

Validity and Reliability of the Assessment of Quality of Life (AQoL)-8D Multi-Attribute Utility Instrument

Jeff Richardson · Angelo Iezzi · Munir A. Khan ·
Aimee Maxwell

Published online: 23 November 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract

Objective The purpose of this paper was to report tests of the validity and reliability of a new instrument, the Assessment of Quality of Life (AQoL)-8D, which was constructed to improve the evaluation of health services that have an impact upon the psychosocial aspects of the quality of life.

Methods Australian and US data from a large multi-instrument comparison survey were used to conduct tests of convergent, predictive and content validity using as comparators five other multi-attribute utility (MAU) instruments—the EQ-5D, SF-6D, Health Utilities Index (HUI) 3, 15D and the Quality of Well-Being (QWB)—as well as four non-utility instruments—the SF-36 and three measures of subjective well-being (SWB). A separate three part Australian survey was used to assess test–retest reliability.

Results Results indicate that AQoL-8D correlates more highly with both the SWB instruments and the psychosocial dimensions of the SF-36, and that it is similar to the other MAU instruments in terms of its convergent and predictive validity. The second Australian survey demonstrated high test–retest reliability.

Conclusions The results indicate that the AQoL-8D is a reliable and valid instrument which offers an alternative to the MAU instruments presently used in economic evaluation studies, and one which is particularly suitable when psychosocial elements of health are of importance.

Key Points for Decision Makers

- Assessment of Quality of Life (AQoL)-8D provides a valid and reliable alternative to existing multi-attribute utility instruments.
- By comparison with existing instruments, AQoL-8D has:
 - greater coverage of mental and social dimensions of health;
 - similar results with respect to convergent and predictive validity;
 - a higher correlation with subjective well-being;
 - a higher correlation with the SF-36 mental health dimensions;
 - a lower correlation with the SF-36 physical health dimensions.
- AQoL-8D ‘levels the playing field’ when services are compared that primarily affect psychosocial health.

1 Introduction

Economic evaluations of health programs commonly use quality-adjusted life-years (QALYs) as the unit of outcome, where QALYs are calculated as the product of time in a health state and the utility of the health state measured on a 0–1 scale. Utility has increasingly been measured with a multi-attribute utility (MAU) instrument—a generic, health-related questionnaire about the quality of life—and an accompanying formula or weights for converting question responses into utility scores.

Utilities measured by these instruments differ significantly. In the only two published studies to date that compare five MAU instruments, the proportion of the variance in the utilities of one instrument explained by another averages 56 % and 47 %, respectively [1, 2]. In a review of empirical studies from 2005 to 2010, Richardson

J. Richardson (✉) · A. Iezzi · M. A. Khan · A. Maxwell
Centre for Health Economics, Monash University, Clayton,
Melbourne, VIC 3800, Australia
e-mail: jeffrey.richardson@monash.edu

et al. [3] identified 392 pairwise comparisons of MAU instruments. Authors commonly concluded that the utilities derived from different instruments were not equivalent, and that comparisons between them warrant caution.

The differences and consequences of them are widely recognized. Commenting on the main instruments, Drummond et al. [4], for example, note that they are “far from identical It is not surprising that comparative studies show the same patient groups can score quite differently depending upon the instrument used” (pp 160–170). Similarly, Brazier et al. [5] report that “generic measures of health have been found to be inappropriate or insensitive for many medical conditions ... no instrument is able to cover all health dimensions” (pp 60–63).

The Assessment of Quality of Life (AQoL)-8D instrument was developed in response to the omissions from the descriptive systems employed by existing instruments. In particular, it sought to increase measurement sensitivity to psychosocial elements of health. This unique emphasis raises the question of the reliability and validity of the new instrument: whether the inclusion of significant new content in its descriptive system compromises its performance as judged by these criteria.

The concepts of ‘validity’ and ‘reliability’ are discussed by a large number of authors [4, 6–10]. As summarized by Streiner and Norman [6] “validation is a process of hypothesis testing ... validating a scale is really a process whereby we determine the degree of confidence we can place on inferences we make about people based upon their score from that scale” (p 174). While tests are variously classified, the present paper presents three forms of testing, namely convergent, predictive and content validity. The first of these—convergent validity—is “how closely the new scale is related to other variables and other measures of the same construct to which it should be related” (p 183) [6].

In the case of an MAU instrument, the construct is ‘utility’, the strength of preference for (in the present case) a health state. Three types of comparator instruments were available, as discussed below, to test the AQoL-8D’s convergent validity. First, the other MAU instruments in the survey all purport to measure utility, although, as noted, measurement from some or all of these instruments is imperfect (*cf* the low correlation between them). Secondly, the preference for health states should correlate with scores obtained from the SF-36, the most widely used non-utility measure of health-related quality of life (HRQoL). Thirdly, people have a preference for happiness or, more generally, subjective well-being (SWB): if they maximised happiness there would be a perfect correlation with utility. It was therefore hypothesized that AQoL-8D scores would correlate with the scores obtained from the three SWB instruments in the survey.

The second concept—predictive validity—is closely related to convergent validity and defined by an

instrument’s “ability to predict observable ‘criterion’ behaviours ... relating it (i.e. a test score) with some ‘outcome’ measure external to it” (p 198) [11].

The third concept—content validity—refers to the coverage of the items in the instrument. As stated in the classic article by Loevinger [7], the pool of items “should be chosen so as to sample all possible contents which might comprise the putative trait according to all known alternative theories of the trait” (quoted from Streiner and Norman (p 22) [6]. The instruments available for the present study provided several different but related measures of the major dimensions of the HRQoL. These were employed in the analysis of content validity.

The reliability of a scale is a measure of “the amount of error, both random and systematic, inherent in any measurement” (p 126) [6]. Two measures are usually included in the assessment of scales. The first is a measure of the homogeneity of each of the items: whether each is “tapping different aspects of the same attribute” (p 68) [6] and is commonly measured by the Cronbach alpha. This indicates whether the same score would be obtained from two split halves of the instrument using every possible combination of ways of splitting the instrument. The second measure—the test–retest reliability—is a measure of the extent to which the same score will be predicted from the same individual at a second point in time.

