CrossMark

ORIGINAL RESEARCH ARTICLE

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

Julie A. Campbell[1] · Andrew J. Palmer[1] · Alison Venn[1] · Melanie Sharman[1] · Petr Otahal[1] · Amanda Neil[1]

Abstract

Background Psychosocial health status is an important and dynamic outcome for bariatric/metabolic surgery patients, as acknowledged in recent international standardised outcomes reporting guidelines. Multi-attribute utility-instruments (MAUIs) capture and assess an individual's health-related quality-of-life (HRQoL) within a single valuation, their utility. Neither MAUIs nor utilities were discussed in the guidelines. Many MAUIs (e.g. EQ-5D) target physical health. Not so the AQoL-8D.

Objectives Our objective was to explore agreement between, and suitability of, the EQ-5D-5L and AQoL-8D for assessing health state utility, and to determine whether either MAUI could be preferentially recommended for metabolic/bariatric surgery patients.

Methods Utilities for post-surgical private-sector patients ($n = 33$) were assessed using both instruments and summary statistics expressed as mean [standard deviation (SD)] and median [interquartile range (IQR)]. Interchangeability of the MAUIs was assessed with Bland–Altman analysis. Discriminatory attributes were investigated through floor/ceiling effects and dimension-to-dimension comparisons. Spearman's rank measured associations between the instruments' utility values and with the body mass index (BMI).

Results Mean (SD) EQ-5D-5L utility value was 0.84 (0.15) and median 0.84 (IQR 0.75–1.00). Mean (SD) AQoL-8D utility value was 0.76 (0.17) and median 0.81 (IQR 0.63–0.88). Spearman's rank was $r = 0.68$; ($p < 0.001$); however, Bland–Altman analysis revealed fundamental differences. Neither instrument gave rise to floor effects. A ceiling effect was observed with the EQ-5D-5L, with 36 % of participants obtaining a utility value of 1.00 (perfect health). These same participants obtained a mean utility of 0.87 on the AQoL-8D, primarily driven by the mental-super-dimension score (0.52).

Conclusions The AQoL-8D preferentially captures psychosocial aspects of metabolic/bariatric surgery patients' HRQoL. We recommend the AQoL-8D as a preferred MAUI for these patients given their complex physical/psychosocial needs.

Key Points for Decision Makers

Psychosocial health status has been increasingly identified as an important health-related quality of life outcome measure for the morbidly obese population who receive bariatric surgery.

Compared with the EQ-5D-5L, the AQoL-8D's descriptive/classification system (and subsequent utility valuation) preferentially captures psychosocial health status for people who have received bariatric surgery.

While the EQ-5D dominates the clinical and economic evaluation literature, choice of multi-attribute utility instrument should be influenced by the innate sensitivities of the instrument to the relevant domains of health for the study population.

✉ Amanda Neil
Amanda.Neil@utas.edu.au

[1] Menzies Institute for Medical Research, University of Tasmania, 17 Liverpool Street, Hobart, TAS 7000, Australia

△ Adis

1 Introduction

Obesity is a worldwide problem. Its extensive health repercussions include a high prevalence of psychological comorbidities, and it also has substantial negative economic impacts [1]. Many clinical and epidemiological studies find the most efficacious therapy for morbid obesity is metabolic/bariatric surgery [2, 3]. A systematic review of the impact of bariatric surgery on health-related quality-of-life (HRQoL) found physical HRQoL was improved to a significantly greater degree than mental HRQoL [4]. Furthermore, the psychosocial health status of bariatric surgery patients is dynamic [5]. This recent study found an initial improvement in mental health followed by deterioration between 4 and 9 years post-surgery. Potential reasons for this diminution of HRQoL were postulated to include disappointment from unrealistic expectations about surgical treatment, unforseen changes in eating behaviour, medical sequelae after surgery, dissatisfaction with body appearance and excess skin, and the reoccurrence of psychiatric disorders [5, 6].

The need to assess the psychosocial health status of bariatric surgery patients in the short, medium and longer terms has been increasingly identified [4–7], and underpins the quality-of-life component of recent guidelines on standardised outcomes reporting for bariatric surgery patients from the American Society for Metabolic and Bariatric Surgeons (ASMBS) [8, 9]. The guidelines made no specific recommendations regarding the most appropriate HRQoL instrument, the recommendation being only to use a "validated instrument(s)". Importantly, the measurement of psychosocial health or any domain of health is wholly dependent on the sensitivity of the instrument employed to assess that domain.

Health state utility values (HSUVs), or utilities, are important health economic metrics that assess the strength of preference for an individual's health state relative to perfect health and death. Utilities are assessed relative to a 0.00–1.00 scale where 1.00 represents perfect health and 0.00 death [10]. The utility value therefore indicates the strength of preference for quality versus quantity of life [11], and quality-adjusted life-years (QALYs) can be calculated as the product of time spent in a health state and its utility. QALYs are a unit of benefit used in economic evaluation, namely cost-utility analysis (CUA) and, in principle, may be used to measure the HRQoL component of the burden of disease [10, 12]. Clinicians have also found that measuring health utilities is of benefit to patient–clinical assessment, relationships, communication and management [13]. Furthermore, utilities have been shown to be independent predictors of patient outcomes, including all-cause mortality and development of complications [14].

Multi-attribute utility instruments (MAUIs) are designed to rapidly and simply assess an individual's HSUV through application of pre-established formulae/weights to the array of responses to the MAUI's questions. Generic and disease-specific non-utility instruments may also be reduced to a single number; however, this number does not have independent meaning [10]. MAUIs thus differ fundamentally from generic HRQoL instruments.

Many MAUIs target physical health. For example, four of the five items in the EQ-5D, a well utilised international measure, relate to physical health. In contrast, 25 of the 35 items in the recently developed Assessment of Quality of Life (AQoL)-8D relate to psychosocial health [12]. Utility values assessed by MAUIs are not equivalent [15, 16], with the difference between the descriptive/classification systems of the MAUIs the principal determinant [15]. Differences in descriptive/classification systems are estimated to explain an average of 66 % of the difference between utilities obtained by MAUIs, and 81 % of the difference between the utilities of the EQ-5D-5L and AQoL-8D [15]. MAUIs are thus 'imprecisely related', a finding that threatens the comparability of economic evaluations that employ different instruments [17].

