Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D

Jeff Richardson · Kompal Sinha · Angelo Iezzi ·
Munir A. Khan

Abstract

Purpose The objective of this paper is to describe the four-stage methodology used to obtain utility scores for the Assessment of Quality of Life (AQoL)-8D, a 35-item 8 dimension multi-attribute utility instrument, which was created to achieve a high degree of sensitivity to psycho-social health.
Methods Data for the analyses were obtained from a representative group of 347 members of the Australian public and from 323 mental health patients each of whom provided VAS and time trade-off valuations of multiple health states. Data were used initially to create multiplicative scoring algorithms for each of the instrument's 8 dimensions and for the overall instrument. Each of the algorithms was then subject to a second-stage econometric 'correction'.
Results Algorithms were successfully created for each of the AQoL-8D's dimensions, for physical and mental 'super-dimensions' and for the overall AQoL-8D instrument. The final AQoL-8D algorithm has good predictive power with respect to the TTO valuations.
Conclusions The AQoL-8D is a suitable instrument for researchers conducting cost utility analyses generally but, in particular, for the analysis of services affecting psycho-social health.

Keywords Quality of life · AQoL · Multi-attribute utility · Cost utility analysis · Psycho-social health

J. Richardson (✉) · K. Sinha · A. Iezzi · M. A. Khan
Faculty of Business and Economics, Centre for Health
Economics, Monash University, Wellington Road, Clayton,
VIC 3800, Australia
e-mail: jeffrey.richardson@monash.edu

Introduction

Economic evaluation studies have increasingly employed multi-attribute utility (MAU) instruments for measuring health state utilities. The instruments consist of a descriptive system (or classification)—a set of health-related questions and response categories—and a scoring system or algorithm, which converts responses into a utility score which measures the preference for the health state on a scale where 0.00 represents death and 1.00 represents the best health state described by the instrument (for descriptions, see Brazier et al. [1]).

A review of published articles listed on the Web of Science between 2005 and 2010 [2] identified 392 cases where a head to head comparison had been made between the results from different instruments. The outcome of these comparisons indicated a relatively low level of agreement between utilities predicted from different instruments. Authors generally concluded that instruments are 'not equivalent' [3] and are 'imprecisely related' [4]. These conclusions are unsurprising as the evaluation methods used to quantify utility, the population groups from which values were obtained and the descriptive systems of the instruments all vary significantly. In using what Dolan et al. [5] describe as a 'deliberately simplified descriptive system' (p. 1036) utility instruments have omitted dimensions nominated by individuals as most important to them [6]. Of the three most widely used instruments—EQ-5D, SF-6D and HUI 3—only SF-6D has a dimension relating to social function and each of the three instruments has only one item to describe the mental health dimension which Helliwell et al. [7] note 'is the single most important determinant of individual happiness (in every case where this has been studied)' (p. 4). While health-related quality of life is a wider concept than mental

health and happiness, their limited description in these instruments suggests the need for a broader descriptive system when these dimensions are important.

The AQoL-8D was constructed in response to this challenge. Its objective was to achieve a high level of sensitivity to psycho-social health states by supplementing the descriptive system of an earlier instrument, the AQoL-6D [8] and to remodel the resulting instrument utilities. Revision of the descriptive system is described in Richardson et al. [9]. A review of ten instruments measuring mental health-related QoL identified 18 relevant health domains (sub-dimensions). Initially, 250 items describing these domains were obtained from the instruments, from four focus groups with mental health patients and (for a small number of domains) from the research team which included, inter alia, two psychiatrists, a psychologist and a personal counsellor. Qualitative and logical analyses reduced these to an item bank of 133. Interviews were

conducted which administered the items to a representative sample of 197 members of the public and to 514 mental health patients. As recommended by McDonald [10], a combination of explanatory and confirmatory factor analyses was used to construct the final instrument which is shown in Fig. 1 along with the goodness-of-fit statistics. These indicate that the descriptive system has good psychometric properties and that each of the three physical and the five psycho-social dimensions derived from the psychometric analyses is associated with an overarching physical or psychological 'super' dimension.

