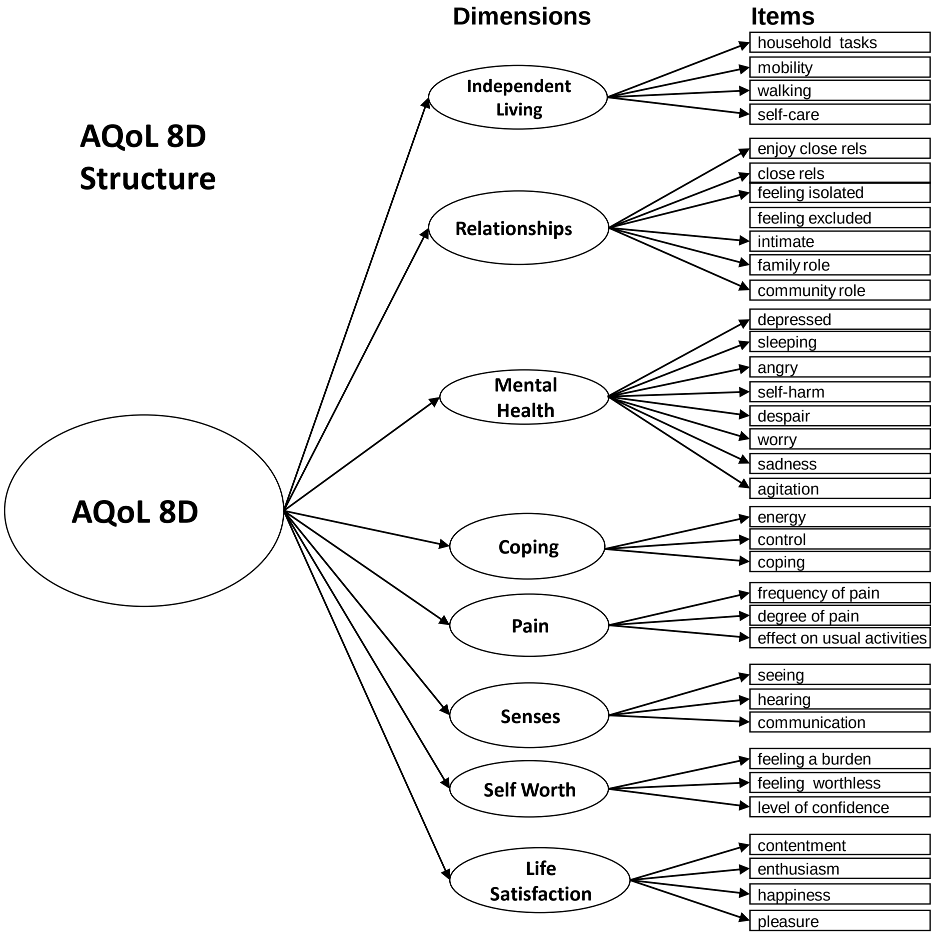
Age	Public			Patient			Grand Total
	Male	Female	Total	Male	Female	Total	
18-24	1	6	7	109	42	151	158
25-34	5	21	26	60	64	124	150
35-44	13	25	38	51	63	114	152
45-54	24	32	56	41	52	93	149
55-65	28	42	70	11	21	32	102
	71	126	197	272	242	514	711

Table 2. Respondents to the construction survey by SEIFA group of residence*

SEIFA	Male	Female	Total	Male	Female	Total	Grand Total
1	14	25	39	47	25	2	111
2	10	24	34	48	28	76	110
3	24	27	51	24	22	46	97
4	13	21	34	34	32	66	100
5	10	29	39	75	89	164	203
	71	126	197	228	196	424	621

SEIFA Missing = 90

Figure 4. Gerry diagram



Results from the weights (scaling) survey are given in Tables 3, 4, and 5. In sum, 670 individuals completed the interviews, 323 patients and 347 members of the public (Table 3). The importance of items varied with the ratio of highest to lowest importance weights varying between 22 and 215 percent (Table 4). The ratio is of interest as it indicates the difference between the treatment of items in the AQoL-8D and the treatment in a 'psychometric' instrument (with equal weights). Table 4 provides similar information for dimensions. The mean TTO of dimension worst health states varied from 4.61 to 6.38 implying disutility scores (which are used as dimension weights) of 0.535 and 0.362 respectively, a variation of 47.8 percent.

