Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data

Aimee Maxwell, Mehmet Özmen, Angelo Iezzi & Jeff Richardson

Quality of Life Research
An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation - Official Journal of the International Society of Quality of Life Research

ISSN 0962-9343

Qual Life Res
DOI 10.1007/s11136-016-1337-z

Qual Life Res
DOI 10.1007/s11136-016-1337-z

Deriving population norms for the AQoL-6D and AQoL-8D multi-attribute utility instruments from web-based data

Aimee Maxwell[1] · Mehmet Özmen[2] · Angelo Iezzi[1] · Jeff Richardson[1]

Accepted: 7 June 2016

Abstract

Objectives (i) to demonstrate a method which ameliorates the problem of self-selection in the estimation of population norms from web-based data and (ii) to use the method to calculate population norms for two multi-attribute utility (MAU) instruments, the AQoL-6D and AQoL-8D, and population norms for the sub-scales from which they are constructed.

Methods A web-based survey administered the AQoL-8D MAU instrument (which subsumes the AQoL-6D questionnaire), to members of the public along with the AQoL-4D which has extant population norms. Age, gender and the AQoL-4D were used as post-stratification auxiliary variables to construct weights to ameliorate the potential effects of self-selection associated with web-based surveys. The weights were used to estimate unbiased population norms. Standard errors from the weighted samples were calculated using Jackknife estimation.

Results For both AQoL-6D and AQoL-8D, physical health dimensions decline significantly with age. In contrast, for the majority of the psycho-social dimensions there is a significant U-shaped profile. The net effect is a shallow U-shaped relationship between age and both the AQoL-6D and AQoL-8D utilities. This contrasts with the almost monotonic decline in the utilities derived from the AQoL-4D and SF-6D MAU instruments.

Conclusions Post-stratification weights were used to ameliorate potential bias in the derivation of norms from web-based data for the AQoL-6D and AQoL-8D. The methods may be used generally to obtain norms when suitable auxiliary variables are available. The inclusion of an enlarged psycho-social component in the two instruments significantly alters the demographic profile.

Keywords CUA · Norms · AQoL · QoL · Multi-attribute utility

Introduction

Numerous 'instruments' have been created which produce indices of the health-related quality of life (HRQoL) [1–3]. A subset of these—multi-attribute utility (MAU) instruments—provide utility weights or a utility algorithm to create indices of health state utilities which are used in conjunction with epidemiological or clinical studies to conduct economic evaluations and, specifically, to calculate quality-adjusted life years (QALYs).

The main MAU instruments, which are described in Brazier et al. [4] and Richardson et al. [5], consist of a descriptive system (or classification)—a set of health-related questions and response categories—and a corresponding set of utility weights, which converts responses to a numerical score on a scale where 0.00 represents death and 1.00 represents the utility of the best health state described by the descriptive system. Seven instruments dominate the field: the EQ-5D [6, 7], HUI 2 [8], HUI 3 [9], SF-6D [10, 11], 15D [12], AQoL-4D [13] and QWB [14].

Electronic supplementary material The online version of this article (doi:10.1007/s11136-016-1337-z) contains supplementary material, which is available to authorized users.

✉ Jeff Richardson
jeffrey.richardson@monash.edu

1 Centre for Health Economics, Monash Business School, Monash University, Melbourne, Australia

2 Department of Econometrics and Business Statistics, Monash Business School, Monash University, Melbourne, Australia

Published online: 25 June 2016

The AQoL instruments were created to increase sensitivity to particular dimensions of health. Like the SF-6D, the AQoL-4D includes items describing both physical and psycho-social health, but it increases the depth of the description. AQoL-6D extends the instrument to achieve greater sensitivity in the vicinity of good health states. AQoL-7D adds a specialised dimension (VisQoL) for vision to the AQoL-6D [15, 16], and AQoL-8D adds two additional dimensions relating to psycho-social dimensions of health [17, 18].

Transformations between the instruments and other MAU instruments have been published [19]. Recent analysis of 6 MAU instruments included in a large cross-national population, and patient survey found AQoL-8D to have greater sensitivity to the majority of the psycho-social dimensions of the SF-36 and greater sensitivity to the psycho-social effects of 7 chronic diseases than alternative MAU instruments [20–23]. By June 2016, 317 research groups had registered to use AQoL-6D or AQoL-8D instruments.

Figure 1 depicts the AQoL-8D. The AQoL-6D is subsumed by AQoL-8D and is shown as the shaded items and dimensions. AQoL-8D consists of five psycho-social and three physical dimensions. With one exception, each of these represents a psychometrically valid sub-scale, i.e. tests indicate they measure a common construct [17]. The exception is the dimension for senses where the three constituent items—vision, hearing and communication—do not reflect a common construct. It was retained in the instrument because of the importance of these items for the quality of life. Six of the eight dimensions of AQoL-8D occur in AQoL-6D including the three physical and three of the psycho-social dimensions. The three physical dimensions are related to a single construct (the 'physical super-dimension') and the five psycho-social dimensions to a single construct (the 'mental super-dimension').