The purpose of the present article is to describe the modelling of the AQoL-8D utilities. This presented a unique challenge because of the size and complexity of the instrument. To date, utility has been modelled using one of two broad approaches. First, decision analytic (DA) procedures have been used by the 15D [11], the three Health Utility Index (HUI) instruments [12–14] and the AQoL-4D
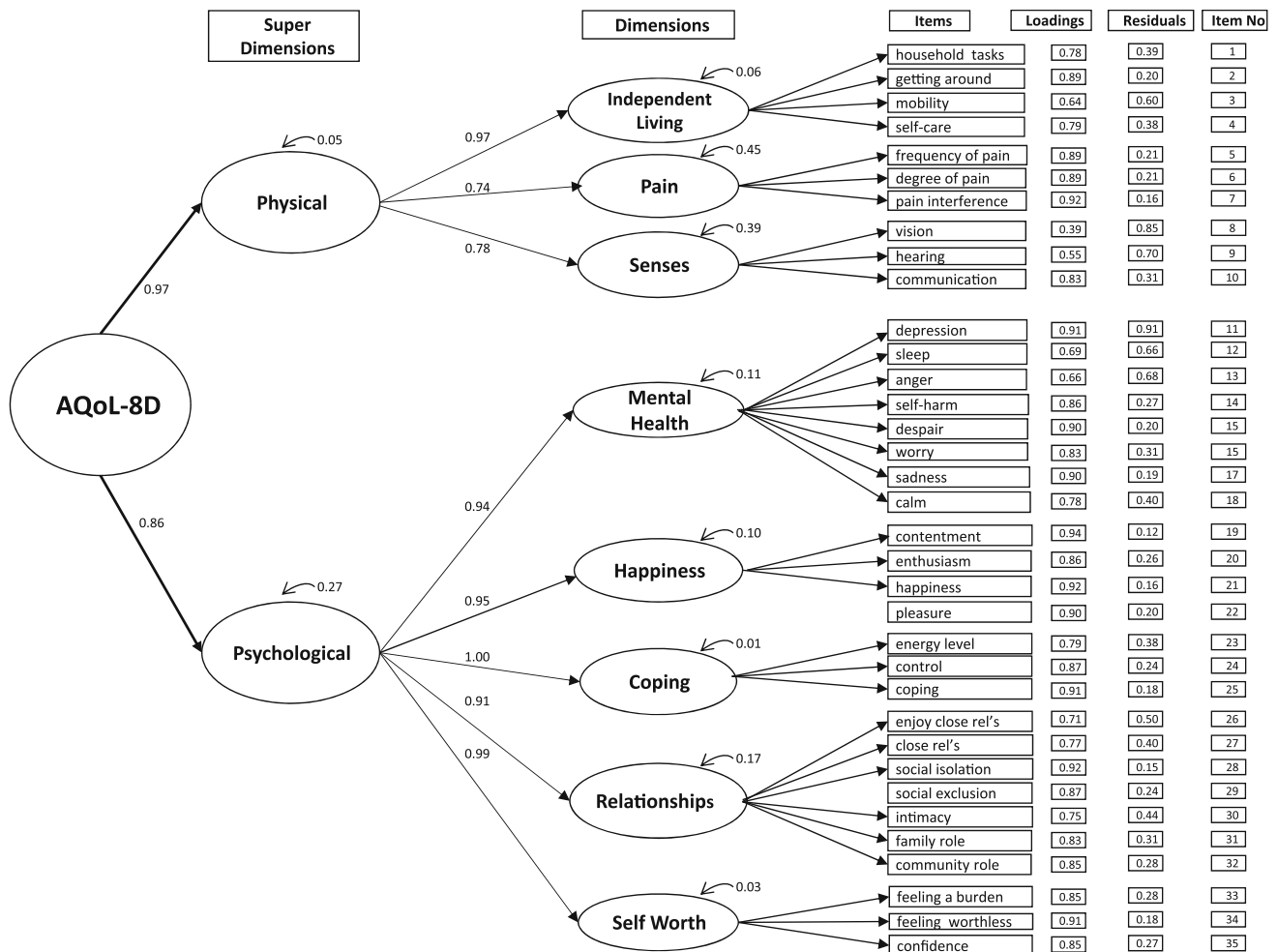

Fig. 1 AQoL-8D structure. *Numbers on joined arrows* indicate correlation. *Numbers on unjoined arrows* are the residual (unexplained) variances of the latent variables. Fit statistics CFI = 0.974, TLI = 0.972, RMSEA = 0.073, WRMR = 1.64. Coefficients were estimated from the combined public patient database ($n = 711$). *Source*: Richardson et al. [9]

[15]. DA theory, however, places strict orthogonality requirements upon items to avoid 'redundancy': the double counting of content which is common to correlated items [16]. This limits the extent to which sensitivity may be increased through the use of multiple, overlapping (but separate) items and dimensions. Secondly, econometric techniques have been used to calibrate the EQ-5D [17] and the SF-6D [18]. The approach avoids the need for strict orthogonality of items but is also subject to a constraint imposed by multi-collinearity as the number of correlated items and dimensions increase.

These constraints highlight a dilemma encountered in MAU modelling in general and a dilemma which the AQoL-8D sought to overcome. Content validity is usually defined by having items (or tasks in a test) which are 'representative of the specified universe' [19, p. 451], and content validity is increased when a measure includes a 'more representative sample of the target behaviour' [20, p. 175]. This commonly implies the need for 3–4 items to achieve a high level of sensitivity within a dimension. However, items describing a single dimension necessarily correlate, violating the strict orthogonality requirement of decision analysis, and with the inclusion of multiple dimensions for psycho-social health, the total number of items is too large for simple econometric modelling of health state utilities.

The approach adopted by the AQoL-8D to overcome this dilemma is new and is outlined in the section below which also describes the survey used to calibrate the model. Survey results and their use to model AQoL-8D utilities are outlined in the 'Results' section and discussed in the 'Discussion' section.

Modelling and data

Modelling

Scores for each of the 8 dimensions of the AQoL-8D were separately modelled and subsequently combined to obtain AQoL-8D utilities. Both the modelling of the dimensions and their combination were carried out in two stages which therefore resulted in a four-stage procedure. Stage 1 used the multiplicative model recommended by DA theory to combine the 35 items into 8 dimension scores. However, each dimension contained multiple, correlated, items which, as discussed above, would result in the double counting of the content which is common to the correlated items. Consequently, in Stage 2, an 'econometric correction' was carried out as described below, to eliminate the effects of double counting. Stages 3 and 4 replicated the methods of Stages 1 and 2 to combine dimension scores into a single utility. Stage 3 therefore consisted of the multiplicative combination of Stage 2 dimensions, and Stage 4 was a final econometric correction. Scores obtained for the

dimensions are referred to here as 'values' to indicate that results for different dimensions are not comparable and to distinguish them from 'utilities' which measure preferences on a 0.00–1.00 (death-best health) scale.