A small number of MAUIs dominate the economic evaluation literature [17]. A review of the Web of Science database (2005–2010) found that, of 1663 studies employing an MAUI, 63 % used the EQ-5D [15, 17]. Arguably, this finding reflects the recommendations of the UK National Institute for Health and Care Excellence (NICE) guidelines to use the EQ-5D as the preferred measure of HRQoL in adults [18]. These guidelines also acknowledge that the EQ-5D "may not be an appropriate measure of health-related utility in all circumstances" [19]. Emerging research is investigating the concept of 'bolt-on' dimensions to the EQ-5D in an attempt to broaden the classification system of this instrument [20, 21].

To inform debate on the choice of instrument for a particular patient group, it is important to compare different preference-based measures of health [22]. In particular, it is necessary to consider the applicability of the descriptive/classification systems. Our study investigated a 'head-to-head' cross-sectional comparison of the EQ-5D-5L [23] and AQoL-8D [24] MAUIs for patients who have previously undergone bariatric surgery. The EQ-5D-5L and AQoL-8D have not been specifically validated for patients who have undergone bariatric surgery. This study explored agreement between, and suitability of, the AQoL-8D and EQ-5D-5L for assessing health state utility in patients who have received bariatric surgery to determine whether either instrument could be preferentially recommended in this study population.

2 Materials and Methods

2.1 Participants

Participants were individuals who had previously received bariatric surgery [predominantly laparoscopic adjustable gastric band (LAGB)] in the private sector ($n = 33$) in Tasmania, Australia. Clinical and socio-demographic data were obtained during recruitment for a focus group designed to explore patient experiences following bariatric surgery. Participants were recruited with the aim of ensuring an appropriate mix of demographic/clinical characteristics. Each participant was sent both MAUIs for self-completion at home 2 weeks before their focus group [13, 25]. All data were de-identified. Questionnaire responses were independently entered into a database by two authors and cross-checked before utilities were generated. Ethics approval was granted by the University of Tasmania's Health and Medical and Social Sciences Human Research Ethics Committees.

2.2 Instruments

The EQ-5D-5L [23] is a recent augmentation of the EQ-5D-3L [26], and the AQoL-8D [27] is the latest in the suite of AQoL instruments (AQoL-4D/6D/7D/8D) [28]. Table 1 provides a detailed comparison of the characteristics of both instruments. The EQ-5D-5L was developed to address the limited sensitivity (lack of descriptive richness and serious ceiling effects [29]) of the EQ-5D-3L. The EQ-5D-5L includes two additional levels for each of the five dimensions in the EQ-5D [30]. Nevertheless, it has the second lowest number of health states of the major MAUIs at 5^5 (3125). The EQ-5D-3L has the lowest, at 243. The EQ-5D-5L retains an optional visual analogue scale (EQ-VAS) in which patients rate their current health state on a scale of 0–100 (worst/best imaginable) [31].

The AQoL-8D is the fourth and most comprehensive of the AQoL suite of instruments, developed to achieve increased sensitivity in psychosocial dimensions of health, which was relatively neglected in other MAUIs, including earlier versions of the AQoL [10]. Both patient and public involvement were utilised during the construction of the AQoL-8D, a key element of robust MAUI development according to a recent systematic review by Stevens [33]. Psychometric principles were also employed during construction of the AQoL instruments, the only MAUIs to do so [32]. These key features of the AQoL-8D were not identified in the Stevens [33] review. The AQoL-8D contains 35 questions and encompasses the largest number of health states of any existing MAUI (2.4×10^{23}).

2.3 Data Analysis

Baseline socio-demographic and clinical data are presented descriptively as mean [standard deviation (SD)] and/or median [interquartile range (IQR)] for continuous variables and frequency (%) for categorical variables. Body mass index (BMI) was calculated as weight (kg)/[height (m)]^2.

HSUVs were generated for the EQ-5D-5L using the UK 'crosswalk' value set with the EQ-5D-5L version mapped (crosswalked) onto the 3L version through the preferred non-parametric model [34]. For the AQoL-8D, we used a scoring algorithm incorporating Australian weights published on the AQoL group's website (http://www.aqol.com.au). We assessed questionnaire completion by measuring the proportion of participants who completed the questionnaire and for whom an individual utility value could be generated.

Summary statistics of the HSUVs for each MAUI were assessed as mean (SD) and median (IQR) given the skewed nature of the data. Strength of correlation between the instruments' utility values for the sample was tested using Spearman's correlation coefficient, with Spearman's rho of greater than 0.50 or less than -0.50 considered strong, values between -0.49 to 0.30 and 0.30 to 0.49 considered moderate; and between -0.30 and 0.30 weak [35]. To determine interchangeability between the instruments, pairwise agreement between the utility values for each instrument for each participant was assessed using a scatterplot and through the Bland–Altman (BA) method of differences [36]. The difference between the two measures was plotted against the mean measurement for those two instruments for each individual, along with the limits of agreement (the range of values that would be expected to include 95 % of individual differences) [31].

An MAUI should be able to produce utility valuations for various health states with a significant degree of accuracy to effectively detect and represent differences between individuals [31]. Discriminatory attributes of the instruments were therefore assessed globally and then at dimensional levels. Globally, the extent of floor (worst health: -0.594 EQ-5D-5L and $+0.09$ AQoL-8D) and ceiling effects (perfect health: 1.0 each instrument) was determined, and then utility values obtained on the alternate instrument were explored. At the dimensional level, summary statistics were obtained for the summary scores for each individual dimension of the AQoL-8D and its super-dimensions. The distribution of responses across the levels (1–5 or 6) of each of three psychosocial-related dimensions within each instrument was then explored. These dimension-to-dimension comparisons [22] encompassed anxiety/depression, self-care and pain/discomfort for the EQ-5D-5L, comprising one item each; and mental

Table 1 Comparison of the dimensions and content of EQ-5D-5L and AQoL-8D multi-attribute utility instruments