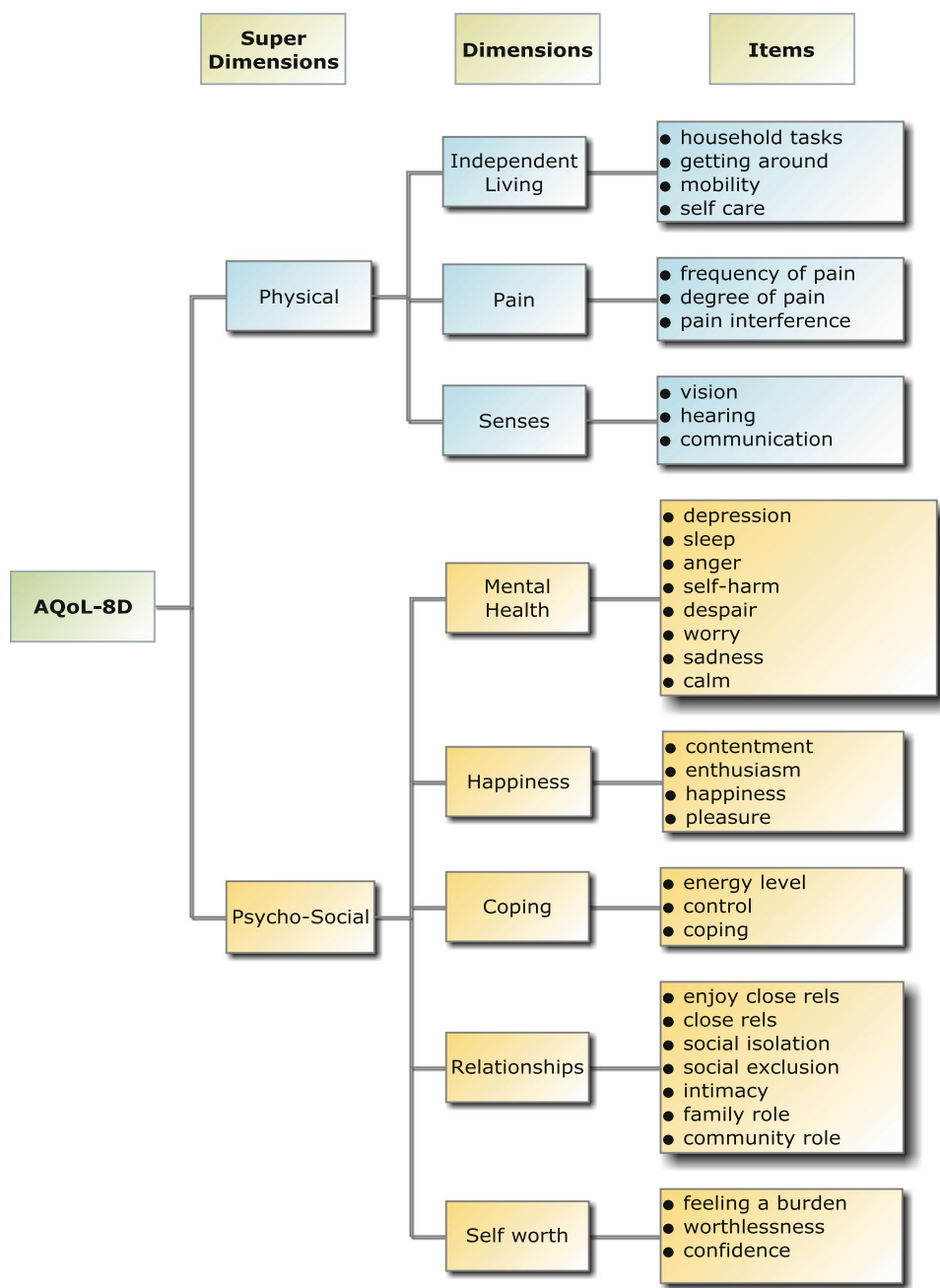
Section 2 below describes the AQoL-8D, its construction, and the comparator instruments used in the study. Section 3 describes the databases employed and the analytical methods used to test the validity and reliability of the AQoL-8D. Results are presented in Sect. 4 and discussed in Sect. 5.

2 Assessment of Quality of Life (AQoL)-8D and Comparator Instruments

2.1 AQoL-8D

The AQoL-8D is an extension of two earlier instruments, the ‘AQoL’ (or AQoL-4D) and AQoL-6D [12, 13]. To achieve an instrument with greater sensitivity to psychosocial health, both the descriptive system and scoring algorithm of the AQoL-6D were revised as detailed elsewhere [14, 15]. Initially, a list of potential items was compiled from the AQoL-6D and from extant mental health instruments. New items were constructed by the research team using results from four focus groups with mental health patients. Items were administered to a representative sample of 195 members of the public and 514 mental health patients. As recommended by McDonald [11], a combination of restrictive and unrestricted factor analyses was used to create the AQoL-8D descriptive system. The resulting instrument, shown in Fig. 1, contains

Fig. 1 AQoL-8D structure.
AQoL Assessment of Quality of Life



35 items which load onto eight dimensions. Three of these are related to a physical ‘super-dimension’ and the remaining five to a psychosocial (‘mental’) super-dimension. Utility weights for each health state described by the instrument were modelled using a two-stage procedure. First, an algorithm was obtained to produce a score for each of the eight dimensions. Secondly, the dimension scores were combined to form final AQoL-8D utilities. Data for the modelling were obtained from a survey/interview of 347 members of the public and 323 mental health patients. Respondents provided visual analogue scale (VAS) valuations of each item response, item and

dimension. Additionally, 3,178 time trade-off (TTO) assessments of 370 multi-attribute health states were obtained during interview (i.e. an average of 8.6 individual assessments per health state). Items were combined into dimensions using the multiplicative modelling recommended by decision analytic theory [16]. Dimensions were subsequently combined, also using a multiplicative model. Finally, the multiplicative score was used to predict, econometrically, the TTO health-state values. The best fitting econometric function was adopted as the AQoL-8D algorithm. This is provided on the AQoL website in both SPSS and Stata, along with a user manual.