Of particular interest, all of the patient dimension weights in Table 5 were significantly below the weights of the general population which is contrary to the common view that patient's scores will be higher because of their adaptation to health states which the general population deems worse because they do not envisage adaptation. Our results suggest that, to the contrary, experiencing mental health states may make them appear to be worse.

To carry out second stage corrections a total of 2787 multi attribute 'within dimension' dimensions were made, an average of 348 per dimension. As reported in Richardson and Iezzi (2009), the frequency distributions of all of these span the range of utility scores from 1.00 to 0.00 (death) and include a small number of worse than death assessments. For the final stage 2 correction 2989 multi attribute TTO scores were collected. The frequency distribution of these is shown in Figure 5. As with the dimension distributions, the range of values is satisfactory for the second stage analysis.

Table 3. Respondents to weights survey

Age Group	Public			Patient			Grand total
	Male	Female	Total	Male	Female	Total	
18 to 24 years	22	31	53	22	23	45	98
25 to 34 years	38	36	74	35	30	65	139
35 to 44 years	32	35	67	35	43	78	145
45 to 54 years	26	43	69	41	29	70	139
55 to 64 years	30	42	72	44	15	59	131
65 years +	7	5	12	3	3	6	18
Total	155	192	347	180	143	323	670

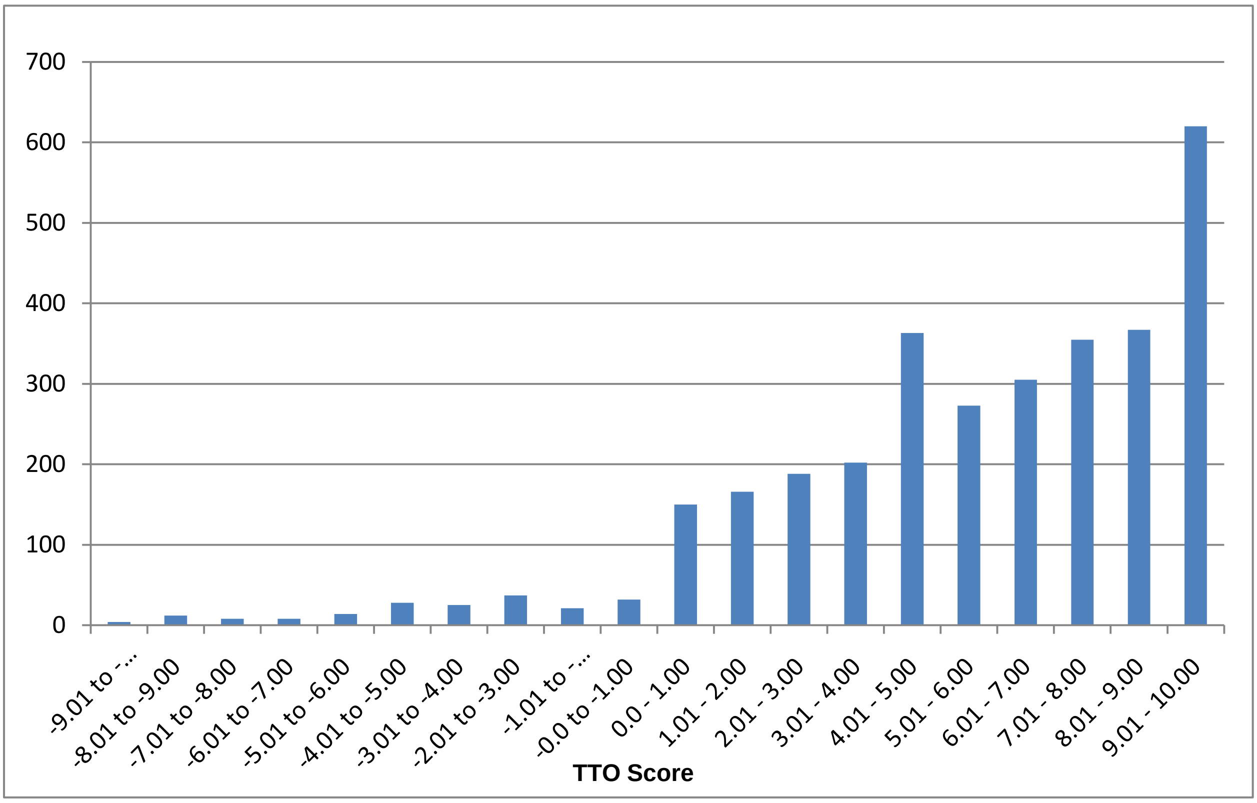
Table 4. Highest/lowest item worst scores by dimension