Characteristics	EQ-5D-5L	AQoL-8D
Development team and year finalised	EuroQol Research Foundation, EuroQol Executive Office, Rotterdam, The Netherlands. (http://www.euroqol.org), 2011	Centre for Health Economics, Monash University, Melbourne, Victoria, Australia. (http://www.aqol.com.au), 2011
Number of health states described	3125	2.4×10^{23}
Number of pages and questions	Two pages: EQ-5D-5L descriptive system five questions and EQ-5D VAS	Three pages: AQoL-1 to AQoL-35
Mean completion time	<1 min	5.45 min [27]
Valuation technique	Non-parametric indirect modelling technique and UK 'crosswalk' set [34]	Time trade-off (Australian sample) [27]
Total number of dimensions	Five dimensions, one item in each. Each item has five levels of severity scored as 1 (best) to 5 (worst). Participants choose one level for each item to describe their health status on the day of interview. Provides health state values as 11111 (best) to 55555 (worst)	Eight dimensions of between three and eight items; 35 items in total. Each item (AQoL-1 to AQoL-35) has four to six levels, scored as 1 (best) to 6 (worst). Participants choose one level for each item to describe health status over the past week
Number of dimensions of physical health	Four dimensions: mobility, self-care, usual activities and pain/discomfort	Three dimensions: (1) independent living, four items (household tasks, getting around, mobility, self-care); (2) senses, three items (vision, hearing, communication); and (3) pain, three items (frequency of pain, degree of pain, pain interference) [10]
Number of dimensions of psychosocial health	One dimension, anxiety/depression, with five levels of severity: (1) I am not anxious or depressed (2) I am slightly anxious or depressed (3) I am moderately anxious or depressed (4) I am severely anxious or depressed (5) I am extremely anxious or depressed	Five dimensions: (4) happiness, four items (contentment, enthusiasm, degree of feeling happiness, pleasure); (5) coping, three items (energy, being in control, coping with problems); (6) relationships, seven items (relationship with family and friends, social exclusion, intimate relationship, family and community role); (7) self-worth, three items (feeling like a burden, worthlessness, confidence); (8) mental health, eight items (feelings of depression, trouble sleeping, feelings of anger, self-harm, feeling despair, worry, sadness, tranquillity/agitation) [10]
Super-dimensions of physical and psychosocial health	No super-dimensions	Two super-dimensions: PSD and MSD. PSD includes independent living, senses and pain; MSD includes happiness, coping, relationships, self-worth and mental health [10]
VAS	Yes	No
Range of reported utility scores from best possible health to worst possible health state	Range: states worse than dead <0 to full health 1.00, anchoring at dead = 0. Scored range: −0.594 to 1.00 (UK crosswalk value set)	Range: full health 1.00 to anchoring at dead = 0. Scored range: +0.09 to 1.00
Available population norms	Ten countries in the crosswalk value set from EQ-5D-3L: Denmark, France, Germany, Japan, Netherlands, Spain, Thailand, UK, USA, Zimbabwe [34]. Australia [50]	Australia

MSD mental super-dimension, *PSD* physical super-dimension, *VAS* visual analogue scale

health, independent living and pain for the AQoL-8D, comprising eight, four and three items each, respectively.

The association between 'current BMI' and utility valuation obtained with each instrument was investigated by testing strength of correlation using the Spearman's correlation coefficient.

Statistical analyses were undertaken using IBM® SPSS® (version 22) or R (version 3.0.2).

3 Results

3.1 Participants' Clinical and Socio-Demographic Characteristics

Table 2 provides the participants' clinical and socio-demographic characteristics. Mean (SD) age was 56 (11) years, and two-thirds ($n = 22$; 67 %) were female. Mean (SD) of the maximum recorded BMI (before surgery) was 43.7 (7.3) kg/m^2, and mean (SD) current BMI (at recruitment) was 32.8 (7.7) kg/m^2. One-third of participants had obtained university qualifications and one-quarter were educated to ≤year 10. Most participants ($n = 32$; 97 %) had received an LAGB, and 12 % ($n = 4$) of these participants had undergone a secondary procedure such as a revision. Median (IQR) number of years since primary surgery was 5.0 (3.0–8.0).

Table 2 Baseline socio-demographic characteristics of participants

Characteristics	Private pts after surgery ($n = 33$)
Age, years	56 ± 11
Sex	
Male	11 (33)
Female	22 (67)
BMI (kg/m^2)	
Maximum (before surgery)	43.7 ± 7.3
Current (at recruitment)	32.8 ± 7.7
Years since primary procedure	6 ± 6
Median (IQR)	5 (3–8)
Highest level of education (%)	
Category 1: year 10 or less	24.5
Category 2: year 11 and/or 12	15
Category 3: certificate, diploma or trade	27.5
Category 4: university	33
LAGB	32 (97)
Laparoscopic	33 (100)
Secondary procedure	4 (12)

Data are presented as mean ± SD or n (%) unless otherwise indicated

BMI body mass index, *IQR* interquartile range, *LAGB* laparoscopic adjustable gastric band, *pts* patients, *SD* standard deviation

3.2 Questionnaire Practicality

All participants completed both MAUIs. The EQ-5D-5L was completed without omissions or additions (such as multiple responses to one question). In contrast, one participant attempted to select two response items to two questions and modify those items when completing the AQoL-8D. These nonconformities had no impact on our ability to assess the utility of this participant. As advised by the AQoL group, we used the worst response for utility generation.

3.3 Construct Validity

Frequency distributions of the individual utility values for both instruments are provided in Fig. 1a, b. Utilities obtained through both MAUIs showed a distribution towards perfect health, more so for the EQ-5D-5L than the AQoL-8D. There was no significant difference in mean and median utility values of both instruments, with a strong correlation overall. The range and IQR for the EQ-5D-5L (0.40–1.00 and 0.75–1.00) and the AQoL-8D (0.35–0.95 and 0.63–0.88) were the same (Table 3), but each was higher for the EQ-5D-5L, reflecting its greater negative skew. In turn, the AQoL-8D's assessed range and IQR for our study population compared with the potential scored range, measured as the difference between the floor to ceiling levels of +0.09 to 1.00, is proportionally larger than for the equivalent measure of the EQ-5D-5L. The inclusion of 1.00 in the EQ-5D-5L's range and IQR also reflect the ceiling effects of this instrument within our study population as detailed below.

The mean (SD) and median utility values tended to be higher [0.84 (0.15); 0.84] for the EQ-5D-5L than for the AQoL-8D [0.76 (0.17); 0.81], respectively (Table 3). A strong correlation was obtained between the utilities for the EQ-5D-5L and AQoL-8D (Spearman's rho 0.68; $p < 0.001$). The EQ-VAS gave rise to mean (SD) and median (IQR) ratings of 76 (17) and 80 (70–90), respectively.