Box 1 Dimensions of the SF-36

Physical	Psychosocial
<i>General health (GH) 5 items:</i>	<i>Mental health (MH) 5 items:</i>
◆ general health perceptions	◆ nervousness
	◆ feeling down
	◆ feeling calm/happiness
<i>Bodily pain (BP) 2 items:</i>	<i>Social functioning (SF) 2 items:</i>
◆ the degree of pain	◆ interference with normal and social activities
◆ interference with normal work due to pain	
<i>Role limit physical (RP) 4 items:</i>	<i>Role limit emotional (RE) 3 items:</i>
◆ time spent on work	◆ work time
◆ difficulty performing work	◆ work accomplished
	◆ care at work
<i>Physical functioning (PF) 10 items:</i>	<i>Vitality (VT) 4 items:</i>
◆ vigorous/moderate activities	◆ energy/tiredness
◆ lifting	
◆ climbing stairs	
◆ bending	
◆ walking	
◆ bathing	

Source: Adapted from McDowell [9]

2.2 Comparator Instruments

Validation of the AqoL-8D used individual results from nine additional instruments: five MAU instruments, three SWB instruments, and the SF-36. The multi-attribute instruments are described and contrasted with the AqoL-8D in Table 1. They vary significantly in size and content. The 35 items of the AqoL-8D define 2.4×10^{23} health states. In contrast, the Quality of Well-Being (QWB) and SF-6D define 945 and 18,000 health states, respectively. The most widely used instrument—the EQ-5D—consists of only five items. Recent revision of the number of response categories from three to five has increased the number of health states described from 243 to 3,125. Four of the five EQ-5D items relate to physical health. In contrast, three of six SF-6D and 25 of 35 AqoL-8D items relate to psychosocial health. Utilities are all measured on a scale where 1.00 represents the instrument ‘all best’ health state and 0.00 represents ‘death’. However, the scoring algorithms predict instrument ‘all worst’ utilities which vary from 0.32 for the QWB to -0.59 for the EQ-5D.

The SF-36 has been cited in more than 14,000 peer review articles and referenced in over 1,900 random control trials, and is the most widely used measure of functional health and quality of life in the world [17]. While it does not have utility weights, its eight dimensions have been shown to represent valid and reliable subscales [9].

These were therefore used to test the content of the AqoL-8D. The eight dimensions are described in Box 1.

The remaining three instruments all seek to measure SWB. However, like the MAU instruments, which all purport to measure utility, their descriptive systems—the questions asked—differ. As discussed in the OECD guidelines on measuring SWB [18], the scales are dominated by the concept of satisfaction. Three of five items in the Satisfaction with Life Scale (SWLS) relate to present satisfaction, and the remaining two relate to past satisfaction [19]. The eight items of the Personal Wellbeing Index (PWI) relate to eight sources of current satisfaction (health, relationships, etc.) [20]. The third instrument developed experimentally in the UK by the Office of National Statistics (ONS) in 2011, includes two current satisfaction items, and single items for happiness and anxiety [21].

3 Methods

3.1 Data

The analysis of validity drew upon results from a Multi-Instrument Comparison (MIC) study. An online survey was administered in Australia, Canada, Germany, Norway, UK and the US by a global panel company, CINT Australia Pty Ltd [22]. For reasons discussed in Sect. 5, the present paper only used results from Australia and the US. Respondents were asked to complete, inter alia, the ten instruments described above—AqoL-8D, the five other MAU and three SWB instruments, and the SF-36.

The personal and medical details recorded by the panel company were used to recruit individuals from seven major disease groups and from the ‘healthy public’, i.e. those who did not report any chronic disease and who obtained a score of at least 70 on a 100-point VAS measuring overall health. Respondents with one of the seven chronic diseases were asked to complete a relevant disease-specific questionnaire. The seven disease groups were arthritis, asthma, cancer, depression, diabetes, hearing loss and coronary heart disease (CHD).

Eight ‘edit criteria’ were employed to determine whether each individual’s answers were unreliable and should be removed from the sample. The criteria were based upon a comparison of duplicated or similar questions. Additionally, results were deleted when an individual’s (recorded) completion time was <20 min, which was judged to be the minimum time in which the 230 questions could be answered. The ‘healthy’ public were recruited to achieve a sample with demographic and educational characteristics that were broadly representative of the total population. Edit procedures, the questionnaire and its administration are described by Richardson et al. [23]. The survey was

Table 1 Description of instrument

Name	Items per dimension										Total	States	Worst health	Market share 2005–2010 ^b % total use (<i>n</i> = 1,682)
	Physical					Psychosocial								
	Mobility/physical function	Pain	Senses	Subtotal, physical	Mental	Social ^b	Vitality	Esteem/SWB	Sub total, psychosocial					
EQ-5D	3	1	–	4	1	–	–	–	–	1	5	3,125	–0.59	63.2
SF-6D	2	1	–	3	1	1	–	–	–	3	6	18,000	0.3	8.8
HUI 3	2	1	2	5	2	1	–	–	–	3	8	972,000	–0.36	14.4 ^c
15D	7	1	3	11	2	2	1	–	–	5	15	3.1 × 10 ¹⁰	0.11	6.9
QWB ^a	2	–	–	*	–	1	–	–	–	4	–	945	0.32	2.4
AQoL-8D	4	3	3	10	8	7	3	7	7	25	35	2.4 × 10 ²³	–0.04	4.3
SF-36	14	2	–	5	8	2	4	–	–	20	35	1.75 × 10 ²⁰	NA	

Source: Brazier et al. [5], Richardson et al. [3]

HUI 3 Health Utility Index 3, QWB Quality of Well-Being, AQoL-8D Assessment of Quality of Life-8D, SWB subjective well-being, NA not applicable

^a QWB includes three items and two symptom groups

^b Includes communication

^c HUI 2 and HUI 3

approved by the Monash University Human Research Ethics Committee (MUHREC), approval CF11/1758: 2011000974.