Multiplicative modelling is similar to the procedure described in Eq. 1.

$$V = \prod_{i=1}^{n} V_i \tag{1}$$

where V_i are the values of the items (dimensions) to be combined and V is the multiplicative score. The actual model is more flexible. It is calculated using disvalues, DV, rather than values, and these are adjusted for the relative importance of each of the items (dimensions) using an importance weight. This results in Eq. 2 in which $DV(x_i)$ is the disvalue of item (dimension) i, w_i is the corresponding item (dimension) importance weight 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 Eq. 3

$$DV = \frac{1}{k} \left[\prod_{i=1}^{n} [1 + kw_i DV(x_i)] - 1 \right] \tag{2}$$

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

The relationship between value and disvalue is given by Eq. 4.

$$V = 1 - DV \tag{4}$$

For both the Stage 2 and Stage 4 econometric correction, a number of multi-attribute (MA) health states were constructed, independently evaluated and regressed upon the multiplicative score and its component items (dimensions). Experimentation with different functional forms resulted in the adoption of linear models in Stage 2 and the exponential model, Eq. 5, in Stage 4. The procedures resulted in 8 dimension specific algorithms and an overall econometric algorithm for the AQoL-8D.

$$U = V^x$$
$$x = a_0 + \sum_i b_i I_i + \sum_i \sum_j c_{ij} I_i I_j \tag{5}$$

where U is the independently assessed utility of an MA health state; V is multiplicative value; $x\,a_0 b_i c_{ij}$ parameters; I_i dimension score for dimension i; $I_i I_j$ score for dimension I_i times score for dimension I_j.

Data and analysis

The relationship between the data collected and the stages of analysis is shown in Fig. 2. Commencing on the left of
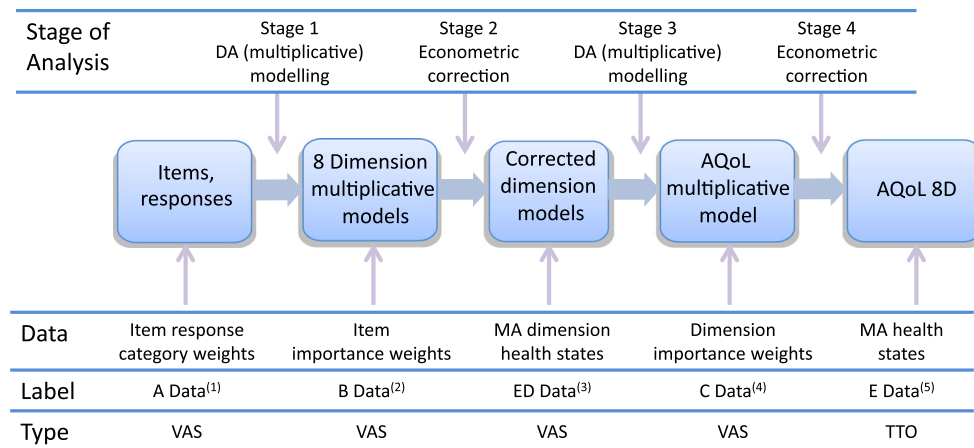
Fig. 2 Summary of data and analysis for the scaling of AQoL-8D.
[1] Within item response on an item best-item worst scale (questionnaire).
[2] Item worst response on a dimension best-dimension worst scale (questionnaire). [3] Within dimension MA health states (interview).

[4] Dimension worst response on AQoL-8D best/death/worse than death scales (interview). [5] Multi-dimensional health state (interview) as described in Iezzi et al. [21]

this figure, item responses for each of the 35 items were assessed on a 0–1 scale (0 = item worst; 1.00 = item best). Item response category weights were set equal to the average assessment ('A data'). Following the recommendation of DA theory, item importance weights ('B data') were obtained from item worst scores measured on a dimension best–worst (1.00–0.00) scale. Multi-attribute health states were created from the items of each dimension ('ED data'). These were used as the dependent variable in the econometric adjustment of dimension scores. Dimension weights for use in the multiplicative modelling were obtained from dimension worst health states ('C data') measured on an AQoL-8D best (1.00)−death (0.00)−worse than death (negative) scale. Multi-attribute health states spanning all combinations of dimensions were created for the Stage 4 adjustment ('E data').

The first four sets of data—A, B, ED and C data—were assessed using a VAS. E data were obtained from TTO valuations. This implies that Stage 4 adjusted for both the effects of the non-orthogonality of dimensions (and the possible double counting of disutility) and also converted the units of analysis from VAS to TTO. VAS valuations were obtained from mail questionnaires administered prior to a subsequent interview which edited data and separately administered the TTO questions. Protocols, props and a model interview are reproduced in Iezzi and Richardson [21].