3.4 Sensitivity

A scatterplot of individual utility values (Fig. 1c) demonstrated two distinct groupings around 0.8 and 1.0 for the EQ-5D-5L. The BA plot (Fig. 1d) revealed a relatively wide limit of agreement (0.55) and systematic variation, notably a negative trend in the difference between individual participant utility values by mean value. No floor effects were identified for either instrument, nor were there ceiling effects for the AQoL-8D (Table 3). However, a ceiling effect was observed for over one-third ($n = 12$; 36 %) of participants with the EQ-5D-5L.
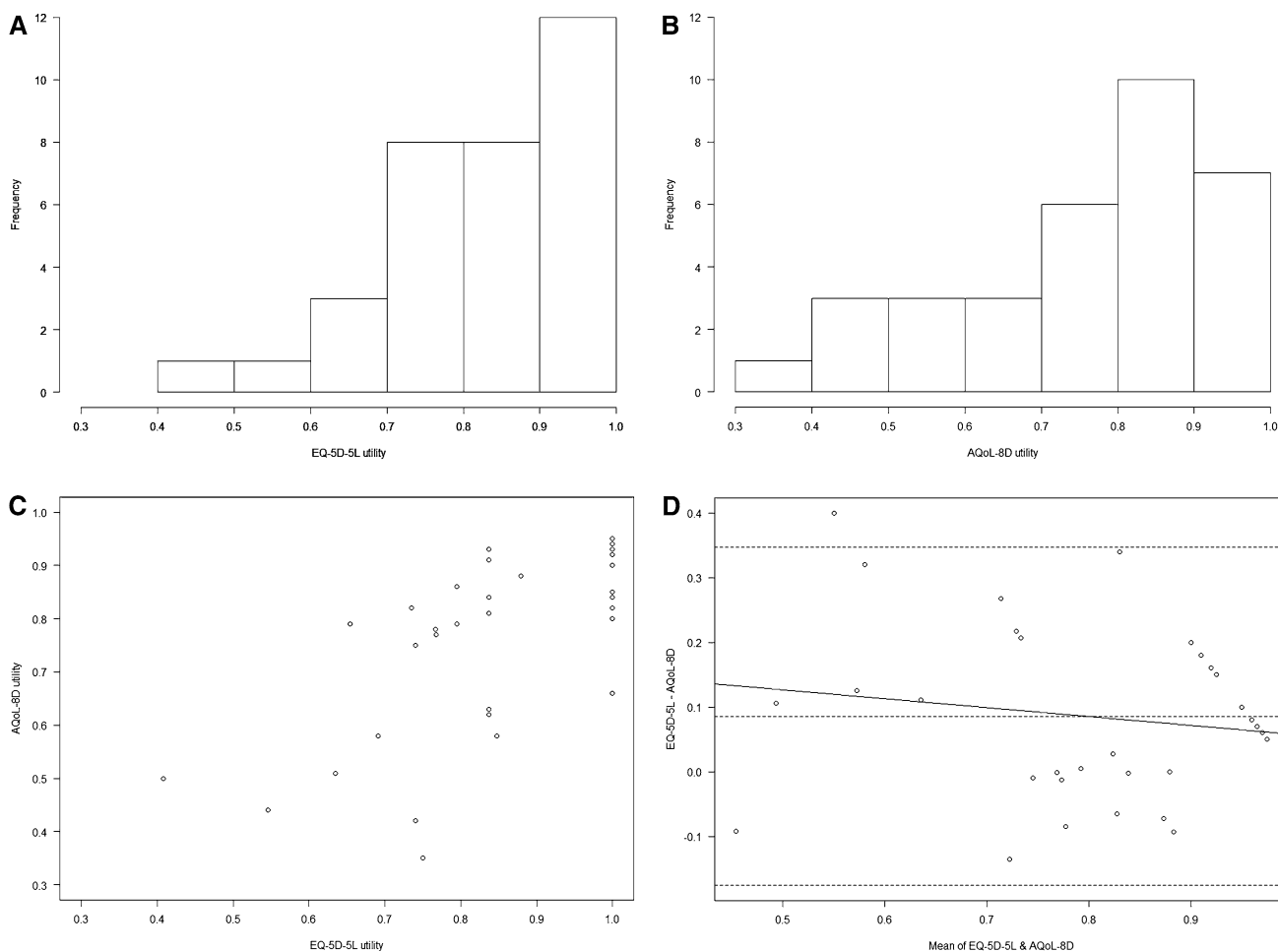
Fig. 1 Distribution of **a** EQ-5D-5L utility scores and **b** AQoL-8D utility scores. **c** Scatterplot of participants' utility scores for EQ-5D-5L and AQoL-8D. **d** Bland–Altman method of differences for utility scores between the EQ-5D-5L and AQoL-8D, all participants ($n = 33$)

Table 3 Descriptive statistics of EQ-5D-5L and AQoL-8D utility valuations, EQ-VAS scores and percent achieving worst and best health states

MAUI ($n = 33$)	Mean (SD)	Median (IQR)	Minimum	% on floor	Maximum	% on ceiling
EQ-5D-5L	0.84 (0.15)	0.84 (0.75–1.00)	0.40	0	1.00	36 % ($n = 12$)
AQoL-8D	0.76 (0.17)	0.81 (0.63–0.88)	0.35	0	0.95	0
EQ-VAS	76 (17)	80 (70–90)	30	0	100	8 % ($n = 1$)

IQR interquartile range, *MAUI* multi-attribute utility instrument, *SD* standard deviation, *VAS* visual analogue scale

Table 4 provides the EQ-VAS rating scores and AQoL-8D global utility values for each participant scoring perfect health using the EQ-5D-5L. One participant (number 11) rated themselves as experiencing perfect health on the EQ-VAS; however, their AQoL-8D utility valuation was high but not perfect (0.93). Overall, the mean (SD) and median (IQR) EQ-VAS ratings were 83 (10) and 84 (79–90), and the mean (SD) and median (IQR) AQoL-8D utility values were 0.87 (0.08) and 0.88 (0.84–0.93). Table 5 provides summary statistics for the individual and super-dimension scores of the AQoL-8D for the entire sample. The maximum score for the mental health dimension at 0.73 was markedly lower than for all other dimensions. The maximum score in the other seven individual dimensions was at least 0.96, six scoring 1.00. In turn, the maximum mental super-dimension score was 0.79. The mental health and mental super-dimensions also recorded the lowest mean (SD) and median (IQR) scores, at 0.62 (0.12) and 0.63 (0.52–0.73) and 0.44 (0.17) and 0.45 (0.27–0.54), respectively. Table 6 provides AQoL-8D individual and super-