In the second, smaller survey to determine test–retest reliability, 285 (different) Australian respondents were invited to complete a baseline survey and to complete two follow-up surveys spaced a fortnight apart. At each of the three stages, the AQoL instruments were administered. Quotas were imposed to ensure that the initial sample was representative of the age, gender and educational profile of the Australian population (MUHREC approval CF11/3192: 2011001748).

3.2 Analysis

Convergent validity was tested conventionally using the Pearson correlation between the AQoL-8D and the other five MAU instruments. However, MAU scores purport to measure utility on the same numerical (0.00–1.00) scale. Consequently, scores were also compared using intra-class correlation (ICC), which tests the correspondence of absolute scores. The criterion set for the AQoL-8D was that its correlation with other MAU instruments should be at least equal to the average correlation between the other widely accepted MAU instruments.

3.3 Validation

In the absence of a gold standard, it is not possible to prove conclusively that an instrument is ‘valid’—that it measures what it purports to measure. Rather, validation is a process of hypothesis testing to increase confidence that a scale has the properties that would be expected if it were valid [10]. Tests are variously classified. The present paper presents tests of convergent, predictive and content validity.

Predictive validity was tested by the ability of AQoL-8D to predict *changes* in the utilities predicted by the other MAU instruments. To carry out this test, pairwise geometric mean squares (GMS) linear regressions were estimated between all combinations of instrument values. In the resulting equation, $MAU_i = a + b MAU_j$, the coefficient ‘*b*’ measures the ratio of the marginal change in MAU_i to the marginal change in MAU_j . Perfect prediction of change would result in $b = 1.00$. The deviation from $b = 1.00$ is a measure of the imperfection of the prediction. The relevant test was therefore that the deviation in the prediction by AQoL-8D should be no greater than the average deviation in the prediction of other instruments. GMS regressions were employed as their results do not vary with the choice of dependent and independent variables [24].

Content validity may be assessed qualitatively by determining whether an instrument includes items directly

describing the major dimension of HRQoL (face validity). The more formal approach adopted here was to determine the significance of the correlation of the final instrument with generally recognized dimensions of the HRQoL. The available data consisted of the dimension scores of the SF-36 and the three indices of SWB.

Internal consistency (reliability) was tested using the Cronbach alpha. This was estimated for each of the eight subscales and two super-dimensions using data from the MIC survey. Test–retest reliability was tested using the ICC between observations at different times using data from the second survey.

4 Results

4.1 Survey 1

In the first survey, editing of data eliminated 14.9 and 11.0 % of the Australian and US respondents, respectively, leaving usable samples of 1,430 and 1,460 respondents, respectively. Age/sex distributions for both ‘public’ and ‘patient’ samples are reported in Table 2. They are almost identical in the two countries, reflecting the use of demographic-based quotas. Unreported results found that the number of public respondents completing only high school, with a diploma or trade certificate, and completing university are almost identical in the Australian sample but skewed towards high-school completions in the US (42.4 %, 23.1 % and 34.5 % for the three US categories, respectively). Because of quoting, the numbers of respondents in each of the seven disease areas are very similar, varying from 148 to 179 per category. By comparison with the US, Australian men are overrepresented in every disease category. However, the differences are unimportant in the context of this study as representative samples are not strictly necessary for a comparison of instruments.

Table 3 reports summary statistics. The scores in the two countries are very similar. The maximum difference between mean scores is 0.03 (EQ-5D, public). Mean scores for the EQ-5D, Health Utilities Index (HUI) 3 and AQoL-8D are also very similar, particularly in the ‘public’ sample. However, the distributions of scores are dissimilar. The standard deviation around the mean varies by more than 100 % between the SF-6D/15D and the HUI 3. The EQ-5D has very significant ceiling effects, with about 40 % of respondents in both countries recording no disutility. In contrast, <10 % of public respondents recorded maximum scores on the SF-6D and AQoL-8D. In the total sample (public plus patients) only 0.5 and 1.4 % of respondents recorded scores below 0.4 on the 15D and SF-6D, whereas more than 10 % were assigned scores below 0.4 by the AQoL-8D and HUI 3, respectively.

4.2 Convergent Validity

The Pearson correlation between MAU instrument scores are reported in the top right-hand side of Table 4. The average of the correlations which included each instrument is shown in the final column of the table. It represents a summary measure of the convergence of each MAU instrument with the remaining five instruments. The results are similar in the two countries. The lowest correlation in both is between the QWB and EQ-5D (0.65 in both countries). The highest correlations are 0.82 and 0.84 between 15D and HUI 3 (Australia) and 15D and AQoL-8D (US). The average correlation with other instruments is highest in both countries for the 15D (0.79, 0.80; Australia/US), followed by AQoL-8D (0.77, 0.79; Australia/US). However, with the exception of the QWB there is little difference between the averages.

While the Pearson correlation is the conventional test of convergent validity, a more stringent test is the use of the ICC, which tests the association between absolute scores. It differs from the Pearson correlation if the line of best fit between the variables is not $Y = X$; that is, the implicit scales of the variables differ. ICC’s between MAU instruments are shown in the bottom left-hand side of Table 4. They are (necessarily) smaller than the Pearson correlations. The average ICC for the 15D drops from the highest to lowest position, reflecting the compressed range of scores it predicts. The largest average ICC in both countries is 0.69 for AQoL-8D, followed by 0.65 (0.67) for the EQ-5D in Australia (US).

4.2.1 Predictive Validity

Pairwise GMS regressions are reported for each combination of instruments for both countries in the Appendix. The country results are again almost identical. There is a maximum difference in the b coefficients between the two countries of only 7.6 % (1.83 vs. 1.70; Australia/US) in the regression of QWB on HUI 3. R^2 coefficients are higher than in the two five-instrument studies reported earlier, reflecting the wider range of observations in the first survey.

Perfect prediction of the marginal change in one MAU instrument by another implies $b = 1.00$ in the relevant pairwise regression. Table 5 reports deviation from this when deviation is measured as the larger divided by the smaller marginal change times 100. The lowest deviation is associated with QWB, AQoL-8D and EQ-5D, indicating greater predictive validity by these instruments when each is judged by the remaining instruments.