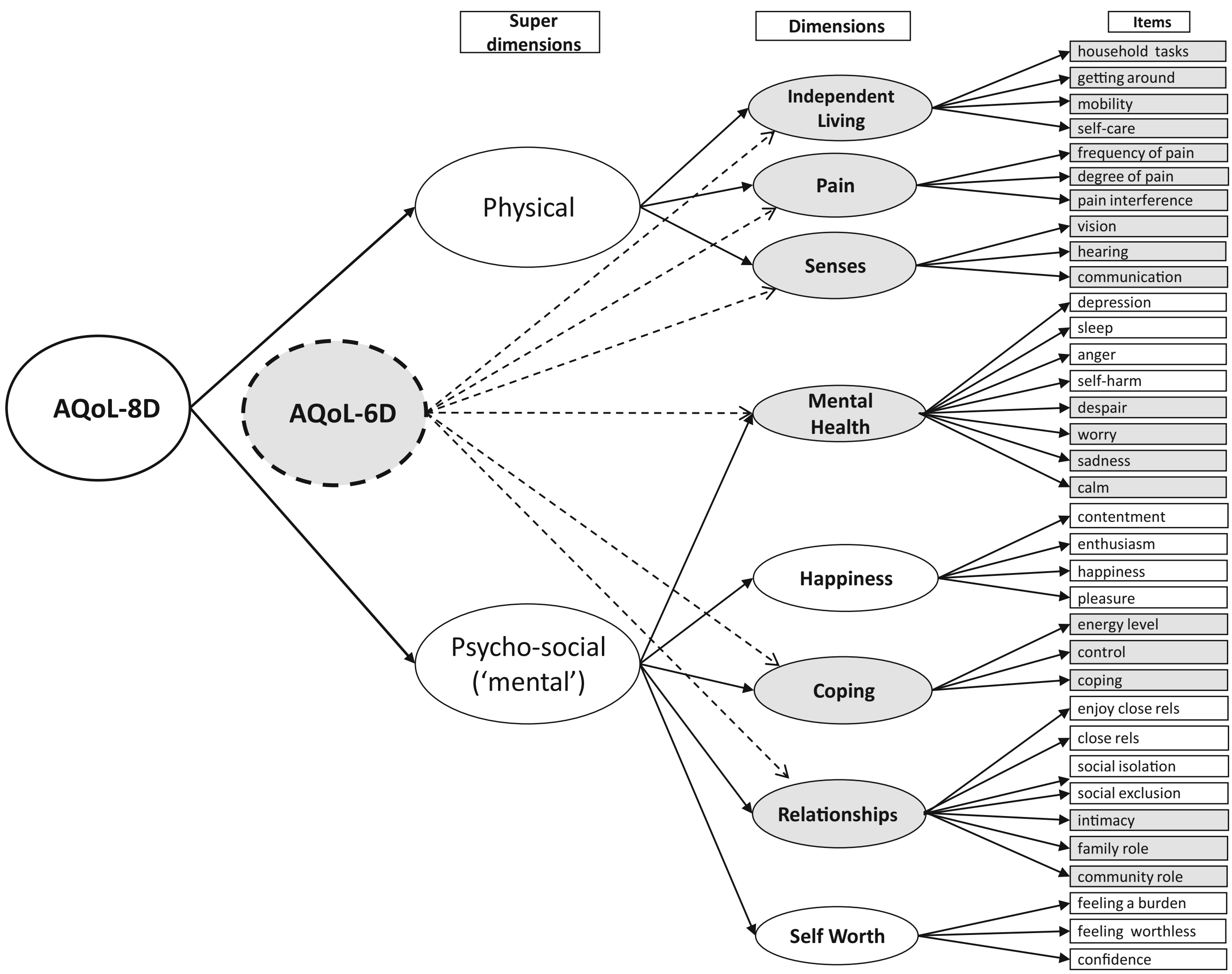


Fig. 1 AQoL-8D and AQoL-6D instruments*.
*AQoL-6D is shown as the shaded items and dimensions. It does not map into psychometrically valid 'super-dimensions' shown for AQoL-8D

To assist with the interpretation of the numbers produced by instruments, it is common practice to provide 'population norms': estimates of the average instrument values for different age–gender cohorts. Norms have multiple purposes. First, and most generally, they help interpret the numbers obtained from an instrument; whether a particular score represents an excellent, normal or poor health relative to the general population. Secondly, in observational studies they permit an estimate to be made of the difference between a population subgroup and the general population. This may be the basis for estimating the burden of disease associated with an illness and the distribution of the burden by demographic cohort. Third, in the absence of longitudinal data, norms may be used when baseline data exist to estimate the improvement in health status from an intervention which will return patients to normal health.

In Australia, norms have been produced for a number of HRQoL-related instruments: the SF-36 version 1 and 2 [24, 25], the K10 [26], the WHOQoL-BREF [27] and the Personal Wellbeing Index [28]. Australian norms for MAU instruments have been published for the SF-6D [29] and for the AQoL-4D [30]. These are summarised in Table 1.

Population norms are generally estimated from the results of a large and reputable national or area-based survey, which achieves a sufficiently high response rate to give confidence in the representativeness of the results. Norms were published for AQoL-4D in 2013 using data from the 2007 National Survey of Mental Health and Wellbeing (NSMHWB) [30, 32].

Data from this source may be compromised for two reasons. First, without the quality control of a personal interview, a non-trivial proportion of web-based respondents may provide inconsistent or arbitrary data. Second, individuals who enrol with panel companies are, by definition, self-selected. Hence, when survey respondents are stratified—grouped into defined categories—data from these categories may not give a true representation of the population. In particular, previous experience indicates that, on average, web survey respondents have quality of life scores that are lower than the general public [33]. However, after initial stratification, bias may be mitigated by weighting with auxiliary variables, if available, which have established population distributions and a strong relationship to the instruments of interest.

The objective of the present paper is twofold. First, it illustrates a general method for estimating population norms for an instrument using web-based data when suitable auxiliary variables are available. In this case, AQoL-4D scores, age and gender are employed as the auxiliary variables. Secondly, it presents the resulting norms for the AQoL-6D and AQoL-8D instruments and for their constituent dimensions. Validity of the method depends upon the extent to which AQoL-4D identifies population attributes which are measured by AQoL-6D and AQoL-8D. This question is examined in Appendix 4 in ESM. Results are presented and described below. The web-based survey was approved by the Monash University Human Research Ethics Committee Approval ID: CF15/2829-2015001164.

Table 1 Population norms: summary of three studies

Instrument	Age					
	16–19[d]	20–29	30–39	40–49	50–59	60–69
AQoL-4D[a]	0.87	0.87	0.85	0.85	0.80	0.79
AQoL-4D[b]	0.87	0.86	0.84	0.81	0.80	0.80
SF-6D[c]	–	0.79	0.79	0.77	0.75	0.74

[a] Hawthorne and Osborne [31]

[b] Hawthorne et al. [30]

[c] Norman et al. [29]

[d] The age range for the 1998 AQoL-4D norm was 15–19

Methods

Members of the public enrolled with a panel company, CINT Pty Ltd, were invited to complete a survey, which included the AQoL-4D and AQoL-8D. As the eight dimensions of the latter instrument subsume the dimensions of the AQoL-6D, the survey data were also used to construct AQoL-6D norms. Respondents were assigned to age–gender cohorts until a pre-selected number was reached. This was determined by the demographic composition of the Australian public and by the initial target of 3000 respondents.