Table 4 EQ-VAS rating and AQoL-8D utility valuation for each individual assessed in perfect health through the EQ-5D-5L

Participant	EQ-5D-5L utility value	EQ-VAS rating	AQoL-8D utility value
1	1.00	85	0.82
2	1.00	80	0.95
3	1.00	90	0.90
4	1.00	80	0.84
5	1.00	65	0.85
6	1.00	90	0.95
7	1.00	83	0.94
8	1.00	90	0.84
9	1.00	75	0.80
10	1.00	90	0.92
11	1.00	100	0.93
12	1.00	70	0.66
Mean (SD)	1.0	83 (10)	0.87 (0.08)
Median (IQR)	1.0	84 (79–90)	0.88 (0.84–0.93)

IQR interquartile range, *SD* standard deviation, *VAS* visual analogue scale

Table 5 AQoL-8D individual and super dimensions scores for the entire sample ($n = 33$)

AQoL-8D dimension or super-dimension ($n = 33$)	Mean (SD)	Median (IQR)	Minimum	Maximum
Dimensions of physical health (3)				
Independent living	0.90 (0.14)	0.96 (0.85–1.00)	0.51	1.00
Senses	0.88 (0.10)	0.85 (0.82–0.95)	0.63	1.00
Pain	0.78 (0.20)	0.80 (0.63–0.95)	0.29	1.00
Dimensions of psychosocial health (5)				
Happiness	0.81 (0.11)	0.85 (0.75–0.89)	0.50	0.97
Coping	0.82 (0.13)	0.87 (0.77–0.90)	0.39	0.96
Relationships	0.77 (0.16)	0.82 (0.60–0.89)	0.49	1.00
Self-worth	0.83 (0.14)	0.86 (0.75–0.92)	0.46	1.00
Mental health	0.62 (0.12)	0.63 (0.52–0.73)	0.37	0.79
Super-dimensions (2)				
PSD	0.73 (0.19)	0.75 (0.63–0.87)	0.34	1.00
MSD	0.44 (0.17)	0.45 (0.27–0.54)	0.10	0.73

IQR interquartile range, *MSD* mental super-dimension, *PSD* physical super-dimension, *SD* standard deviation

dimension scores for those recording perfect health using the EQ-5D-5L. One of these participants (Table 4, participant 11) achieved the maximum score (1.00) for the physical super-dimension (PSD). Given their AQoL-8D utility valuation was 0.93, this participant's overall health status was diminished due to psychosocial impacts. The maximum mental super-dimension (MSD) score within this group was 0.71. The mean (SD) scores for the AQoL-8D PSD and MSD were 0.89 (0.07) and 0.52 (0.13), respectively. The findings are also reflected at the individual dimensions level of physical and psychosocial health. The physical health dimensions gave rise to the highest scores [independent living 0.97 (0.04), senses 0.92 (0.06), pain 0.95 (0.09)], and the psychosocial dimensions the lowest

scores [happiness 0.85 (0.07), coping 0.87 (0.08), relationships 0.85 (0.12), self-worth 0.90 (0.08), mental health 0.65 (0.09)].

Table 7 provides a dimension-to-dimension comparison for each of three individual psychosocial-related dimensions of the EQ-5D-5L and AQoL-8D. The EQ-5D-5L showed a larger proportion of participants at Level 1 than the AQoL-8D for each dimension, and less dispersion overall. There were no participants rated at Level 4 or above within the psychosocial dimensions for the EQ-5D, unlike the AQoL-8D.

A moderate association was found between 'current BMI' and utility valuations for both the EQ-5D-5L and AQoL-8D with Spearman's rho -0.37; $p = 0.03$ and -0.39; $p = 0.02$, respectively.

Table 6 AQoL-8D individual dimension and super-dimension scores for each individual assessed in perfect health through the EQ-5D-5L

AQoL-8D dimension or super-dimension ($n = 12$)	Mean (SD)	Minimum	Maximum
Dimensions of physical health (3)			
Independent living	0.97 (0.04)	0.90	1.00
Senses	0.92 (0.06)	0.84	1.00
Pain	0.95 (0.09)	0.72	1.00
Dimensions of psychosocial health (5)			
Happiness	0.85 (0.07)	0.73	0.97
Coping	0.87 (0.08)	0.71	0.96
Relationships	0.85 (0.12)	0.66	1.00
Self-worth	0.90 (0.08)	0.70	1.00
Mental health	0.65 (0.09)	0.43	0.78
Super-dimensions (2)			
PSD	0.89 (0.07)	0.80	1.00
MSD	0.52 (0.13)	0.27	0.71

MSD mental super-dimension, *PSD* physical super-dimension, *SD* standard deviation

Table 7 Distribution of levels of response for EQ-5D-5L individual dimensions of anxiety/depression, self-care and pain/discomfort with the AQoL-8D individual dimensions of mental health, independent living and pain

EQ-5D-5L				AQoL-8D		
Dimension	Anxiety/depression (1)	Self-care (2)	Pain/discomfort (3)	Mental health (1)	Independent living (2)	Pain (3)
Level						
1	64 %	97 %	49 %	24 %	63 %	43 %
2	24 %	3 %	30 %	39 %	24 %	39 %
3	12 %	0	18 %	30 %	8 %	15 %
4	0	0	3 %	5 %	5 %	3 %
5	0	0	0	2 %	0	0
6	NA	NA	NA	0	0	NA

All columns add to 100 %

NA not applicable

4 Discussion

To the best of our knowledge, our study is the first to investigate a 'head-to-head' comparison of the EQ-5D-5L and AQoL-8D MAUIs in patients who have undergone bariatric surgery. Our study's key finding was the divergent sensitivity of the instruments in assessing health state utility in this patient group, a difference arguably due to their ability to assess and capture psychosocial HRQoL impacts. This finding is crucial because psychosocial health status has been identified as a significant outcome for the morbidly obese population who receive bariatric surgery [4–6, 8, 9].