4.2.2 Content Validity

From Table 1, AQoL-8D has high face validity and particularly in the psychosocial dimensions, which include 24

Table 2 Respondents by age and gender (survey 1)

Respondents	Age group (years)	Australia			US			Total		
		Male (%)	Female (%)	<i>N</i>	Male (%)	Female (%)	<i>N</i>	Male (%)	Female (%)	<i>N</i>
Healthy public ^a	18–24	4.5	6.8	30	4.7	5.6	33	4.6	6.1	63
	25–34	8.7	9.4	48	8.1	9.7	57	8.4	9.6	105
	35–44	8.3	10.6	50	7.5	10.6	58	7.8	10.6	108
	45–54	8.3	10.2	49	9.3	10.9	65	8.9	10.6	114
	55–64	7.2	7.5	39	7.8	8.4	52	7.5	8.0	91
	65+	9.4	9.1	49	7.8	9.7	56	8.5	9.4	105
	<i>Total</i>	<i>46.4</i>	<i>53.6</i>	<i>265</i>	<i>45.2</i>	<i>54.8</i>	<i>321</i>	<i>45.7</i>	<i>54.3</i>	<i>586</i>
Patient	18–24	0.7	1.5	25	1.1	3.7	55	0.9	2.6	80
	25–34	3.0	5.0	93	2.5	6.3	100	2.7	5.6	193
	35–44	4.2	6.1	120	3.8	9.3	149	4.0	7.7	269
	45–54	8.2	11.2	227	9.4	15.6	285	8.8	13.4	512
	55–64	17.8	14.8	380	8.8	16.7	290	13.3	15.8	670
	65+	16.5	11.0	320	10.9	11.9	260	13.7	11.5	580
	<i>Total</i>	<i>50.4</i>	<i>49.6</i>	<i>1,165</i>	<i>36.4</i>	<i>63.6</i>	<i>1,139</i>	<i>43.5</i>	<i>56.5</i>	<i>2,304</i>
Total	18–24	1.4	2.4	55	1.9	4.1	88	1.7	3.3	143
	25–34	4.1	5.8	141	3.7	7.1	157	3.9	6.4	298
	35–44	5.0	6.9	170	4.6	9.6	207	4.8	8.3	377
	45–54	8.3	11.0	276	9.4	14.6	350	8.8	12.8	626
	55–64	15.8	13.5	419	8.6	14.9	342	12.1	14.2	761
	65+	15.2	10.6	369	10.2	11.4	316	12.7	11.0	685
	<i>Total</i>	<i>49.7</i>	<i>50.3</i>	<i>1,430</i>	<i>38.4</i>	<i>61.6</i>	<i>1,460</i>	<i>43.9</i>	<i>56.1</i>	<i>2,890</i>

^a Healthy = no chronic disease; visual analogue scale overall health $\geq 70/100$

Table 3 Summary statistics: survey 1

Instrument	EQ-5D		HUI 3		SF-6D		15D		QWB		AQoL-8D	
	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US
Public												
Mean ^a	0.87	0.90	0.88	0.89	0.79	0.80	0.94	0.94	0.74	0.76	0.86	0.87
SD ^a	0.12	0.10	0.15	0.15	0.11	0.12	0.06	0.07	0.14	0.14	0.18	0.13
Ceiling (%) ^a	39.60	42.00	18.50	21.50	2.00	7.20	18.10	23.40	7.20	9.00	2.60	8.10
Min	0.36	0.14	0.14	0.03	0.49	0.53	0.69	0.46	0.35	0.28	0.44	0.33
Total												
Mean ^a	0.73	0.73	0.69	0.70	0.70	0.70	0.84	0.84	0.62	0.63	0.72	0.72
SD ^a	0.23	0.23	0.28	0.27	0.13	0.14	0.13	0.13	0.15	0.16	0.22	0.22
Floor (%) ^b	7.90	8.90	16.4	14.5	1.40	1.40	0.50	0.50	6.30	7.40	10.60	10.10
Min	-0.38	-0.32	-0.28	-0.33	0.30	0.30	0.34	0.31	0.15	0.15	-0.06	0.03

HUI Health Utilities Index, QWB Quality of Well-Being, AQoL-8D Assessment of Quality of Life-8D, AUS Australia

^a Ceiling: $U = 1.00$

^b Floor: $U < 0.4$

of its 35 items. The more formal evidence of content validity is presented in Table 6, which reports the Pearson correlation between the dimensions of the SF-36, the three SWB, and the MAU instruments. The table excludes the SF-6D. As it is derived from the SF-36, its correlation with the SF-36 dimensions is an invalid comparator. From

Table 6, the AQoL-8D has the highest correlation with each of the psychosocial dimensions. The difference is particularly significant for mental health where the AQoL-8D correlation is 0.27 and 0.22 points above the average correlation coefficient in the two countries, respectively. In the physical domain, the correlation is higher for general

Table 4 Correlations between MAU instruments

Instrument	Top right-hand side: Pearson correlation							
	Country	EQ-5D	HUI 3	SF-6D	15D	QWB	AQoL-8D	Average
EQ-5D	Australia	–	0.78	0.76	0.81	0.65	0.73	0.75
	US	–	0.80	0.74	0.82	0.65	0.76	0.75
HUI 3	Australia	0.76	–	0.73	0.82	0.65	0.79	0.75
	US	0.78	–	0.73	0.83	0.66	0.82	0.77
SF-6D	Australia	0.66	0.57	–	0.79	0.67	0.77	0.79
	US	0.65	0.59	–	0.79	0.68	0.80	0.80
15D	Australia	0.57	0.49	0.49	–	0.73	0.81	0.67
	US	0.60	0.53	0.52	–	0.71	0.84	0.67
QWB	Australia	0.53	0.53	0.57	0.32	–	0.67	0.76
	US	0.54	0.55	0.60	0.34	–	0.65	0.77
AQoL-8D	Australia	0.73	0.77	0.68	0.57	0.55	–	0.79
	US	0.76	0.79	0.72	0.61	0.55	–	0.80
Average	Australia	0.65	0.62	0.59	0.49	0.50	0.69	–
	US	0.67	0.65	0.62	0.52	0.52	0.69	–