The Survey

The survey was administered by a speaking avatar which introduced the questions in the following way:

> Welcome to the AQoL norming project… We are relying upon you to give truthful and considered answers…Please read this explanatory statement carefully. Feel free to ask questions about any information in the document. You may also wish to discuss the project with a relative or friend…Participation will involve completing an online questionnaire. The questionnaire is simple and will take about 10 min to complete. The project will be carried out according to the national statement on ethical conduct in research involving humans.

The introduction was followed by the administration of the AQoL-4D. Utility scores were calculated, and the respondent was assigned to the relevant demographic cohort. The survey continued if the quota for that cohort was not full. The order of the AQoL-8D questions was randomised to offset bias from an ordering effect. The questionnaire ended with a limited number of questions regarding education and current economic status.

Editing

Four criteria were applied to ensure data quality. These were based upon methods outlined by Meade and Craig [34] and are discussed in greater detail in Appendix 1 in ESM. They were: (i) giving answers to any of the three repeated questions which differed by more than one level on the four response level scale; (ii) reporting their age to be 1 + years different from their age calculated from their reported year of birth; (iii) providing answers, which gave scores on the AQoL-4D and AQoL-8D, which differed by more than 0.5; and (iv) claiming their highest educational attainment was 'part primary school' (a category included to assist editing as it is less than the compulsory minimum).

Analysis

It is common for survey data to exhibit disproportionate sampling, whether by design or as a result of bias in the sampling procedure. While the reasons behind self-selection and non-response bias are complex, post-stratification survey weights can mathematically align the survey to the population along a small number of dimensions, to reduce many of these complex biases.

In the present study, age, gender and AQoL-4D utility were selected as post-stratification auxiliaries to adjust results for the bias arising from the self-selection of survey participants. They were selected as they are good predictors of the AQoL-6D and AQoL-8D (See Appendix 4 in ESM) and because there is reliable information on their distribution in the Australian population. Age and gender were obtained from the Australian Bureau of Statistics (ABS) census [32]. Estimates for the cohort-specific distribution of the AQoL-4D distribution were obtained from another ABS survey. The 2007 National Survey of Mental Health and Wellbeing (NSMHWB) was a nationally representative cross-sectional household survey. Sampling was based on random selection from a stratified multi-stage area probability sample of private dwellings. In total, 8441 participants were interviewed. Oversampling resulted in 1269 respondents aged 70 or above, which allowed reliable estimates of the AQoL-4D to be made for the elderly. The relation between AQoL-4D, AQoL-6D and AQoL-8D is outlined in Appendix 4 in ESM.

For each age (i) and gender (j) cohort, the NSMHWB provides the distribution of AQoL-4D utilities across k categories, where each category is defined by a range of utility scores [−0.04 to <0.2; 0.2 to <0.4; 0.4 to <0.6; 0.6 to <0.8; 0.8 to <1.0 and 1.0]. Age and gender cohort counts available from the national census data (N_{ij}) were further split into AQoL-4D categories in the same proportion as occurred in the NSMHWB data to produce counts by age, gender and AQoL-4D categories (N_{ijk}).

The population in each cell N_{ijk} was converted to a proportion, p_{ijk}, of the total Australia population, N aged 15–74. Therefore, $p_{ijk} = N_{ijk}/N$. Similarly p^*_{ijk} was calculated as $p^*_{ijk} = n_{ijk}/n$ where n is the total number of valid web-based survey respondents and n_{ijk} the respondent number in cell ijk. Consequently, p^*_{ijk} is the proportion of web-based survey respondents in the cell ijk. These proportions were used to calculate the proportional weights $w_{ijk} = p_{ijk}/p^*_{ijk}$ and cohort-specific population weights $W_{ijk} = w_{ijk} \times N/n = N_{ijk}/n_{ijk}$. The sum of the weights W_{ijk} equals the population sizes within each post-stratum and were used to estimate stratum means and standard errors. With complex survey designs, derivation of the standard error of a weighted mean is complex. Various methods are discussed by Gatz and Smith [35]. Here, standard errors were estimated using a 'design-based' method that makes no assumptions about the model that is generating the data, but instead uses available sample information. The population weight, W_{ijk}, scales cohort counts to the correct population level, while maintaining information found in the proportional weights. This is necessary since the standard error of the mean estimate is the standard deviation of the sample means of all possible samples drawn from the population and hence is a property associated with the population size across cohorts. The data are scaled to the population level and standard errors estimated using the Jackknife methodology.

Jackknife replication simulates the distribution of repeated samples by taking a series of random, but unbiased, sub-samples from the observed sample and measures the variability between these sub-samples to estimate the sampling variance. Each sub-sample gives an unbiased estimate of the population mean, and therefore, the variance between the sub-sample means gives an estimate of the true sampling variance which is the definition of the standard error. The Stata 'svy' package was used to estimate standard errors using Jackknife variance estimation with weights, W_{ijk}, defined as described above.

Results

A letter of invitation was emailed to 95,975 individuals of whom 72 per cent did not open the email. Of the 27,159 people who read the letter, 9420 (35 per cent) clicked the link and read the study information. Of these, 6488 (69 per cent) commenced the survey, but 3019 (47 per cent) were screened out as cell quotas were filled. Of those passing the quota process, 99 per cent finished the survey and were paid approximately $1.08 upon completion. There were 3228 individuals aged 74 or less. Data editing removed 497 cases (or 15.4 per cent). The deletions by criteria and by age–gender cohort are reported in Appendix 1 in ESM. The remaining 2731 respondents are classified by age and gender in Table 2.