To obtain ED and E data for the econometric corrections in Stage 2 and Stage 4, MA health states were constructed manually to ensure simplicity and plausibility. (Those with unbearable pain cannot, plausibly, be independent and happy.) Additional criteria were that MA states include all combination of items and that each dimension included mild, moderate and severe health states. All of the Stage 2

and 3 MA states are reported in Richardson et al. [22], along with the frequency distributions of dimension scores for the MA health states. These indicate that the full range of health states was employed in the analysis.

Multiplicative dimension scores were obtained from Eq. 2 using 'A data' to measure $DV(x_i)$. Parameters w_i and k were solved by inserting 'B data' in Eq. 3 (Stage 1). Final dimension scores were obtained from OLS linear regression of ED data upon each dimension multiplicative score, V_i, and the dimension items (Stage 2). A multiplicative AQoL score, V, was obtained by inserting these corrected dimension scores and dimension weights—C data—in Eqs. 2 and 3 (Stage 3). The final AQoL-8D algorithm was estimated in Stage 4 from Eq. 5 using the latter, multiplicative score, and each of the dimension scores as independent variables, and the multi-attribute (E data) as the dependent variable.

Various combinations of independent variables were tested in Stage 4. The final algorithm was selected on the basis of the usual goodness-of-fit statistics (R^2, F, RMSE) but also using the 'internal validity' of the models; that is, their ability to provide an unbiased estimate of the TTO scores from which the model was derived. This was tested using the regression: TTO $= a + b$ AQoL$_i$; where AQoL$_i$ is the predicted utility, U, from Eq. 5. With no bias, $a = 0$ and $b = 1.00$.

Main survey

The survey and data collection are outlined in Richardson et al. [22]. It had two components: (a) a sample of the Australian population and (b) a sample of patients with a clinically diagnosed mental health disorder. The target sample was

based on the minimum sampling error of 5 %, which results in a target of 400 cases in each strata [23]. The population sample was drawn from a computer readable phone directory, using a stratified, clustered two-stage design, similar to Hawthorne et al.'s [15] 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 and contacted by letter and subsequently by telephone.

People with mental disorders ('neuroses', depression and psychosis) were accessed via Melbourne's St Vincent's Hospital Mental Health Service, The Melbourne Clinic (a private hospital) and centres for post-traumatic stress disorder. Treatment providers were approached to assist with recruitment, to obtain informed consent and to ensure that people were well enough to participate in the study. Interviews were conducted with the public respondents at the Centre for Health Economics at Monash University. Patients were interviewed at their treatment service site. The protocol was approved by the Monash University Human Research Ethics Committee (MUHREC Approval No: 2004000750).

Results

Survey respondents are described in Table 1. In total, 670 individuals participated: 323 patients and 347 members of the public. The survey sought to obtain a representative sample of the public between the ages of 18 and 65. Comparison of the final columns indicates that this was achieved in the public sample but also amongst patients despite less scope for targeting patients according to their demographic characteristics. In contrast to the 60 % of the Australian population living in the lowest three SEIFA (SES) groups, only 10.8 and 26.1 % of the public and patient samples were from these postcodes. However, subsequent analyses failed to detect differences in responses between individuals from the top two and bottom three SEIFA groups.

Table 2 summarises the types and number of data collected. From the final column, average item category weights were obtained for the 176 response categories of the 35 items for both the public and patients; average importance weights were obtained for each of the 35 items and 8 dimensions for both groups. A total of 174 dimension health states and 370 AQoL-8D health states were constructed and evaluated for the econometric corrections.

Table 1 Public and patient respondents to main survey by age and gender

Age	Public			Patient			Total	% (aged 18–65)		
	Male	Female	Total	Male	Female	Total		Public	Patient	Aust[a]
18–24	22	31	53	22	23	45	98	15.8	14.2	16.3
25–34	38	36	74	35	30	65	237	21.9	20.5	22.7
35–44	32	35	67	35	43	78	145	20.0	24.6	22.1
45–54	26	43	69	41	29	70	139	20.7	22.1	21.2
55–64	30	42	72	44	15	59	131	21.6	18.6	17.7
18–65	148	187	335	177	140	317	317	100	100	100
65+	7	5	12	3	3	6	18	3.5	1.9	17.5
Total n	155	192	347	180	143	323	670	–	–	–

[a] ABS [28]

Table 2 Summary of data collected by stage of analyses

Stage	Type of data	Type	Number of data
1	Item response category weights	VAS	$\sum_{i=1}^{35}$ no. of response categories i : total = 176
1	Item importance weights	VAS	1 per item: total = 35
2	MA Health state values: dimensions[a]	VAS	Average of 22/dimension: total = 174
3	Dimensions importance weights	VAS	1 per dimension: total = 8
4	MA health state utilities: AQoL-8D[a]	TTO	Total = 370

Source: Richardson et al. [22, Table 2]

All data are average values

[a] Valuation of the constructed MA health states

Data analysis

Initial data analysis and editing are described in Richardson et al. [22]. With few exceptions, all item response categories were used by respondents and resulting frequency distributions indicated the absence of content 'gaps'. Public and patient VAS ratings of item responses (A data) were very similar. 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 patients), it was the public, not patients, who gave the higher rating. These differences occurred in the mental health, self-worth and happiness dimensions. There was, similarly, little difference in the public and patient valuation of item worst health states (B data). These differed with 5 % significance for only 4 of the 35 items. Consequently, subsequent modelling used total (public plus patient) data.