Bottom left hand side: intra-class correlation

MAU multi-attribute utility, HUI Health Utilities Index, QWB Quality of Well-Being, AQoL-8D Assessment of Quality of Life-8D

Table 5 Percent deviation from perfect prediction ($b = 1$)^a in pairwise regression^{b,c}

Independent variable	EQ-5D		SF-6D		HUI 3		15D		QWB		AQoL-8D	
	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US
EQ-5D	–	–										
SF-6D	69	62	–	–								
HUI 3	23	20	108	96	–	–						
15D	79	71	6	6	120	106	–	–				
QWB	48	42	12	14	83	70	20	20	–	–		
AQoL-8D	2	4	67	56	25	25	75	63	45	33	–	–
Average ^c	42	40	52	47	72	63	66	53	42	36	43	36

Source: [Appendix](#)

HUI Health Utilities Index, QWB Quality of Well-Being, AQoL-8D Assessment of Quality of Life-8D, AUS Australia, MAU multi-attribute utility

^a $b = \text{change in MAU}_i / \text{change in MAU}_j$

^b Percent deviation is calculated as the larger divided by the smaller incremental utility

^c Arithmetically, calculating the percent 'deviation' from the larger divided by the smaller marginal change results in a larger number than calculating it from the smaller divided by the larger marginal change (3 is 50 % above 2; 2 is 33.3 % below 2). The average increase from substituting instruments with the larger for instruments with the smaller marginal utility (the average of the final row) is 49.9 %. Substituting instruments with the lower marginal utility results in an average reduction in marginal utility of 32.9 %. Overall, therefore, substituting one instrument chosen at random for another chosen at random causes an average percentage change of 41.4 %

health but below the average for physical function and pain. However, in these cases the correlation is still sufficient to indicate sensitivity to these dimensions. The correlation between the AQoL-8D physical super-dimension and the SF-36 physical component summary (PCS) was 0.80, and indicates that AQoL-8D is sensitive to the

physical dimensions, but that the overall correlation with the full AQoL-8D is reduced because of the increased breadth of the content.

The correlation between the MAU instruments and the three SWB instruments reported in the last three lines of Table 6 is lower than between the MAU instruments.

Table 6 Pearson correlation with dimensions of the SF-36^a

SF-36 dimension	EQ-5D		HUI 3		15D		QWB		AQoL-8D		Average ^b	
	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US	Aus	US
Physical QoL												
General health	0.63	0.63	0.63	0.64	0.75	0.74	0.61	0.61	0.71	0.70	0.65	0.65
Bodily pain	0.77	0.75	0.69	0.67	0.71	0.70	0.60	0.60	0.62	0.65	0.69	0.68
Physical function	0.67	0.67	0.65	0.66	0.71	0.71	.058	0.59	0.52	0.57	0.65	0.66
Role limit	0.49	0.56	0.46	0.56	0.53	0.63	0.43	0.56	0.46	0.57	0.48	0.58
PCS	0.68	0.66	0.62	0.62	0.71	0.69	0.59	0.66	0.50	0.54	0.65	0.66
Mental QoL												
Vitality	0.61	0.61	0.63	0.64	0.73	0.72	0.61	0.60	0.80	0.80	0.65	0.64
Social function	0.63	0.66	0.63	0.67	0.66	0.71	0.54	0.55	0.69	0.74	0.62	0.65
Role limit	0.37	0.50	0.38	0.50	0.41	0.56	0.31	0.47	0.48	0.61	0.37	0.51
Mental health	0.50	0.58	0.55	0.65	0.55	0.63	0.45	0.47	0.76	0.80	0.49	0.58
MCS	0.41	0.50	0.45	0.55	0.47	0.57	0.38	0.44	0.69	0.75	0.44	0.44
Subjective Well-Being												
SWLS	0.40	0.43	0.49	0.49	0.48	0.47	0.42	0.37	0.70	0.62	0.45	0.44
PWI	0.42	0.45	0.49	0.50	0.49	0.49	0.43	0.3	0.69	0.64	0.46	0.45
ONS	0.37	0.36	0.46	0.44	0.43	0.39	0.38	0.29	0.64	0.58	0.41	0.37

HUI Health Utilities Index, *QWB* Quality of Well-Being, *AQoL-8D* Assessment of Quality of Life-8D, *AUS* Australia, *QoL* quality of life, *PCS* physical component summary, *MCS* mental component summary, *SWLS* Satisfaction with Life Scale, *PWI* Personal Wellbeing Index, *ONS* Office for National Statistics

^a Correlations between the SF-6D and the SF-36 dimensions from which it is constructed are not shown. The average (Aus + US) correlation between the SF-6D and the SWLS, PWI and ONS are 0.48, 0.51 and 0.40, respectively

^b Average of the EQ-5D, HUI 3, 15D and QWB

While the three SWB instruments measure closely related constructs, they differ. Nevertheless, the correlation between them and the MAU instruments is similar. The lowest correlation in the Australian sample occurs with the EQ-5D, and in the US with the QWB. The highest correlation in both countries with all three instruments is with AQoL-8D. Its average correlation across the three instruments of 0.65 is 48 % above the average correlation of 0.44 for the remaining instruments.

4.2.3 Reliability

In the second survey, 385 (different) Australian public respondents were invited to complete a baseline survey and to complete two further surveys spaced a fortnight apart. A total of 224 people completed the second-stage survey and all of these respondents completed the third-stage survey. Overall, therefore, 58 % of initial respondents completed all three surveys. The sample contained the same number of men and women (112); approximately 20 % were from the age cohorts below 34, 35–44, 45–54, 55–64 and 65+ years. Educational status was also spread: 35 % had completed only high school; 35 % had additional non-university qualifications and 30 % had a bachelor's degree or above from a university.