Respondents had a mean age of 46.6 years (SD = 16.05), 46.1 per cent were males and 53.9 per cent, females. Seventy-five per cent of participants were born in Australia, 25 per cent born elsewhere. In total, 58 per cent of the sample were married or in a de-facto relationship, 28 per cent were single; 31.9 per cent had only completed primary or secondary school; 35.9 and 31.1 per cent, respectively, held a vocational certificate or diploma or had completed a university degree. Twenty-eight per cent of respondents were employed full time in the labour force; 31.8 per cent were retired/pensioners/homemakers; and 6.3 per cent were students. Median reported income was in the range of $52,000 to $72,748 per annum; 11 per cent reported incomes less than $18,200 per annum (10 per cent refused to answer the income question). Seventy per cent of respondents rated their own current health as good, very good or excellent, 30 per cent reported that their current health was fair, poor or very poor.

Dimensions

Population norms for the dimensions of the AQoL-8D are shown in Fig. 2 and reported in full in Appendix 2 in ESM. Dimension scores are obtained from item responses weighted for self-selection, but not by utility weights. The number of response categories per item, the number of items per dimension and the minimum and maximum scores per dimension vary. Consequently, each dimension scale is unique and scores across different dimensions cannot be compared. Differences between scores on a single scale may be compared, and the significance of a change in the score assessed in relation to the standard error.

Psycho-social dimensions

For both men and women, the age profile of the psycho-social super-dimension (MSD) is U-shaped against age with the lowest scores for men and women in the cohorts aged 35–44 and 45–54, respectively. The age-related variation is greater for men than women primarily because of the high male scores in the age range 16–24 and 65–74, particularly, the former. The profile is largely driven by happiness and coping. The mean values for 16- to 24-year-old men for these dimensions are 0.084 and 0.089 points above the mean value in the 45–54-year-old cohort. Mean values for mental health relationships and self-worth in the 16–24 age cohort exceed mean values for men at their psycho-social nadir (age 35–44) by 0.082, 0.058 and 0.042 points, respectively. By age 65–74, the mean MSD score for men has risen from its nadir (0.463) by 0.113 and is only 0.029 below the mean value for 18–24-year-old males. At age 65–74 male, self-worth reaches the highest score for any cohort.

The U-shaped profile for women is more muted. Self-worth rises monotonically except for a small decline in the overall nadir years of 45–54. Happiness, coping, mental health and relationships all reach their minimum in the age range 45–54 and are, respectively, 0.032, 0.023, 0.032 and 0.029 points below their score in the youngest cohort. However, all psycho-social dimensions subsequently rise, and the oldest cohort has an MSD score which is 0.094 above the nadir cohort (age 45–54) and a higher score for every psycho-social dimension than women aged 16–24.

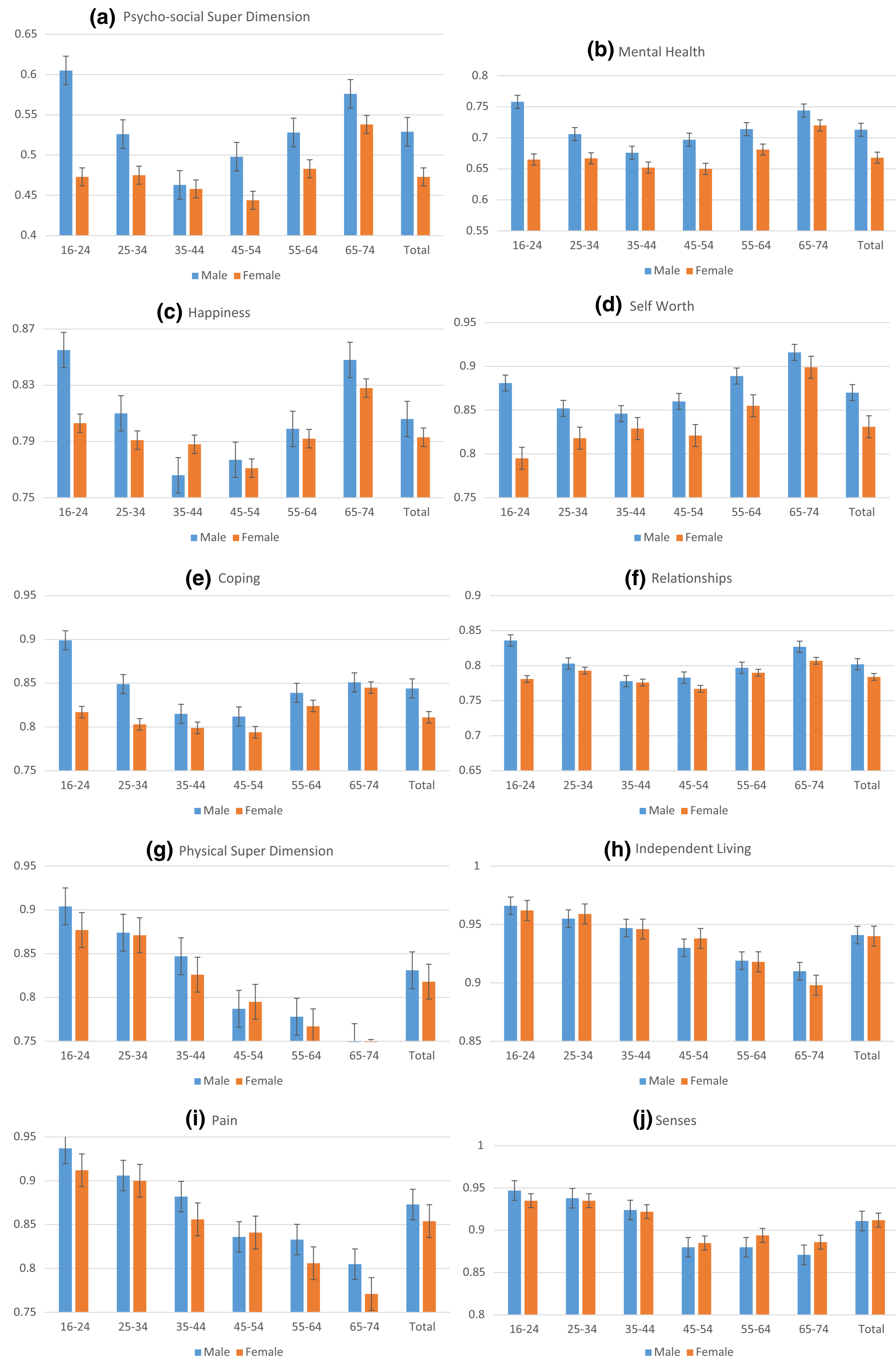
Physical dimensions

The physical super-dimension (PSD) declines monotonically with age for both men and women. The decrease is driven primarily by pain. Individual living and senses both decline monotonically for both genders, but the decline is less steep. Males obtain higher values in most cohorts for pain and senses with the exception of men aged 35–44.