We found 36 % of participants were assessed as having perfect health on the EQ-5D-5L, but none on the AQoL-8D. The mean utility valuation of the patient group scoring perfect health on the EQ-5D-5L was 0.87 using the AQoL-8D, the lower utility driven by less than perfect scores on

the AQoL-8D MSD and, in all but one instance, the PSD. The assessed range for the EQ-5D-5L as a proportion of the potential scored range was less than for the AQoL-8D at 16 % [0.25/(1 − (−0.594))] and 27 % [0.25/(1 − 0.09)], respectively, indicating greater discriminatory attributes of the latter for this study population. These findings are underpinned by differences in the classification/descriptive systems and scoring algorithms of the two instruments.

EQ-5D-5L and AQoL-8D utility values were highly correlated by rank ordering (Spearman's rho 0.68); however, high correlation does not imply close agreement and is blind to the possibility of systematic bias [36]. We observed pairwise disagreement in utility values assessed for a given individual and evidence of systematic bias. In turn, the utility valuations obtained with these instruments in the population who underwent bariatric surgery are non-interchangeable. Our finding of non-interchangeability between the EQ-5D-5L and the AQoL-8D is consistent

with a lack of pairwise agreement between the EQ-5D-3L and the AQoL-4D [25].

One of the key drivers for the development of the EQ-5D-5L was to address serious ceiling effects of the EQ-5D-3L [23], with over 45 % of participants scoring perfect health in some studies [37, 38]. The severe ceiling effects of the EQ-5D-3L reflected difficulties in its ability to measure small and medium changes in health [23]. In an investigation of the EQ-5D-5L compared with the EQ-5D-3L across eight patient groups, the ceiling effect was reduced from 20 % (EQ-5D-3L) to 16 % (EQ-5D-5L), on average. Importantly, this study found that ceiling effects were higher for chronic diseases such as diabetes. In this population, the ceiling effect reduced from 34 % (EQ-5D-3L) to 28 % (EQ-5D-5L). In contrast, the ceiling effects for depression were reduced from 12 % (EQ-5D-3L) to 6 % (EQ-5D-5L) [30]. Arguably, this is a direct reflection of the specific question on depression/anxiety in the EQ-5D and underpins the importance of the descriptive systems employed.

Whilst floor/ceiling effects were not investigated in studies of bariatric surgery patients that employed the EQ-5D-3L [39–41], over one-third of participants reported perfect health on the EQ-5D-5L in our study. This is a finding comparable to the extent of ceiling effects reported in recent studies that used the EQ-5D-5L for chronic conditions, including diabetes ($n = 117$ [42] and $n = 289$ [43]), end-stage renal disease ($n = 150$ [44]), and chronic hepatic disease ($n = 1088$ [45]), and consistent with the comparative findings above.

The ongoing ceiling effects measured in this and other studies indicate the limitations of the breadth of the EQ-5D. Furthermore, research concerning the development of 'bolt-on' items for the EQ-5D has argued that these items could facilitate greater sensitivity for specific conditions, and further research has been encouraged [20]. However, it has also been noted that the use of 'bolt-on' items may lead to "some variations in measurement between conditions and detract from the advantages of using a generic instrument" [20]. We postulate that inclusion of one or more 'bolt-on' items may render results non-interchangeable, even with other 'EQ-5D' analyses and, in turn, the current dominance of this instrument irrelevant.

We found the mean, median and maximum scores of the AQoL-8D mental health and MSD were low relative to other AQoL-8D dimension scores for both the entire sample and the ceiling effect's subgroup for the EQ-5D. We also found greater dispersion for the AQoL-8D than the EQ-5D-5L across the three most comparable individual dimensions potentially impacting psychosocial health. We contend that together these findings support the greater sensitivity of the AQoL-8D than the EQ-5D towards psychosocial health.

In regard to the moderate correlations observed between utilities obtained from each instrument and 'current BMI', we contend that this finding is reflective of weight status being just one factor contributing to the HRQoL of people who have received bariatric surgery. This position is consistent with the most recent evidence, which does not support a direct link between long-term weight reduction and continued improvement/decline in mental health after bariatric surgery [5, 6]. Psychosocial support, alongside weight loss maintenance, are important management components for the HRQoL of this group of individuals in the longer term.

Economic evaluations of interventions that affect HRQoL commonly employ CUA that prioritise interventions according to the costs per QALY gained [15]. We found that significant differences in the EQ-5D-5L and AQoL-8D descriptive systems impact their sensitivity towards psychosocial domains of health. We also found that the utility values obtained cannot be used interchangeably. Impacts on psychosocial health for bariatric surgery patients have been identified as a vital outcome. Our findings thus have implications for the choice of utility instrument employed for clinical assessment and/or economic evaluation in the population for whom bariatric surgery is a consideration.

As noted previously, NICE's recommendation to use the EQ-5D for utility assessment is tempered by whether use of the EQ-5D is considered appropriate; a lack of content validity, including missing key health dimensions, is a primary concern [18]. If the nominated choice of instrument lacks sensitivity within a particular health context (or health domain), interventions affecting health states where the instrument's sensitivity is low will be disadvantaged [32], a potential bias of particular importance for healthcare decision makers. For people who are morbidly obese considering or having undergone bariatric surgery, the impact of any intervention will not be fully captured unless the nominated MAUI is sensitive to psychosocial health.

In turn, while the EQ-5D dominates the clinical/economic evaluation literature, its prevalence should not influence the choice of instrument in this (or other) study population(s). Rather, the choice of MAUI should be influenced by the sensitivity of the instrument to a patient group's health profile. In turn, we argue that the AQoL-8D should be preferred to the EQ-5D-5L within the morbidly obese population, including those undergoing or having undergone bariatric surgery, given its sensitivity to the psychosocial dimension of HRQoL.

Within the ASMBS's recently published outcomes reporting guidelines for bariatric and metabolic surgery [8, 9], the EQ-5D was classified as one of several frequently used generic HRQoL instruments within this population; however, the ASMBS was unable to provide specific

guidance as to a preferred HRQoL instrument(s), as previously noted. No reference was made to MAUIs per se; a situation we believe is an important oversight. If a MAUI and associated utility valuation comprehensively assesses and captures the physical and psychosocial domains of health for bariatric surgery patients, use of such an instrument could fulfil ASMBS HRQoL requirements. Related economic evaluations would also be underpinned by robust utility valuation, and thus facilitate defensible resource allocation.