Table 7 reports the mean scores of the AQoL-8D and its dimensions at each stage of the survey and the ICC coefficients between the three stages. The standard error of each mean was 0.01. Mean values are relatively stable over the 4-week retest period but increase by a small statistically significant amount for AQoL-8D and each of its dimensions, with the exception of independent living and happiness. The largest increases are for mental health (4.9 %) and senses (4.7 %). AQoL-8D increases by 4.1 %. For group data, a correlation of at least 0.7 is recommended as evidence of satisfactory reliability [8], and each of the ICC coefficients in Table 6 exceeds this threshold, with the exception of the dimensions for senses. Coefficients of 0.9 are considered satisfactory at the individual level for clinical purposes [8]. The AQoL-8D coefficient of 0.89 is close to this higher threshold.

Cronbach alpha coefficients were calculated from the MIC database and reported in the last two columns of Table 7. AQoL-8D alphas are very high in both countries—0.96. The recommended value of 0.7 is also achieved by each of the AQoL-8D dimensions, with the exception of senses. This truncated dimension includes vision, hearing and communication, and the results suggest that there is not a strong underlying construct

Table 7 Survey 2: mean (SE) and ICC coefficients

Dimensions	Mean score (SE) ^a			ICC: baseline–follow-up		Cronbach's alpha	
	Baseline	2-week	4-week	Base–2-week	Base–4-week	Aus	US
Independent living	0.90	0.91	0.90	0.86	0.86	0.89	0.90
Happiness	0.75	0.76	0.76	0.86	0.85	0.90	0.90
Mental health	0.61	0.63**	0.64**	0.87	0.84	0.89	0.89
Coping	0.76	0.77	0.77*	0.82	0.80	0.79	0.79
Relationships	0.69	0.71	0.72**	0.78	0.73	0.86	0.88
Self-worth	0.81	0.82	0.83**	0.86	0.85	0.82	0.81
Pain	0.84	0.85	0.85*	0.85	0.85	0.87	0.86
Senses	0.85	0.87**	0.89**	0.64	0.69	0.51	0.51
Physical super dimension	0.37	0.40**	0.42**	0.84	0.87	0.88	0.88
Mental super dimension	0.75	0.77**	0.78**	0.90	0.86	0.96	0.96
AQoL-8D	0.74	0.76**	0.77	0.91	0.89	0.96	0.96

SE standard error, ICC intra-class correlation, AUS Australia, AQoL Assessment of Quality of Life-8D

**, * indicate significantly different from baseline at 1 % and 5 %, respectively (two-tailed test)

^a The SE around every mean is 0.01

corresponding with these. However, the items were retained due to their intrinsic importance.

5 Discussion

The AQoL-8D extends the range and detail of the psychosocial items in MAU instruments. The ‘opportunity cost’ of this has been a relative reduction in the correlation between the predicted utilities and the physical dimensions which dominate the other MAU instruments. The primary purpose of the present article was, therefore, to determine whether the increased psychosocial content has resulted in convergent and predictive *invalidity* as compared with the comparator instruments. The second objective was to present results from tests of reliability of the instrument and its dimensions.

The tests of content validity reported above confirmed that AQoL-8D is more closely related to psychosocial health and SWB than the five other MAU instruments in the study. However, with one exception, the tests of convergent validity did not produce results that distinguished AQoL-8D from other instruments. Pearson and ICC between AQoL-8D and other MAU instruments resulted in above average coefficients, with the former technique and the highest average correlation using the ICC; however, differences were generally small.

Since MAU instruments all purport to measure the same quantity—the utility of health states—there should be a high level of predictive validity. This has not been found in other studies. Consistent with these earlier results, the present study found that the prediction of differences

between health state utilities was very imperfect. Across all pairwise comparisons, in Table 5, the discrepancy averaged 49.2 %. Prediction by the AQoL-8D is associated with an average 43 and 36 % deviation (Australia/US), which is less than the overall average. Using the comparative criteria adopted here its predictive validity is at least as great as the predictive validity of other MAU instruments.

The main data used for these tests drew upon results from only two of the six countries in the MIC study. As illustrated in all of the present results, the relationships between instruments found in Australia and the US were very similar. This pattern is repeated in the other countries [22]. Repetition of six sets of results would not have altered the conclusions of this paper (but may have changed the focus of interest to country-specific differences). The similarities are unsurprising. While cultural differences may alter responses to a cluster of symptoms, there are no strong reasons for believing that the relationship between instruments—the subject of the present study—should vary between (relatively homogeneous) cultures.

Despite its strengths, there are limitations with the data used in the study. They were obtained from a web-based survey, which means that respondents were from the subset of the population who enrol with a panel company. There are corresponding problems with more conventional survey techniques which typically obtain a response rate <40 %. Nevertheless, the risk of frivolous responses led to a stringent edit procedure. It is still likely respondent panel surveys differ somewhat from the general population. However, it is unlikely that this would generate correlations between instruments. A more important consideration was that the data contained a wide range of observations.

This was achieved by the deliberative sampling of a diverse group of chronically ill respondents in addition to the inclusion of a demographically representative sample of the healthy population. The tests drew upon a larger and more diverse range of observations than previously reported in the literature.

6 Conclusions

The AQoL-8D is the most recent MAU instrument to be constructed. With its emphasis upon psychosocial dimensions of health, it offers significant advantages for evaluation studies where these dimensions are important. However, these advantages may have been at the expense of convergent and predictive validity, as judged by the most widely used MAU instruments. The tests reported here indicate performance by the AQoL-8D that is at least equal to that of other MAU instruments. The tests also indicate good reliability. It may therefore be concluded that the AQoL-8D is a suitable instrument for use in economic evaluation studies, and particularly suitable when psychosocial elements of health are of importance.

Acknowledgments This research was supported by Project Grant ID 1006334 issued from the National Health and Medical Research Council (NHMRC), Australia.

Jeff Richardson was responsible for the design of the study and for drafting the paper.

Angelo Iezzi was responsible for preparation of the questionnaire, including the authorizations for the use of instruments.

Munir A Khan was responsible for analysis of the data and preparation of the tables.

Aimee Maxwell was responsible for the online administration of the questionnaire to the respondents provided by the survey company, CINT, and editing of the data.

Conflicts of interest There were no conflicts of interest affecting any of the four authors.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix

See Table 8.