The male cohort aged 16–24 have a mean value for every physical dimension which is greater than any other cohort; the 25–34-year-old cohort mean values are

Table 2 Number and per cent of respondents by age and gender after editing

	n	16–24	25–34	35–44	45–54	55–64	65–74	Total
Males	1258	8.4	15.5	17.7	18.2	21.5	18.7	100
Females	1473	14.1	17.0	17.6	18.3	18.5	14.5	100
Total	2731	11.5	16.3	17.7	18.2	19.9	16.4	100


(a) Psycho-social Super Dimension
0.65
0.6
0.55
0.5
0.45
0.4
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(b) Mental Health
0.8
0.75
0.7
0.65
0.6
0.55
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(c) Happiness
0.87
0.83
0.79
0.75
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(d) Self Worth
0.95
0.9
0.85
0.8
0.75
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(e) Coping
0.95
0.9
0.85
0.8
0.75
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(f) Relationships
0.9
0.85
0.8
0.75
0.7
0.65
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(g) Physical Super Dimension
0.95
0.9
0.85
0.8
0.75
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(h) Independent Living
1
0.95
0.9
0.85
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(i) Pain
0.95
0.9
0.85
0.8
0.75
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female
(j) Senses
1
0.95
0.9
0.85
0.8
16-24
25-34
35-44
45-54
55-64
65-74
Total
Male Female

◀**Fig. 2** Dimension scores by age and gender. **a** Psycho-social Super-Dimension (MSD) **b** Mental Health **c** Happiness **d** Self-worth **e** Coping **f** Relationships **g** Physical Super-Dimension **h** Independent Living **i** Pain **j** Senses *Source*: 2(a) to 2(j) Appendix 2 in ESM

similarly greater than for any other older cohort and those aged 35–44 also achieve greater mean values for all physical dimensions than more elderly men. In contrast with the psycho-social dimensions, there is little difference between average scores for men and women on the physical dimensions. Men have higher scores for pain (less pain is experienced), but no significant differences exist between genders for independent living or senses. The PSD score for men is greater than for women in every age cohort, with the exception of the age group 45–54. Differences, however, are small, and the mean value of the PSD for men exceeds the mean for women by only 0.013.

Norms for AQoL-6D versus AQoL-8D

The pattern of AQoL-6D and AQoL-8D norms reported in Table 3 is very similar. The profile for both genders for both instruments is U-shaped with the nadir for both genders for both instruments occurring in the 45–54-year-old cohort. The decline to the nadir is almost identical for the two instruments (0.06), with a greater decline for men than for women. The subsequent recovery on the AQoL-8D is insignificantly greater for both men and women. This is because the AQoL-6D does not include happiness or self-worth both of which exhibit pronounced U-shaped profiles for both men and women. The importance of the

Table 3 Weighted AQoL-8D and AQoL-6D population norms by age and gender

Gender	Age group	Age, gender and AQoL-4D weighted										Age and gender weighted		Differences in estimates between weighting schemes	
		AQoL-6D					AQoL-8D								
		Mean	SE	95 % CI LB	95 % CI UB	Est. Pop. SD	Mean	SE	95 % CI LB	95 % CI UB	Est. Pop. SD	AQoL-6D Mean	AQoL-8D Mean	AQoL-6D	AQoL-8D
Male	16–24	0.904	0.010	0.884	0.924	0.076	0.869	0.013	0.843	0.895	0.093	0.826	0.778	0.078*	0.091*
	25–34	0.865	0.010	0.845	0.885	0.134	0.807	0.012	0.784	0.830	0.218	0.785	0.725	0.066*	0.082*
	35–44	0.846	0.009	0.828	0.864	0.145	0.820	0.012	0.797	0.844	0.154	0.745	0.686	0.120*	0.134*
	45–54	0.824	0.012	0.801	0.847	0.182	0.791	0.011	0.770	0.813	0.161	0.718	0.664	0.128*	0.127*
	55–64	0.851	0.010	0.832	0.870	0.187	0.780	0.013	0.754	0.807	0.198	0.747	0.684	0.077*	0.096*
	65–74	0.869	0.009	0.852	0.886	0.186	0.836	0.010	0.816	0.856	0.217	0.797	0.751	0.072*	0.085*
	Total	0.859	0.004	0.851	0.868	0.152	0.816	0.005	0.806	0.826	0.173	0.768	0.713	0.091*	0.103*
Female	16–24	0.840	0.012	0.816	0.863	0.147	0.786	0.014	0.758	0.814	0.167	0.709	0.642	0.131*	0.144*
	25–34	0.834	0.011	0.812	0.855	0.166	0.789	0.013	0.764	0.815	0.191	0.696	0.636	0.138*	0.153*
	35–44	0.819	0.010	0.800	0.839	0.165	0.777	0.012	0.754	0.800	0.187	0.699	0.641	0.120*	0.136*
	45–54	0.803	0.011	0.782	0.825	0.188	0.757	0.012	0.734	0.781	0.204	0.687	0.631	0.116*	0.126*
	55–64	0.837	0.009	0.818	0.855	0.184	0.785	0.011	0.764	0.806	0.203	0.708	0.645	0.129*	0.140*
	65–74	0.850	0.010	0.831	0.869	0.191	0.814	0.011	0.793	0.836	0.213	0.770	0.722	0.080*	0.092*
	Total	0.828	0.004	0.820	0.837	0.174	0.782	0.005	0.772	0.792	0.194	0.707	0.647	0.122*	0.135*
All	16–24	0.873	0.008	0.857	0.888	0.114	0.828	0.010	0.809	0.848	0.133	0.768	0.711	0.105*	0.117*
	25–34	0.849	0.007	0.835	0.864	0.151	0.805	0.009	0.787	0.822	0.173	0.740	0.680	0.109*	0.125*
	35–44	0.832	0.007	0.819	0.846	0.156	0.784	0.008	0.768	0.799	0.174	0.722	0.663	0.111*	0.121*
	45–54	0.813	0.008	0.798	0.829	0.185	0.769	0.009	0.751	0.786	0.202	0.702	0.647	0.111*	0.122*
	55–64	0.844	0.007	0.831	0.857	0.186	0.796	0.008	0.780	0.811	0.211	0.728	0.665	0.117*	0.131*
	65–74	0.859	0.007	0.846	0.872	0.189	0.825	0.007	0.810	0.840	0.216	0.784	0.736	0.075*	0.089*
	Total	0.844	0.003	0.838	0.850	0.164	0.799	0.004	0.792	0.806	0.185	0.737	0.680	0.107*	0.119*