Dimension models

Item worst scores (B data) and their use in the calculation of item weights are reported in Supplementary Table S.1. Inserting the weights for each dimension in Eq. 2, in conjunction with the dimension scaling constant calculated from Eq. 3, creates the (Stage 1) multiplicative dimension formulae reported in Box 1.

The 174 'within dimension' multi-attribute health states created for the econometric correction of dimensions resulted in 2,787 individual observations (ED data), an average of 348 per dimension. The frequency distributions of these data spanned the full range of scores for each dimension, i.e. from 1.00 to 0.00 (dimension best to dimension worst). In Stage 2, these data were regressed upon the predicted multiplicative score and the average item scores for each dimension. Various models were tested, but linear regressions proved to be the most satisfactory. They are reported in Box 2.

Box 1 Multiplicative equations for dimensions (Stage 1)

General formula	$\mathrm{MULT}_d = \frac{1}{k}\left[1 - \prod_{i=1}^{n}(1 - kw_i\mathrm{DV}_i); k_d > 0\right]$
Ind living	$\mathrm{MULT\,(IL)} = 1.02[1 - (1 - 0.54\mathrm{dv}_1)(1 - 0.59\mathrm{dv}_2)(1 - 0.87\mathrm{dv}_3)(1 - 0.81\mathrm{dv}_4)]$
Pain	$\mathrm{MULT\,(PAIN)} = 1.04[1 - (1 - 0.69\mathrm{dv}_5)(1 - 0.67\mathrm{dv}_6)(1 - 0.68\mathrm{dv}_7)]$
Senses	$\mathrm{MULT\,(SEN)} = 1.04[1 - (1 - 0.63\mathrm{dv}_8)(1 - 0.61\mathrm{dv}_9)(1 - 0.69\mathrm{dv}_{10})]$
Mental health	$\mathrm{MULT\,(MH)} = 1.00\begin{bmatrix}1 - (1 - 0.74\mathrm{dv}_{11})(1 - 0.63\mathrm{dv}_{12})(1 - 0.70\mathrm{dv}_{13})(1 - 0.86\mathrm{dv}_{14}) \\ (1 - 0.74\mathrm{dv}_{15})(1 - 0.66\mathrm{dv}_{16})(1 - 0.69\mathrm{dv}_{17})(1 - 0.76\mathrm{dv}_{18})\end{bmatrix}$
Happiness	$\mathrm{MULT\,(HaP)} = 1.01[1 - (1 - 0.66\mathrm{dv}_{19})(1 - 0.60\mathrm{dv}_{20})(1 - 0.77\mathrm{dv}_{21})(1 - 0.77\mathrm{dv}_{22})]$
Coping	$\mathrm{MULT\,(COP)} = 1.03[1 - (1 - 0.60\mathrm{dv}_{23})(1 - 0.69\mathrm{dv}_{24})(1 - 0.78\mathrm{dv}_{25})]$
Relationships	$\mathrm{MULT\,(REL)} = 1.00\begin{bmatrix}1 - (1 - 0.67\mathrm{dv}_{26})(1 - 0.69\mathrm{dv}_{27})(1 - 0.69\mathrm{dv}_{28})(1 - 0.74\mathrm{dv}_{29}) \\ (1 - 0.65\mathrm{dv}_{30})(1 - 0.67\mathrm{dv}_{31})(1 - 0.66\mathrm{dv}_{32})\end{bmatrix}$
Self-worth	$\mathrm{MULT\,(SW)} = 1.02[1 - (1 - 0.69\mathrm{dv}_{33})(1 - 0.74\mathrm{dv}_{34})(1 - 0.70\mathrm{dv}_{35})]$

MULT multiplicative model score

k, w_i DV are defined in Eqs. (1) and (2), dv_i disvalue for item i where i is the order of the item in the final column of Fig. 1. *IL* independent living, *PAIN* pain, *SEN* senses, *MH* mental health, *HAP* happiness, *COP* coping, *REL* relationships, *SW* self-worth

---

Box 2 Econometric correction to multiplicative scores (Stage 2)

Ind living	$\mathrm{DV\,(IL)} = 0.123\mathrm{dv}_4 + 0.525\mathrm{DMULT\,(IL)}; R^2 = 0.85, n = 204$
Pain	$\mathrm{DV\,(PAIN)} = 0.205\mathrm{dv}_5 + 0.277\mathrm{dv}_6 + 0.205\mathrm{dv}_7 + 0.205\mathrm{MULT\,(PAIN)}; R^2 = 0.79, n = 396$
Senses	$\mathrm{DV\,(SEN)} = 0.084\mathrm{dv}_8 + 0.113\mathrm{dv}_{10} + 0.582\mathrm{MULT\,(SEN)}; R^2 = 0.87, n = 306$
Mental health	$\mathrm{DV\,(MH)} = 0.103\mathrm{dv}_{11} + 0.092\mathrm{dv}_{13} + 0.169\mathrm{dv}_{14} + 0.443\mathrm{MULT\,(MH)}; R^2 = 0.90, n = 204$
Happiness	$\mathrm{DV\,(HAP)} = 0.176\mathrm{dv}_{19} + 0.175\mathrm{dv}_{20} + 0.198\mathrm{dv}_{21} + 0.106\mathrm{dv}_{22} + 0.137\mathrm{MULT\,(HAP)}; R^2 = 0.88, n = 204$
Coping	$\mathrm{DV\,(COP)} = 0.204\mathrm{dv}_{23} + 0.090\mathrm{dv}_{24} + 0.198\mathrm{dv}_{25} + 0.232\mathrm{MULT\,(COP)}; R^2 = 0.86, n = 192$
Relationships	$\mathrm{DV\,(REL)} = 0.176\mathrm{dv}_{26} - 0.103\mathrm{dv}_{28} + 0.107\mathrm{dv}_{29} - 0.079\mathrm{dv}_{31}$ $\qquad - 0.109\mathrm{dv}_{32} + 0.533\mathrm{MULT\,(REL)}; R^2 = 0.77, n = 186$
Self-worth	$\mathrm{DV(SW)} = 0.184\mathrm{dv}_{33} + 0.115\mathrm{dv}_{34} + 0.258\mathrm{dv}_{35} + 0.190\mathrm{MULT\,(SW)}; R^2 = 0.78, n = 168$