Respondent burden is also a necessary consideration in instrument choice. The ASMBS document argues that HRQoL instruments with more items are less likely to be completed by patients, whereas instruments with fewer items are completed at higher rates. We expected the EQ-5D-5L would achieve a higher level of completion given that it comprises 30 fewer items than the AQoL-8D. Additionally, the average time for completion for the EQ-5D-5L (1 min), is approximately 4 min faster than that for the AQoL-8D (Table 1). Our study showed a 100 % response rate for both instruments and subsequent generation of individual utility values. Nevertheless, we acknowledge that our study participants were fully engaged through focus group involvement and that this may have influenced the completion rate of the MAUIs for our study. As participant levels of education were relatively evenly spread, this should not confound questionnaire completion.

The ASMBS document also recommends, with reference to a 2011 review of HRQoL instruments measuring bariatric surgery [46], the use of a combination of HRQoL instruments to capture psychosocial impacts. The 2011 review found that while several generic and obesity-specific instruments have been developed and/or used in bariatric surgery, all have limitations [46]. The review investigated the content validity of one MAUI (EQ-5D) and other generic and disease-specific instruments, including the SF-36, Nottingham Health Profile and IWQoL-lite. The review consequently proposed a conceptual framework for a bariatric surgery-specific HRQoL instrument that comprised 20 items, 19 of which, including all of the psychosocial domains of health, are included in the AQoL-8D. The item not included in the AQoL-8D pertained to eating. This conceptual framework subsequently underpinned the development of the disease-specific quality-of-life instrument, the 'bariatric and obesity-specific survey' (BOSS) [47]. The BOSS is not an MAUI. The BOSS-42 (the final version of this instrument) contains 42 items, seven more than the AQoL-8D.

Thus, our study found that a single MAUI instrument, the AQoL-8D, is sensitive to the psychosocial as well as the physical domains of health for people who have undergone bariatric surgery, and it captures the vast majority of domains considered crucial in this population.

While the length of the AQoL-8D may be an initial deterrent, this concern must be balanced against the sensitivity of this instrument to mental health [48] and physical health dimensions. Further, the use of a combination of up to three or four HRQoL instruments could be more burdensome and time consuming for the study population than the use of a single comprehensive instrument.

The major strength of this study is the use of a homogeneous group of bariatric surgery patients to minimise confounding due to patient characteristics in the identification of similarities and key differences between the EQ-5D-5L and AQoL-8D. The key limitation of this study is the sample size ($n = 33$). Nevertheless, we found that about one-third of the participants scored perfect health on the EQ-5D-5L, which is consistent with other studies of chronic disease with larger samples. Another limitation is that we did not include a disease-specific instrument because of concerns about the potential impact of respondent burden on both the quantitative and the qualitative components of the broader study. In lieu of a disease-specific instrument, we compared the utility valuations of 'current BMI'. One further limitation could be attributed to the utilities estimated from the EQ-5D-5L crosswalk value set [49]. Given this study was exploratory, larger confirmatory studies are justified. We also suggest that a comparison between the AQoL-8D and SF-6D would be of value.

5 Conclusions

Before selecting a generic MAUI, researchers should fully understand the instruments' descriptive/classification systems and the innate sensitivities of the MAUI in their context. Given the relative importance of the psychosocial health in the population contemplating or having undergone bariatric surgery, the choice of MAUI may be crucial. For bariatric surgery, the AQoL-8D more fully captured and assessed the psychosocial aspects of these patients' HRQoL as compared with the EQ-5D-5L. Additionally, the AQoL-8D was sensitive to the physical aspects of these patients' HRQoL. We recommend the AQoL-8D as a preferred MAUI to the EQ-5D-5L for patients undergoing bariatric surgery given their complex physical and psychosocial needs.

References

1. Gortmaker SL, Swinburn BA, Levy D, Carter R, Mabry PL, Finegood DT, et al. Changing the future of obesity: science, policy, and action. Lancet. 2011;378(9793):838–47.

2. Angrisani L, Santonicola A, Iovino P, Formisano G, Buchwald H, Scopinaro N. Bariatric surgery worldwide 2013. Obes Surg. 2015;25(10):1822–32.

3. Buchwald H, Oien DM. Metabolic/bariatric surgery worldwide 2011. Obes Surg. 2013;23(4):427–36.

4. Lindekilde N, Gladstone BP, Lübeck M, Nielsen J, Clausen L, Vach W, et al. The impact of bariatric surgery on quality of life: a systematic review and meta-analysis. Obes Rev. 2015;16(8):639–51.

5. Herpertz S, Müller A, Burgmer R, Crosby RD, de Zwaan M, Legenbauer T. Health-related quality of life and psychological functioning 9 years after restrictive surgical treatment for obesity. Surg Obes Relat Dis. 2015;11(6):1361–70. doi:10.1016/j.soard.04.008.

6. Karmali S. The impact of bariatric surgery on psychological health. J Obes. 2013. doi:10.1155/2013/837989 **(Epub 28 Mar 2013)**.

7. Burgmer R, Legenbauer T, Müller A, de Zwaan M, Fischer C, Herpertz S. Psychological outcome 4 years after restrictive bariatric surgery. Obes Surg. 2014;24(10):1670–8.

8. Brethauer SA, Kim J, El Chaar M, Papasavas P, Eisenberg D, Rogers A, et al. Standardized outcomes reporting in metabolic and bariatric surgery. Surg Obes Rel Dis. 2015;11(3):489–506.

9. Brethauer S, Kim J, el Chaar M, Papasavas P, Eisenberg D, Rogers A, et al. Standardized outcomes reporting in metabolic and bariatric surgery. Obes Surg. 2015;25(4):587–606.

10. Khan MA, Richardson J, O'Brien P. The effect of obesity upon health related quality of life (HRQoL): a comparison of the AQoL-8D and SF-36 instruments. Farmeconomia Health Economics Therapeutic Pathways. 2012;13(2):69–82.

11. Drummond MF, Sculpher M, Torrance G, O'Brien B, Stoddart G. Methods for the economic evaluation of health care programmes. 3rd ed. New York: Oxford University Press; 2005.

12. Richardson J, Iezzi A, Khan MA, Maxwell A. Validity and reliability of the Assessment of Quality of Life (AQoL)-8D multi-attribute utility instrument. Patient. 2014;7(1):85–96.