* 95 % confidence intervals for estimates across the weighting schemes do not overlap. The estimates are significantly different

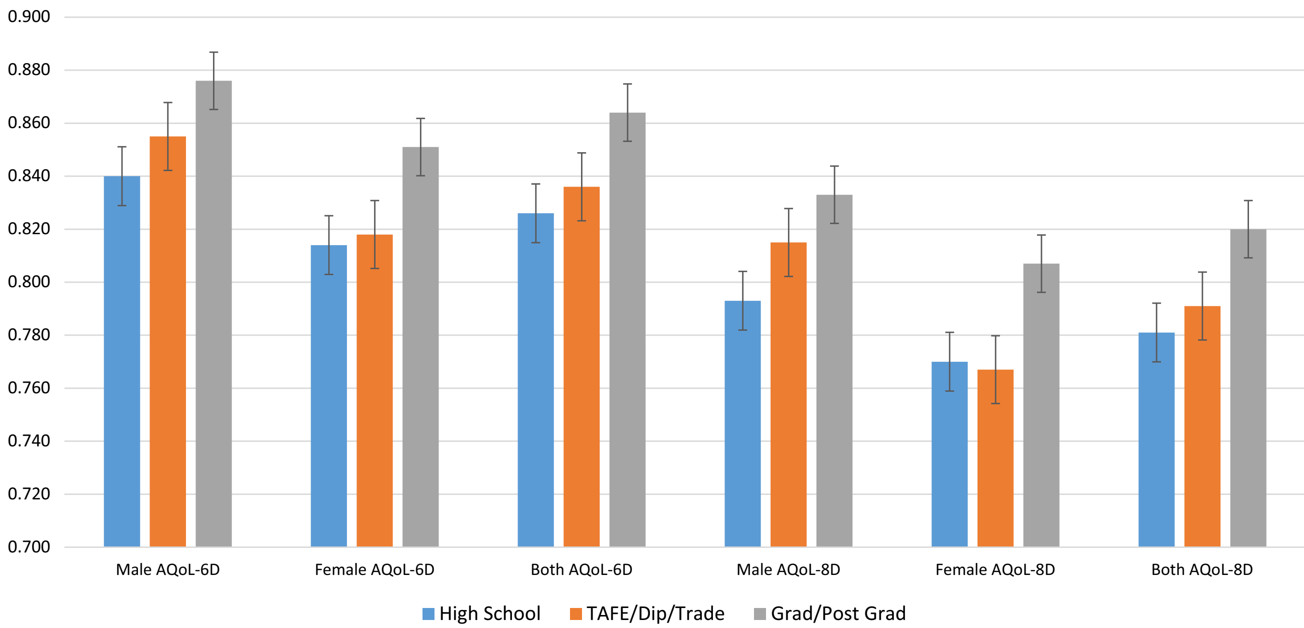


Fig. 3 Mean weighted norms by gender and education

adjustment for self-selection employed in the study is indicated in the final two columns of Table 3. These indicate that, as expected, web-based respondents had atypically low utilities so adjustment led to an increase in both AQoL-6D and AQoL-8D norms.

Education

Figure 3 presents norms by education status for both instruments. Norms by dimensions and gender are reported on the AQoL website [36]. Unweighted norms for the AQoL-6D and 8D and norms for the AQoL-8, an 8-item brief instrument based upon a reduction of the AQoL-4D by Hawthorne [37] can also be found on the website.

Discussion

To date, researchers wishing to use a particular instrument have been restricted to the use of norms which have often been created in another country and generally for a representative sample of the entire (non-institutionalised) population. The available norms for a country may not correspond with the most sensitive instrument for a particular population group. For example, the 15D MAU instrument displays greater sensitivity to a number of disease-related health states than other commonly used MAU instruments [20]. However, its Finnish norms undoubtedly inhibit its use. The present paper has demonstrated a low-cost method for overcoming this limitation using a web-based survey.

To obtain a reliable sample of 2731 an initial 95,975 individuals were approached. This implies that the principles of sampling which apply to normal surveys have no application: the scale of self-selection introduces an unknown but potentially very large bias. The methods described in this paper illustrate how the problem may be overcome or significantly ameliorated. The analytical basis of this method is theoretically plausible. To the extent that the ABS profile of the age–gender distribution and the profile of AQoL-4D utilities reflect the true profiles in the Australian population and to the extent that AQoL-4D utilities reflect the relevant attributes of the AQoL-6D and AQoL-8D then adjusting our sample to replicate the ABS profile ensures a valid set of AQoL-6D and AQoL-8D norms. ABS age–gender distributions were derived from the national census. AQoL-4D norms were derived from an ABS survey, which employed best available survey techniques. [32].