See Box 1 for notation

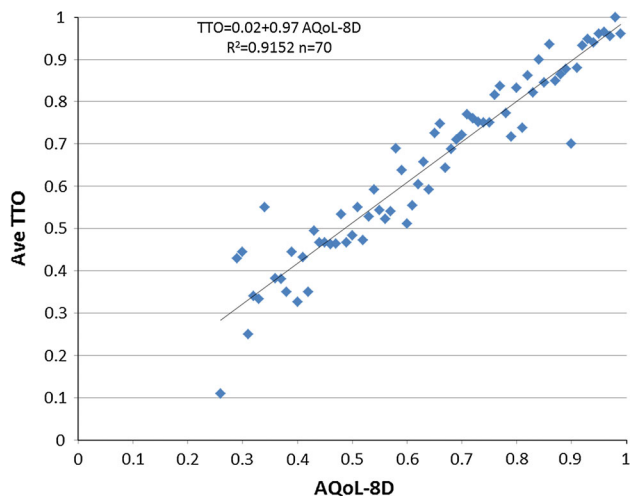| Box 3 AQoL-8D and super dimension (SD) multiplicative scores (Stage 3)	
SD (Physical)	$SD(PHYSICAL) = 1 - (1 - 0.65DV(IL))(1 - 0.72DV(PAIN))(1 - 0.69DV(SEN))$
SD (Mental)	$SD(MENTAL) = 1 - (1 - 0.74DV(MH))(1 - 0.71DV(HAP))(1 - 0.64DV(COP))$
	$(1 - 0.72DV(REL))(1 - 0.66DV(SW))$
AQoL	$MULTAQoL = 0.999[1 - (1 - 0.65DV(IL))(1 - 0.72DV(PAIN))(1 - 0.69DV(SEN))(1 - 0.74DV(MH))$
	$(1 - 0.71DV(HAP))(1 - 0.64DV(COP))(1 - 0.72DV(REL))(1 - 0.66DV(SW))]$

Fig. 3 Average TTO versus (predicted) AQoL-8D

AQoL-8D model

Dimension worst health states (C data) are reported in Supplementary Table S.2 for both the public and patients. Average scores differ significantly only for independent living (5 % significance) and coping (1 % significance). In both cases, the public score is higher, not lower, than the patient score. In Stage 3, the average of both scores was used as the dimension weight, w_i, in Eqs. 2 and 3. The resulting multiplicative models for the two super-dimensions and for the AQoL-8D are reported in Box 3.

For the final econometric correction (Stage 4), 3,178 TTO observations were obtained from the 370 MA health states. These successfully spanned all dimensions and combinations of dimensions (E data). Mean TTO scores were used in the final econometric correction, Eq. 5. As noted, since previous valuations employed a VAS, the Stage 4 'correction' took account of both double counting of disutilities and differences between VAS and TTO. Results are presented in Table 3.

The test of internal validity for the five models is reported in the bottom block of Table 3. 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. Using this criterion model,

4 marginally outperforms model 5. However, the quantitatively large but marginally significant negative coefficients in this model result in a number of anomalous scores (in excess of 1.00). Consequently, model 5 was adopted for the final algorithm. The OLS test regression was separately estimated for the public and patient samples. Results reported as footnote to Table 2 are very similar and suggest that the preferences of the two groups do not diverge substantially.

Predicted AQoL-8D utilities from the final algorithm are plotted against average TTO scores in Fig. 3. To simplify the figure, data were aggregated to create a single observation for each percentile of the AQoL-8D utilities. The vertical axis therefore measures the average TTO for AQoL-8D utilities in the range (1.00–0.995), (0.985 to <0.995), (0.975 to <0.985), etc. The increased aggregation did not alter the OLS regression reported for model 5 in Table 3, but increased the R^2 from 0.73 to 0.92.

Discussion

AQoL-8D methods were innovative in several respects. Its descriptive system was derived using psychometric methods [9]. As described here, its modelling employed a four-stage methodology incorporating econometric corrections. This was necessary as the descriptive system was too large to model with the single-stage econometric analyses used in some other MAU instruments [1]. But the use of DA methods to combine items necessarily violated orthogonality requirements [12]. The econometric correction reconciled the need for a larger descriptive system with the problems this created for modelling utility.