13. Skinner EH, Denehy L, Warrillow S, Hawthorne G. Comparison of the measurement properties of the AQoL and SF-6D in critical illness. Critic Care Resusc. 2013;15(3):205.

14. Clarke PM, Hayes AJ, Glasziou PG, Scott R, Simes J, Keech AC. Using the EQ-5D index score as a predictor of outcomes in patients with type 2 diabetes. Med Care. 2009;47(1):61–8.

15. Richardson J, Iezzi A, Khan MA. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility'effects. Qual Life Res. 2015;24:1–9.

16. Chen G, Iezzi A, McKie J, Khan MA, Richardson J. Diabetes and quality of life: comparing results from utility instruments and Diabetes-39. Diabetes Res Clin Pract. 2015;109(2):326–33. doi:10.1016/j.diabres.2015.05.011.

17. Chen G, Khan MA, Iezzi A, Ratcliffe J, Richardson J. Mapping between 6 multiattribute utility instruments. Med Decis Making. 2015. doi:10.1177/0272989x15578127 **(Epub 3 Apr 2015)**.

18. National Institute for Health and Care Excellence. Process and methods guides: guide to the methods for technology appraisal 2013. London: NICE; 2013. https://www.nice.org.uk/article/pmg9/resources/non-guidance-guide-to-the-methods-of-technology-appraisal-2013-pdf. Accessed 15 June 2015.

19. Richardson J, McKie J, Bariola E. Review and critique of health related multi attribute utility instruments. Melbourne: Monash University, Business and Economics, Centre for Health Economics; 2011.

20. Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. Value Health. 2015;18(1):52–60.

21. Lin F-J, Longworth L, Pickard AS. Evaluation of content on EQ-5D as compared to disease-specific utility measures. Qual Life Res. 2013;22(4):853–74.

22. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ. 2004;13(9):873–84.

23. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.

24. Richardson. J. Increasing the sensitivity of the AQoL inventory for the evaluation of interventions affecting mental health. Melbourne: Monash University, Centre for Health Economics; 2011.

25. Holland R, Smith RD, Harvey I, Swift L, Lenaghan E. Assessing quality of life in the elderly: a direct comparison of the EQ-5D and AQoL. Health Econ. 2004;13(8):793–805.

26. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.

27. Richardson J, Sinha K, Iezzi A, Khan M. Modelling utility weights for the Assessment of Quality of Life (AQoL)-8D. Qual Life Res. 2014;23(8):2395–404.

28. Richardson J, Atherton Day N, 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.

29. Jia Y, Cui F, Li L, Zhang D, Zhang G, Wang F, et al. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual Life Res. 2014;23(8):2355–63.

30. Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22(7):1717–27.

31. Turner N, Campbell J, Peters TJ, Wiles N, Hollinghurst S. A comparison of four different approaches to measuring health utility in depressed patients. Health Qual Life Outcomes. 2013;11:81.

32. Culyer AJ. Encyclopaedia of Health Economics. 1st ed. : vol. 2. 1st ed. Oxford: Elsevier Science; 2014. p. 353.

33. Stevens KJ. How well do the generic multi-attribute utility iIncorporate patient and public views into their descriptive systems? Patient. 2015. doi:10.1007/s40271-015-0119-y (Epub 8 Feb 2015).

34. van Hout B, Janssen M, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15(5):708–15.

35. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med. 2005;60(7):1571–82.

36. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. Ultrasound Obstet Gynecol. 2003;22(1):85–93.

37. Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. Med Care. 2005;43(7):736–49.

38. Cunillera O, Tresserras R, Rajmil L, Vilagut G, Brugulat P, Herdman M, et al. Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. Qual Life Res. 2010;19(6):853–64. doi:10.1007/s11136-010-9639-z.

39. Ribaric G, Buchwald JN, d'Orsay G, Daoud F. 3-year real-world outcomes with the Swedish adjustable gastric band in France. Obes Surg. 2013;23(2):184–96.

40. Ackroyd R, Mouiel J, Chevallier JM, Daoud F. Cost-effectiveness and budget impact of obesity surgery in patients with type-2 diabetes in three European countries. Obes Surg. 2006;16(11):1488–503.

41. Lin VW, Wong ES, Wright A, Flum DR, Garrison LP Jr, Alfonso-Cristancho R, et al. Association between health-related quality of life and body mass after adjustable gastric banding: a nonlinear approach. Value Health. 2013;16(5):823–9.

42. Pattanaphesaj J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. Health Qual Life Outcomes. 2015;13(1):14. doi:10.1186/s12955-014-0203-3.

43. Pan C-W, Sun H-P, Wang X, Ma Q, Xu Y, Luo N, et al. The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Qual Life Res. 2014;24(7):1767–74.

44. Yang F, Lau T, Lee E, Vathsala A, Chia KS, Luo N. Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). Eur J Health Econ. 2015;16(9):1019–26. doi:10.1007/s10198-014-0664-7.

45. Scalone L, Ciampichini R, Fagiuoli S, Gardini I, Fusco F, Gaeta L, et al. Comparing the performance of the standard EQ-5D 3L with the new version EQ-5D 5L in patients with chronic hepatic diseases. Qual Life Res. 2013;22(7):1707–16.

46. Tayyem R, Ali A, Atkinson J, Martin CR. Analysis of health-related quality-of-life instruments measuring the impact of bariatric surgery: systematic review of the instruments used and their content validity. Patient. 2011;4(2):73–87.

47. Tayyem R, Atkinson J, Martin C. Development and validation of a new bariatric-specific health-related quality of life instrument "bariatric and obesity-specific survey (BOSS)". J Postgrad Med. 2014;60(4):357–61.

48. Mihalopoulos C, Chen G, Iezzi A, Khan MA, Richardson J. Assessing outcomes for cost-utility analysis in depression: comparison of five multi-attribute utility instruments with two depression-specific outcome measures. Br J Psychiatry. 2014;205(5):390–7.

49. Scalone L, Cortesi PA, Ciampichini R, Cesana G, Mantovani LG. Health related quality of life norm data of the Italian general population: results using the EQ-5D-3L and EQ-5D-5L instruments. Epidemiol Biostat Public Health. 2015;12(3):e11457-1-15.

50. Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Appl Health Econ Health Policy. 2013;11(3):287–98.