Dimension	Highest / Lowest	=	Factor of variation within dimension
Independent living	46.6 / 14.8	=	3.15
Life satisfaction	39.1 / 20.3	=	1.93
Mental health	31.2 / 14.9	=	2.09
Coping	38.5 / 21.6	=	1.78
Relationships	39.7 / 25.5	=	1.56
Self worth	32.3 / 25.7	=	1.26
Pain	32.6 / 26.7	=	1.22
Senses	35.3 / 25.9	=	1.36

Table 5. Dimension worst TTO scores best-death (10-0) scale

Dimension	Mean			Sig
	Public	Patient	Total	
Independent living	5.8	4.74	5.30	.001
Life satisfaction	6.48	5.20	5.84	.000
Mental health	5.06	4.38	4.72	.025
Coping – TTO	7.18	5.64	6.38	.000
Relationships	5.77	4.57	5.18	.000
Self worth	6.89	5.49	6.18	.000
Pain	5.02	4.29	4.65	.021
Senses	5.29	4.64	4.98	.076

Figure 5. Frequency Distribution of Multi-Attribute (MA) TTO Scores, N = 3178



5 Endnote

This paper has summarised the rationale and methodology of the AQoL-8D project. In particular it has sought to show the relationship between the methods used and the data collected. Further details of the project are provided elsewhere. Survey results are summarised in (Khan, Richardson et al. 2009), survey methodology and databases in Iezzi (2009a, b), psychometric analysis and item selection in Hawthorne et al. (forthcoming) and modelling of final utility scores in Sinha (2009).

References

Awad, A. and L. Vorunganti (2000). "Intervention research in psychosis: issues relating to the assessment of quality of life." Schizophrenia Bulletin **26**(3): 557-564.

Bennett, K., G. Torrance, et al. (2000). "Development and testing of a utility mesure for major, unipolar depression (McSad)." Quality of Life Research **9**(1): 109-120.

Bowling, A. (1995). Measuring Disease. Philadelphia, Open University Press.

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.

Chouinard, G. and P. Albright (1997). "Economic and health state utility determinations for schizophrenic patients treated with risperidone or haloperidol." Journal of Clinical Psychopharmacology **17**(4): 298-307.

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.

Froberg, D. and R. Kane (1989). "Methodology for measuring health state preferences." Journal of Clinical Epidemiology **42**(5): i-iv.

Gladis, M., E. Gosch, et al. (1999). "Quality of life: Expanding the scope of clinical significance." Journal of Consulting and Clinical Psychology **67**(3): 320-331.

Guyatt, G., D. Feeny, et al. (1993). "Measuring health-related quality of life." Annals of Internal Medicine **118**(8): 622-629.

Hawthorne, G., H. McNeil, et al. (1996). Endpoints: Issues in the Meaning of "Good Health" in Health Related Quality of Life Measurement. Melbourne, Centre for Health Program Evaluation.

Hawthorne, G. and J. Richardson (1995). An Australian MAU/QALY Instrument: Rationale and Preliminary Results. Health Outcomes and Quality of Life Measurement. Canberra.

Hawthorne, G. and J. Richardson (2001). "Measuring the value of program outcomes: A review of multi attribute utility measures." Expert Review of Pharmacoeconomics Outcomes Research **1**(2): 215-228.

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. (2001). "A comparison of the Assessment of Quality of Life (AQoL) with four other generic utiliyt instruments." Annals of Medicine **33**(5): 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.

Herrman, H., G. Hawthorne, et al. (2002). "Quality of life assessment in people living with psychosis." Soc Psychiatry Psychiatr Epidemiol **37**: 510-518.

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

Kaplan, R., J. Bush, et al. (1976). "Health status: Types of validity and the index of wellbeing." Health Services Research **11**(4): 478-507.