The methods were necessarily based upon a number of limiting assumptions. Utility was measured using the TTO technique. Coefficients at all stages of the analysis were derived from average responses. This implies that the instrument will not properly represent the preferences of particular individuals. The average utilities reflected the preferences of an Australian sample which may differ from preferences in other countries. The sample did not include many respondents aged above 65 whose valuation of ill health may differ from other age groups. However, these problems are general. Even a perfectly representative cross

Table 3 Regression results for Eq. 5 (Stage 4)

Independent variables[a]	M1	M2	M3	M4	M5
Regression: TTO on MULT AQoL[b] and its dimensions ($n = 370$)					
Constant	0.145	0.25	0.33	0.29	0.29
IL (D1)			ns	ns	
PAIN (D7)			−0.04	−0.21*	−0.04
SEN (D8)			−0.06		
MH (D3)			−0.06	ns	0.14
HAP D2			−0.04**	ns	0.29
COP (D4)			−0.06		−0.08
REL (D5)			0.05*	0.41*	0.32
SW (D6)			−0.07	−0.35*	−
SUPER (phys)		−0.08			
SUPER (MH)		−0.18			
IL × REL				−0.24***	
IL × SW				0.20***	
IL × PAIN				0.10***	
IL × HAP × REL				−0.31	−0.54
MH × PAIN				0.17**	
MH × SEN				−0.10**	−0.15
COP × SEN				−0.05	
MH × SW					−0.24
R^2 (adj.)	0.84	0.87	0.88	0.89	0.90
F	1,944	800	299	166	366
ROOT MSE	0.24	0.21	0.20	0.20	0.103
Regression: TTO on AQoL-8D[c] ($n = 370$)					
a	−0.21	0.005 ns	0.05 ns	0.01 ns	0.02 ns
b	1.41	1.03	1.03	1.00	0.97
R^2	0.62	0.66	0.68	0.72	0.73
F	606	717	802	928	864
ROOT MSE	0.12	0.11	0.11	0.10	0.10

[a] 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)$^{constant + \Sigma bi\ variable\ i}$. All coefficients are significant at 0.1 % level unless shown; * significant at 1 % level; ** significant at 5 % level; *** significant at 10 % level

[b] MULT AQoL = multiplicative combination of dimensions (see Box 3)

[c] TTO are the average from public and patient assessments for each health state. Separating the two groups results in public: TTO = 0.06 + 0.96 TTO; $R^2 = 0.88$ ($n = 370$). Patient: TTO = 0.01 + 0.95 TTO; $R^2 = 0.86$ ($n = 370$)

section of a national population may not capture preferences of a particular socio-demographic or disease subgroup in the same country. It would, nevertheless, be desirable for future research to test the validity of AQoL-8D for use with other populations and particularly the elderly who represent an increasingly important section of the population.

Despite these limitations, AQoL-8D is unique in having employed both a representative cross section of the general public and patients with clinically defined mental health problems for constructing both the descriptive system and the utility algorithm. The objective was to achieve a 'mixed perspective' as a compromise between the perceived differences between public and patient perspectives. Its success is qualified by the non-random selection of the patients of participating psychiatrists and sampling according to predetermined criteria was, therefore, not feasible. However, there were no grounds for believing that these patients were otherwise atypical. In the event, very few differences emerged between the valuations from the two groups. Contrary to the common belief that adaptation results in higher patient scores, when differences were observed, it was the public, not the patient group which had the higher score. This suggests that the common view of adaptation may apply chiefly to physical disability where the public can envisage the consequences of the disability.

A lesser ability to envisage and empathise with psychological health states would also explain the higher valuations placed upon them by the public when differences occurred.

A consequence of the enlarged psycho-social component of the descriptive system is that the AQoL-8D is sensitive to changes in the dimensions which drive adaptation and which, when absent, differentiate patient and public valuations. As information relating to self-esteem, coping, happiness, etc. is included in the health state description of the AQoL-8D, it is less likely that patient and public valuations will diverge because of a differing assessment of these elements. The similar evaluations by the two groups in the AQoL-8D surveys suggest that differences between public and patient assessments are, in part, attributable to the absence of relevant information in the health state description and that when appropriate psycho-social information is provided, the valuations of items, dimensions and health states will be relatively similar. However, by increasing the relative importance of psycho-social dimensions, the AQoL-8D decreased the relative importance of physical dimensions. This is possibly an inescapable consequence of expanding the scope of measurement.

By September 2013, the AQoL register (www.aqol.com.au) had recorded 99 research teams which had indicated an intention to use AQoL-8D, 59 of which provided details as published on the AQoL website [24]. The instrument has been included in an 8,022 person multi-instrument online comparative survey [25]. This found that, despite its greater length, there was an average completion time of 5.45 min. The instrument demonstrated a high degree of sensitivity to all the health states but particularly to health states with poor psycho-social health. The survey was used, inter alia, to test AQoL-8D reliability and validity [26]. Reliability (Cronbach alpha) coefficients from the analysis are reported in Supplementary Table S.3, and the frequency distribution of AQoL scores for the healthy public shown as Supplementary Figure S.1. Population norms for the instrument have been published [27].

Conclusions

The objective of the AQoL-8D project was the creation of an MAU instrument with increased sensitivity to psychosocial health states. The construction of a descriptive system with content validity in these dimensions resulted in an instrument which is large by the standards of extant MAU instruments and too large for the use of past methods to create a flexible utility algorithm. The present article has described the methods adopted to meet this challenge.