However, as outlined in Appendix 4 in ESM the content of AQoL-4D, AQoL-6D and AQoL-8D differ (or multiple instruments would be unnecessary). The differences are not random. Both AQoL-6D and AQoL-8D increased the psycho-social content of the construct measured. Consequently, individuals with the same AQoL-4D score may have dissimilar but compensating physical and psycho-social scores. If web-based respondents, on average, have poorer mental health, this will not be fully reflected in the AQoL-4D scores and the methodology will not fully take account of self-selection by those with poor mental health. Psycho-social dimension norms therefore have the possibility of downward bias and physical dimension norms of upward bias. The net effect of this secondary self-selection on AQoL-6D and AQoL-8D norms is unlikely to be significant as the two effects are offsetting, and the replication of the national profile of a highly correlated utility instrument suggests confidence in the results.

A further caveat to the present results is that while ABS interviews of elderly respondents may produce reliable answers from sufficient respondents, web-based surveys of this cohort are more problematical, partly because the lower computer literacy in this age range exacerbates the problem of self-selection but also because, without a personal assessment of competency, confidence in the validity of responses declines with age. For these reasons, the present norms were not extended to people above the age of 74 and, consequently, the present norms cannot be used for analyses involving this cohort.

The existence or extent of residual bias in the present norms cannot be determined as there is no gold standard against which to validate results. Norms for the two new AQoL instruments differ significantly from those obtained from other MAU instruments. This reflects the greater psycho-social content of the instruments and the improvement in these dimensions which occurs beyond a certain age. The difference implies that they have a comparative advantage for the measurement of health states where psycho-social health is of importance.

The contrast between the norms obtained for Australian utility instruments is illustrated in Fig. 4. Differences reflect the different methods for assigning utilities. However, the primary difference is the type and balance of questions in the instruments. Relative to the AQoL-8D, the SF-6D and AQoL-4D have a preponderance of physical items, resulting in a monotonic decline in utilities as occurs with the AQoL-8D physical super-dimension. In contrast, the AQoL-6D and AQoL-8D profiles reflect the age-related improvements in psycho-social health. In the context of happiness, this age-related improvement has been controversial [38]. The present results strongly endorse the observation and illustrate that age-related improvement in psycho-social health is more broadly based than the happiness dimension.

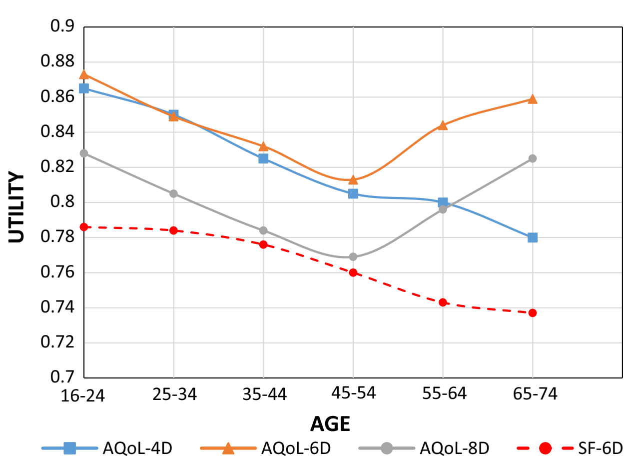


Fig. 4 Mean utility (male and female) AQoL-4D (Hawthorne et al. [30]), SF-6D (Norman et al. [29]), AQoL-6D and AQoL-8D (Table 3)

In practice, all instruments which purport to measure utility give numerical values which differ significantly. The largest six instrument comparative study to date found, on average, absolute differences between utilities predicted for individuals of 0.135. The discrepancy was attributable to differences in questions in the descriptive system and to differences in the measurement scale, which compressed or inflated responses differently [39]. The AQoL instruments are subject to the same caveats. Their properties have been compared and documented [40]. A statistical transformation exists on the AQoL website, which aligns the scales. Likewise, transformations have been published between both AQoL-6D and AQoL-8D and each of the major MAU instruments: EQ-5D-5L, SF-6D, HUI 3, 15D, QWB [19]. Transformations align instrument scales but largely preserve the sensitivity of the original instrument.

Conclusion

This paper had two objectives: the first was to demonstrate the use of post-stratification weights to mitigate the effects of self-selection in a web-based survey when reliable auxiliary variables are available; the second objective was to apply this method to produce norms for the AQoL-6D, AQoL-8D and their dimensions. The objectives were achieved using ABS demographic data and AQoL-4D results from the ABS NSMHWB survey as auxiliary variables. Norms for both instruments give a U-shaped age profile, reflecting the positive effect of ageing upon the psycho-social dimensions which are heavily represented in these instruments. Results do not represent a gold standard, but the rigour of the techniques employed, and plausibility of the profiles should give confidence that the method might be used more generally when there is a need for norms for a particular instrument or for a particular population or sub-population.

Acknowledgments This research was funded by a National Health and Medical Research Council Project Grant ID: 1006334.

Compliance with ethical standards The research has been approved by the Monash University Human Research Ethics Committee Approval ID: CF15/2829-2015001164.

Conflict of interest The authors declare that they have no conflict of interest.

Funding This research was funded by National Health and Medical Research Council Project Grant ID: 1006334.