Katschnig, H. (1997). How useful is the concept of quality of life in psychiatry? Quality of Life in Mental Disorders. H. Katschnig, H. Freeman and N. Sartorius. West Sussex, John Wiley & Sons**:** 3-6.

Khan, M., J. Richardson, et al. (2009). A New Method for Measuring Quality of Life: The AQoL 8D (PsyQoL) Instrument, Research Paper XX. Melbourne, Centre for Health Economics, Monash University.

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.

Lehman, A. (1988). "A quality of life interview for the chronically mentall ill." Evaluation Program Planning **11**: 51-62.

Lehman, A. (1997). Instruments for measuring quality of life in mental illness. Quality of Life in Mental Disorders. H. Katschnig, H. Freeman and N. Sartorius. West Sussex, John Wiley & Sons**:** 79-107.

Lenert, L., C. Sherbourne, et al. (2000). "Estimation of utilities for the effects of depression from the SF-12." Medical Care **38**(7): 763-770.

McKenna, S. and S. Hunt (1992). "A new measure of quality of life in depression: Testing the reliability and construct validity of the QLDS." Health Policy **22**(3): 321-330.

Mendlowicz, M. and M. Stein (2000). "Quality of life in individuals with anxiety disorders." American Journal of Psychiatry **157**(5): 669-682.

Murray, C. J. and A. D. Lopez (1996). The Global Burden of Disease. Boston Massachusetts, The Harvard School of Public Health on behalf of The World Health Organization and The World Bank.

Narud, K. and A. Dahl (2002). "Quality of life in personality disorders." Current Opinion in Psychiatry **15**(2): 131-133.

Nord, E., J. Richardson, et al. (1993). "Social evaluation of health care versus personal evaluation of health states: evidence on the validity of four health state scaling instruments using Norwegian and Australian surveys." International Journal of Technology Assessment in Health Care **9**: 463-478.

Osborne, R., G. Hawthorne, et al. (2000). "Measurement of rapid changes in health outcomes in people with influenza symptoms." Journal of Health Outcomes Research **4**: 15-30.

Richardson, J. (1994). "Cost utility analysis: What should be measured." Social Science & Medicine **39**(1): 7-21.

Richardson, J. (2002). The conceptual basis for summary measures of population health: Qualifying DALYs, dallying with QALYs. Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications. C. J. Murray, J. Salomon, C. D. Mathers and A. D. Lopez. Geneva, The World Health Organization.

Richardson, J., S. Peacock, et al. (2007). Construction and Validation of the Assessment of Quality of Life (AQoL) Mark 2 Instrument. Research Paper 24. Melbourne, Centre for Health Economics, Monash University.

Rosser, R. and P. Kind (1978). "A scale of valuations of states of illness: Is there a social consensus." International Journal of Epidemiology **7**: 4-15.

Sintonen, H. (2001). "The 15D instrument of health related quality of life: Properties and applications." Annals of Medicine **33**(5): 328-336.

Smith, K. W., N. E. Avis, et al. (1999). "Distinguishing between quality of life and health status in quality of life research: A meta-analysis." Quality of Life Research **8**: 447-459.

Sturm, J. W., R. H. Osborne, et al. (2002). "Brief comprehensive quality of life assessment after stroke: The assessment of quality of life instrument in the North East Melbourne Stroke Incidence Study (NEMESIS)." Stroke **33**: 2888-2894.

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

Voruganti, L. and A. Awad (2000). "Assessing health utilities in schizophrenia: A feasibility study." Pharmacoeconomics **17**(3): 273-286.