Table 8 Pairwise geometric mean squares regressions^a

	EQ-5D	SF-6D	HUI 3	15D	QWB	AQoL-8D
EQ-5D	Australian US	EQ = 0.46 + 1.69SF EQ = -0.41 + 1.62SF	EQ = 0.17 + 0.81H EQ = 0.15 + 0.83H	EQ = -0.78 + 1.79D EQ = -0.71 + 1.71D	EQ = -0.2 + 1.48Q EQ = -0.16 + 1.42Q	EQ = 0.2 + 1.02AQ EQ = -0.2 + 1.04AQ
SF-6D	0.58 0.55	Australian US	SF = 0.37 + 0.48H SF = 0.35 + 0.51H	SF = -0.19 + 1.06D SF = -0.19 + 1.06D	SF = 0.16 + 0.89Q SF = 0.15 + 0.88Q	SF = 0.27 + 0.60AQ SF = 0.24 + 0.64AQ
HUI 3	0.61 0.64	0.53 0.53	Australian US	H = -1.16 + 2.20D H = -1.04 + 2.06D	H = -0.45 + 1.83Q H = -0.38 + 1.70Q	H = -0.22 + 1.25AQ H = -0.21 + 1.25AQ
15D	0.66 0.67	0.59 0.62	0.66 0.69	Australian US	D = 0.32 + 0.83Q D = 0.32 + 0.83Q	D = 0.43 + 0.57AQ D = 0.40 + 0.61AQ
QWB	0.42 0.42	0.45 0.45	0.42 0.44	0.53 0.56	Australian US	Q = 0.12 + 0.69AQ Q = 0.10 + 0.74AQ
AQoL-8D	0.53 0.58	0.59 0.62	0.62 0.67	0.66 0.71	0.45 0.42	Australian US

Bottom LHS: R^2 , Australia, US

HUI Health Utilities Index, QWB Quality of Well-Being, AQoL-8D Assessment of Quality of Life-8D

^a Geometric mean squares regressions are obtained from the geometric mean of the coefficients on the two regressions $Y = a_1 + b_1X + e_1$ and $X = a_2 + b_2Y + e_2$. Results are therefore independent of the distribution between choice of dependent and independent variables (see Tofallis [24])

References

- Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med*. 2001;33:358–70.
- Fryback DG, Palta M, Cherepanov D, Bolt D, Kim J. Comparison of 5 health related quality of life indexes using item response theory analysis. *Med Decis Mak*. 2010;30(1):5–15.
- Richardson J, McKie J, Bariola E. Multi attribute utility instruments and their use. In: Culyer AJ, editor. *Online encyclopaedia of health economics*. San Diego: Elsevier Science; 2014.
- Drummond M, Sculpher M, Torrance G, O'Brien B, Stoddart G. *Methods for the economic evaluation of health care programs*. 3rd ed. Oxford: Oxford University Press; 2005.
- Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press; 2007.
- Streiner D, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press; 2003.
- Loevinger J. Objective tests as instruments of psychological theory. *Psychol Rep*. 1957;3:634–94.
- Patrick D, Erickson P. Health status and health policy: quality of life in health care evaluation and resource allocation. Oxford: Oxford University Press; 1993.
- McDowell I. *Measuring health: a guide to rating scales and questionnaires*. Oxford: Oxford University Press; 2006.
- McDonald RP. *Test theory: a unified treatment*. London: Lawrence Erlbaum Associates, Publishers; 1999.
- McDonald RP. Semiconfirmatory factor analysis: the example of anxiety and depression. *Struct Equ Model Multidiscip J*. 2005;12(1):163–72.
- Hawthorne G, Richardson J, Osborne R. The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of health related quality of life. *Qual Life Res*. 1999;8:209–24.
- Richardson J, Day NA, Peacock S, Iezzi A. Measurement of the quality of life for economic evaluation and the Assessment of Quality of Life (AQoL) Mark 2 Instrument. *Aust Econ Rev*. 2004;37(1):62–88.
- Richardson J, Elsworth G, Iezzi A, Khan MA, Mihalopoulos C, Schweitzer I, et al. Increasing the sensitivity of the AQoL inventory for evaluation of interventions affecting mental health. Research paper 61. Melbourne: Centre for Health Economics, Monash University; 2011.
- Richardson J, Sinha K, Iezzi A, Khan M. Modelling the utility of health states with the Assessment of Quality of Life (AQoL) 8D instrument: overview and utility scoring algorithm. Research paper 63. Melbourne: Centre for Health Economics, Monash University; 2011.
- Von Winterfeldt D, Edwards W. *Decision analysis and behavioral research*. Cambridge: Cambridge University Press; 1986.
- Simmons CA, Lehmann P. *Tools for strengths-based assessment and evaluation*. New York: Springer; 2013.
- OECD. *Guidelines on measuring subjective well-being*. Paris: OECD Publishing. 2013. <http://dx.doi.org/10.1787/9789264191655-en>. Accessed 22 Apr 2013.
- Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. *J Personal Assess*. 1985;49(1):71–5.
- International Wellbeing Group. *Personal Wellbeing Index*. Melbourne: Australian Centre on Quality of Life, Deakin University. 2006. http://www.deakin.edu.au/research/acqol/instruments/wellbeing_index.htm. Accessed 6 Nov 2012.
- ONS. Findings from the National Well-Being Debate. UK Office for National Statistics paper, July 2011. 2011 <http://www.ons.gov.uk>. Accessed 10 Oct 2013.
- Richardson J, Khan MA, Iezzi A, Maxwell A. 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. Centre for Health Economics, Monash University. 2012. <http://www.buseco.monash.edu.au/centres/che/che-publications.html>. Accessed 25 Jan 2013.
- Richardson J, Iezzi A, Khan MA, Maxwell A. Cross-national comparison of twelve quality of life instruments: MIC paper 1: background, questions, instruments. Research paper 76. Melbourne: Centre for Health Economics, Monash University. 2012a. <http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper76.pdf>. Accessed 29 Jul 2013.
- Tofallis C. Model fitting for multiple variables by minimising the geometric mean deviation. In: Van Huffel S, Lemmerling P, editors. *Total least squares and errors-in-variables modeling: algorithms, analysis and applications*. Kluwer Academic; 2002. <http://ssrn.com/abstract=1077322>.