References

1. Simmons, C. A., & Lehmann, P. (2013). *Tools for strengths-based assessment and evaluation*. New York: Springer.
2. Bowling, A. (2005). *Measuring health: A review of quality of life measurement scales* (3rd ed.). Maidenhead, Berkshire: Open University Press.
3. McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires*. Oxford: Oxford University Press.
4. Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2007). *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press.
5. Richardson, J., McKie, J., & Bariola, E. (2014). Multi attribute utility instruments and their use. In A. J. Culyer (Ed.), *Encyclopedia of health economics* (pp. 341–357). San Diego: Elsevier Science.
6. EuroQol Group. (1990). EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy, 16*, 199–208.
7. Dolan, P., Gudex, C., Kind, P., Williams, A. (1995). *A social tariff for EuroQoL: Results from a UK general population survey*. Discussion Paper No 138. York: Centre for Health Economics, University of York.
8. Torrance, G., Feeny, D., Furlong, W., Barr, R., Zhang, Y., & Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system: Health utilities index mark II. *Medical Care, 34*(7), 702–722.
9. 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*(2), 113–128.
10. 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.
11. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13*, 873–884.
12. Sintonen, H., & Pekurinen, M. (1989). A generic 15 dimensional measure of health-related quality of life (15D). *Journal of Social Medicine, 26*, 85–96.
13. 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.
14. Kaplan, R., Bush, J., & Berry, C. (1976). Health status: Types of validity and the index of wellbeing. *Health Services Research, 11*(4), 478–507.
15. Misajon, R., Hawthorne, G., Richardson, J., Barton, J., Peacock, S., Iezzi, A., & Keeffe, J. (2005). Vision and quality of life: The development of a utility measure. *Investigative Ophthalmology and Visual Science, 46*(11), 4007–4015.
16. Richardson, J., Iezzi, A., Peacock, S., Sinha, K., Misajon, R., & Keeffe, J. (2012). Utility weights for the vision related Assessment of Quality of Life (AQoL) 7D instrument. *Ophthalmic Epidemiology, 19*(3), 172–182.
17. Richardson, J., Elsworth, G., Iezzi, A., Khan, M. A., Mihalopoulos, C., Schweitzer, I., Herrman, H. (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.
18. Richardson, J., Sinah, K., Iezzi, A., Khan, M. A. (2014). Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. *Quality of Life Research, 23*, 2395–2404.
19. Chen, G., Khan, M. A., Iezzi, A., Ratcliffe, J., & Richardson, J. (2016). Mapping between 6 multi attribute utility instruments. *Medical Decsion Making, 36*(2), 160–175.
20. Richardson, J., Khan, M. A., Iezzi, A., Maxwell, A. (2015). Measuring the sensitivity and construct validity of six utility instruments in seven disease states. Medical Decision Making, Accepted 22 Sep 2015.
21. Richardson, J., Khan, M. A., Iezzi, A., & Maxwell, A. (2015). Comparing and explaining differences in the content, sensitivity and magnitude of incremental utilities predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB and AQoL-8D multi attribute utility instruments'. *Medical Decision Making, 35*(3), 276–291.
22. Richardson, J., Chen, G., Khan, M. A., & Iezzi, A. (2015). Can multi attribute utility instruments adequately account for subjective well-being? *Medical Decision Making*, *35*(3), 292–304. doi:10.1177/0272989X14567354
23. Campbell, J. A., Palmer, A. J., Venn, A., Sharman, M., Otahal, P., Neil, A. (2016). A head-to-head comparison of the EQ-5D-5L and AQoL-8D multi-attribute utility instruments in patients who have previously undergone bariatric surgery. *The Patient—Patient-Centered Outcomes Research, 2016* 1–12. doi:10.1007/s40271-015-0157-5
24. Hawthorne, G., Osborne, R., Sansoni, J., & Taylor, A. (2007). The SF-36 version 2: critical analysis of population weights, scoring algorithms and population norms. *Quality of Life Research, 16*(4), 661–673.
25. ABS. (1995). Australian Bureau of Statistics, National Health Survey SF-36 Population Norms Australia ABS Catalogue No. 4399. Canberra: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4399.01995?OpenDocument. Accessed 19 Apr 2012.
26. Slade, T., Johnston, A., Oakley Brown, M. A., Adnrews, G., & Whitefor, H. (2009). National survey of mental health and wellbeing: Methods and key findings. *Australian and New Zealand Journal of Psychiatry, 43*(7), 594–605.
27. Hawthorne, G., Herrman, H., & Murphy, B. (2006). Interpreting the WHOQoL-Brèf: Preliminary population norms and effect size. *Social Indicators Research, 77*, 37–59.
28. Cummins, R. A., Knapp, T. M., Woerner, J., Walter, J., Page, K. (2005). The personal Wellbeing of Australians living within federal electoral divisions. Report No: 13.1. Melbourne: Deakin University.
29. Norman, R., Church, J., van den Berg, B., & Goodall, S. (2013). Australian health-related quality of life population norms derived from the SF-6D. *Australian and New Zealand Journal of Public Health, 37*(1), 17–23.
30. Hawthorne, G., Korn, S., & Richardson, J. (2013). Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing. *Australian and New Zealand Journal of Public Health, 37*(1), 17–23.
31. Hawthorne, G., & Osborne, R. (2005). Population norms and meaningful differences for the Assessment of Quality of Life (AQoL) measure. *Australian and New Zealand Journal of Public Health, 29*(2), 136–142.
32. 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 12 Aug 2013.
33. Richardson, J., Iezzi, A., Khan, M. A., Chen, G. (2014). *Interim population norms for the AQoL-6D and AQoL-8D multi attribute utility instruments.* Research Paper 87. Melbourne: Centre for Health Economics, Monash University.
34. Meade, A. W., & Craig, B. S. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.
35. Gatz, D. F., & Smith, L. (1995). The standard error of a weighted mean concentration-I: Bootstrapping vs other methods. *Atmospheric Environment, 29*(11), 1185–1193.
36. AQoL. (2016) Assessment of Quality of Life (AQoL). http://www.aqol.com.au.

37. Hawthorne, G. (2009). Assessing utility where short measures are required: Development of the short assessment of quality of life-8 (AQoL-8) instrument. *Value in Health, 12*(6), 948–957.
38. Frijters, P., & Beatton, T. (2012). The mystery of the U-shaped relationship between happiness and age. *Journal of Economic Behavior and Organization, 82*, 525–542.
39. Richardson, J., Iezzi, A., & Khan, M. A. (2015). Why do multi attribute utility instruments produce different utilities: The relative importance of the descriptive systems, scale and 'micro utility' effects. *Quality of Life Research*. doi:10.1007/s11136-015-0926-6.
40. Iezzi, A., & Richardson, J. (2016). *A comparison of AQoL-4D, AQoL-6D, AQoL-7D and AQoL-8D multi attribute utility instruments*. Research Paper 93. Melbourne: Centre for Health Economics, Monash University.