Appendix.
Multiplicative Disutility Equations

Dimensions	
General Formula	$DU_d = \frac{1}{k}\left[1-\prod_{i=1}^{n}\left(1-kw_i DU_i\right)\right]; k_d > 0$
Independent Living	$DU_1 = 1.02\left[1-\left(1-0.38du_1\right)\left(1.-0.58du_2\right)\left(1-0.62du_3\right)\left(1-0.77du_4\right)\right]$
Social and Family	$DU_2 = 1.08\left[1-\left(1-0.59du_5\right)\left(1.-0.65du_6\right)\left(1-0.47du_7\right)\right]$
Mental Health	$DU_3 = 1.02\left[1-\left(1-0.63du_8\right)\left(1.-0.66du_9\right)\left(1-0.64du_{10}\right)\left(1-0.70du_{11}\right)\right]$
Coping	$DU_4 = 1.08\left[1-\left(1-0.39du_{12}\right)\left(1.-0.60du_{13}\right)\left(1-0.72du_{14}\right)\right]$
Pain	$DU_5 = 1.08\left[1-\left(1-0.69du_{15}\right)\left(1-0.57du_{16}\right)\left(1-0.57du_{17}\right)\right]$
Senses	$DU_6 = 1.18\left[1-\left(1-0.49du_{18}\right)\left(1-0.39du_{19}\right)\left(1-0.51du_{20}\right)\right]$
AQoL General Formula	$DU_{AQoL} = \frac{W}{k}\left[1-\prod_d\left(\left(1-kw_d DUx_i\right)\right)\right]; k > 0$
$DU_{AQoL} = 1.150\left[1-\left(1-0.462DU_1\right)\left(1-0.442DU_2\right)\left(1-0.472DU_3\right)\left(1-0.344DU_4\right)\left(1-0.581DU_5\right)\left(1-0.630DU_6\right)\right]$	

Key: W = the conversion factor between the 0-1 (death, full health) model
k_d = scaling constant
wi = dimension weights
du_{ij} = disutility for the i[th] item of dimension $_j$

Scaling the Multiplicative Model: An Example

Assigning a utility score to a health state involves the following steps.

(i) Complete the AQoL questionnaire and determine the 20 response levels which define the health state.

(ii) Read the 20 item disutility scores, du_i, which correspond with the response levels from Table 3. These 'disutilities' are measured on a (1-0) scale with the item best and worst defining the endpoints.

(iii) Enter the item disutility scores, du_i, into the corresponding equation in Box 1. Calculate the six dimension disutility scores DU_d. These disutilities are measured on a (0-1) scale where the endpoints are the dimension best and dimension 'all worst' (all items at their worst level).

(iv) Enter the six dimension DU_d scores into the final AQoL equation in Box 1. The score obtained is the predicted disutility for the health state.

(v) Convert disutilities into utilities using the equation U = 1 - DU.

These steps are illustrated for a randomly chosen health state (see below).

Calculating a utility score: A numerical example

Step (i) Complete the AQoL questionnaire to obtain 20 response levels; 1 per item

Example: Response levels are:

D 1(1,1,2,1); D 2(2,2,1); D 3 (3,2,3,1); D 4(1,1,1); D 5(2,1,1); D 6(2,1,2)

Step (ii) Read the 20 disutility scores from Table 3

In the example:

$$D1(0,0.04):D2(.07,.19,.65):D3(.39,.14,.33,.00)D4(0,0,0):D5(.13,.0,.0)D6(.03,00,.19)$$

Step (iii) Enter the 20 disutility scores into the equations in Box 1

$$DU_1 = 1.02[1-(1-38*0)(1-.58*0)(1-.62*.04)(1-.77*0)] \quad = 0.03$$

$$DU_2 = 1.08[1-(1-.59*.07)(1-.65*.19)(1-.47*.0) \quad = 0.17$$

$$DU_3 = 1.02[1-(1-.63*.39)(1-.66*.14)(1-.64*.33)(1-.7*0) \quad = 0.40$$

$$DU_4 = 1.08[1-(1-.39*0)(1-.60*0)(1-.72*.0) \quad = 0.00$$

$$DU_5 = 1.08[1-(1-.69*.13)(1-.57*0)(1-.57*.0) \quad = 0.10$$

$$DU_6 = 1.18[1-(.1-.4*0.03)(1-.39*0)(1-.51*.19) \quad = 0.12$$

Step (iv) Enter the DU_i scores into the AQoL formula (above)

$$DU_{AQol} = 1.15[1-(1-.462*.03)(1-.442*.17)(1-.472*.4)$$

$$(1-.344*0.0)(1-.581*.1)(1-.63*.12)] = .42$$

Step (v) Convert disutility to utilities from the equation U = 1 - DU_i

Dimension Utilities = 0.97; 0.83; 0.6; 1.00; 0.9; 0.88

Global U = 0.58

Computerised algorithms for the AQoL instruments are available from the AQoL website: http://www.buseco.monash.edu.au/centres/che/aqol/