The resulting algorithm predicts utilities which are in the same range as extant instruments. The predicted utilities are an unbiased estimate of the TTO valuations of the health states. The instrument has high internal and test–retest reliability and good external validity. It exhibits high sensitivity to the health dimensions which it was designed to measure. These results imply that the AQoL-8D is a useful addition to the extant instruments and particularly for analyses requiring measurement where psycho-social health is affected.

References

1. Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2007). *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press.
2. Richardson, J., McKie, J., & Bariola, E. (2014). Multi attribute utility instruments and their use. In A. J. Culyer (Ed.), *Encyclopedia of Health Economics*. San Diego: Elsevier Science.
3. Moock, J., & Kohlmann, T. (2008). Comparing preference-based quality-of-life measures: Results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Quality of Life Research, 17*, 485–495.
4. Fryback, D. G., Palta, M., Cherepanov, D., Bolt, D., & Kim, J. (2010). Comparison of 5 health related quality of life indexes using item response theory analysis. *Medical Decision Making, 30*, 5–15.
5. Dolan, P., Lee, H., & Peasgood, T. (2012). Losing sight of the wood for the trees: Some issues in describing and valuing health, and another possible approach. *Pharmacoeconomics, 30*, 1035–1049.
6. Bowling, A. (1995). What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Social Science and Medicine, 41*, 1447–1462.
7. Helliwell, J., Layard, R., & Sachs, J. (2012). *World happiness report*. New York: Earth Institute, Columbia University. http://www.earthinstitute.columbia.edu/articles/view/2960 (Accessed April 30, 2012).
8. Richardson, J., Day, N. A., Peacock, S., & Iezzi, A. (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*, 62–88.
9. Richardson, J., Elsworth, G., Iezzi, A., Khan, M. A., Mihalopoulos, C., Schweitzer, I., et al. (2011). *Increasing the sensitivity of the AQoL inventory for evaluation of interventions affecting mental health, research paper 61*. Melbourne: Centre for Health Economics, Monash University.
10. McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling: A Multidisciplinary Journal, 12*, 163–172.
11. Sintonen, H., & Pekurinen, M. (1989). A generic 15 dimensional measure of health-related quality of life (15D). *Journal of Social Medicine, 26*, 85–96.
12. Torrance, G., Boyle, M., & Horwood, S. (1982). Application of multi-attribute utility theory to measure social preferences for health status. *Operations Research, 30*, 1043–1069.
13. Feeny, D., Furlong, W., & Torrance, G. (1996). *Health Utilities Index Mark 2 and Mark 3 (HUI 2/3) 15 item questionnaire for self-administered, self assessed usual health status*. Hamilton,

ON: Centre for Health Economics and Policy Analysis, McMaster University.

14. Feeny, D., Furlong, W., Torrance, G., Goldsmith, C., Zhu, Z., DePauw, S., et al. (2002). Multi attribute and single attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care, 40*, 113–128.

15. Hawthorne, G., Richardson, J., & Osborne, R. (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.

16. Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

17. Dolan, P. (1997). Modeling valuations for EuroQol health states. *Medical Care, 35*, 1095–1108.

18. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*, 271–292.

19. Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

20. Streiner, D., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use*. Oxford: Oxford University Press.

21. Iezzi, A., & Richardson, J. (2009). *Measuring quality of life at the CHE: Description of instruments, interview props and their administration, research paper 41*. Melbourne: Centre for Health Economics, Monash University.

22. Richardson, J., Iezzi, A., Khan, M. A., Sinha, K., Mihalopoulos, C., Herrman, H., 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.

23. Hoinville, G., Jowell, R., Airey, C., Brook, J., Courtenay, G., Hedges, B., et al. (1977). *Survey research practice*. London: Heinemann Educational Books.

24. Richardson, J., & AQoL Group. (2013). *Assessment of Quality of Life (AQoL) users* (Vol. 2013). http://www.aqol.com.au/index.php/aqoluserinfo. (Melbourne: http://www.aqol.com.au/).

25. Richardson, J., Khan, M. A., Iezzi, A., & Maxwell, A. (2012). *Cross-national comparison of twelve quality of life instruments*. Research Papers 78, 80–83, 85. MIC Report: 2: Australia; 3: UK; 4: USA; 5: Canada; 6: Norway; 7: Germany. Melbourne: Centre for Health Economics, Monash University. http://www.buseco.monash.edu.au/centres/che/che-publications.html (Accessed January 25, 2013).

26. Richardson, J., Iezzi, A., Khan, M. A., & Maxwell, A. (2013). Validity and reliability of the Assessment of Quality of Life (AQoL-8D) multi attribute utility instrument. *The Patient: Patient-Centered Outcomes Research*. doi:10.1007/s40271-013-0036-x.

27. Richardson, J., Khan, M. A., Chen, G., Iezzi, A., & Maxwell, A. (2012). *Population norms and Australian profile using the Assessment of Quality of Life (AQoL) 8D utility instrument, research paper 72*. Melbourne: Centre for Health Economics, Monash University.

28. ABS. (2013). *Australian demographic statistics, population by age and sex, Cat 3201.0*. Canberra: Australian Bureau of Statistics. http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3201.0Jun%202010?OpenDocument (Accessed August 12, 2013).