Comparing and Explaining Differences in the Magnitude, Content, and Sensitivity of Utilities Predicted by the EQ-5D, SF-6D, HUI 3, 15D, QWB, and AQoL-8D Multiattribute Utility Instruments

Jeff Richardson, PhD, Munir A. Khan, PhD, Angelo Iezzi, MSc, Aimee Maxwell, BBNSc (Hons)

Background. *Cost utility analysis permits the comparison of disparate health services by measuring outcomes in comparable units, namely, quality-adjusted life-years, which equal life-years times the utility of the health state. However, comparability is compromised when different utility instruments predict different utilities for the same health state. The present paper measures the extent of, and reason for, differences between the utilities predicted by the EQ-5D-5L, SF-6D, HUI 3, 15D, QWB, and AQoL-8D.* ***Methods.*** *Data were obtained from patients in seven disease areas and members of the healthy public in six countries. Differences between public and patient utilities were estimated using each of the instruments. To explain discrepancies between the estimates, the measurement scales and content of the instruments were compared. The sensitivity of instruments to independently measured health dimensions was measured in pairwise comparisons of all combinations of the instruments.* ***Results.*** *The difference between public and patient utilities varied with the choice of instrument by more than 50% for every disease group and in four of the seven groups by more than 100%. Discrepancies were associated with differences in both the instrument content and their measurement scales. Pairwise comparisons of instruments found that variation in the sensitivity to physical and psychosocial dimensions of health closely reflected the items in the instrument's descriptive systems.* ***Discussion.*** *Results indicate that instruments measure related but different constructs. They imply that commonly used instruments systematically discriminate against some classes of services, most notably mental health services. Differences in the instrument scales imply the need for transformations between the instruments to increase the comparability of measurement.* ***Key words:*** *utility measurement; multiattribute utility; cost utility analysis; economic evaluation.* ***(Med Decis Making 2015;35:276–291)***

Cost utility analysis (CUA) allows the comparison of interventions that have a different impact on the health-related quality of life (HRQoL). This is achieved by measuring quality-adjusted life-years (QALYs), which are defined as life-years times an index of utility. If the utilities in these calculations are measured on a commensurable scale, then QALYs represent a commensurable unit of outcome. Health state utilities have increasingly been estimated using a multiattribute utility instrument (MAUI): a generic health-related questionnaire about the HRQoL (the ''descriptive system'' or ''classification'') and an accompanying formula or set of weights for converting responses into utility scores. In practice, a limited number of MAUIs dominate the literature. A review of articles listed on the Web of Science between 2005 and 2010 found 1663 studies that had used an MAUI.[1] Of these, 63% used the EQ-5D (EuroQoL, 5 Dimension, 3-item instrument—precursor to the EQ-5D-5L, which has 5 levels); 15% used the HUI 2 (Health Utilities Index Mark 2) or HUI 3 (Mark 3), 9% used the SF-6D (Short Form, 6 Dimension); and the remaining 13% used the 15D (15 Dimension), AQoL (Assessment of Quality of Life), or Quality of Wellbeing (QWB) index (6.9%, 4.3%, and 2.4%, respectively).

DOI: 10.1177/0272989X14543107

The reliability of CUA is compromised if different MAUIs produce different utilities for the same person, and the validity of comparisons between different classes of services (physical, psychosocial) is compromised if the instruments have different sensitivities to the dimensions of health. Discrepancies between instruments have been noted for some time.[2,3] In the two-large scale surveys comparing five MAUIs published to date, it was found that on average, only 42% and 56% of the variance in one instrument could be explained by another instrument.[4,5] The 2005–2010 review identified 392 head-to-head comparisons of the main instruments. The authors generally found a low correspondence between instruments and concluded that their comparison ''warrants caution''[6] and that instruments are ''not equivalent''[7] and are ''imprecisely related.''[5] The problems that arise from these differences have been recognized in the literature.[8]

The discrepancies in measurement are unsurprising. The instruments described by Brazier and colleagues[9] and Richardson and colleagues[1] differ in their conceptualization, content, size, and the methods used for converting health state descriptions into utilities. Descriptive systems are contrasted in Table 1, which divides the dimensions of health into the two domains of physical and psychosocial health. Three of the instruments—EQ-5D, HUI 3, and 15D—have a preponderance of items that relate to physical health. The SF-6D has an equal number of items in the two broad domains, and the AQoL-8D has a preponderance of items in the psychosocial domain. The QWB, whose descriptive system is constructed largely from symptom/problem clusters, is focused primarily upon the physical domain. Conceptually, HUI 3 has a ''within the skin'' descriptive system: It focuses on an individual's body functions. The other instruments are conceptualized primarily, but not exclusively, in terms of handicap (more recently described by the World Health Organization[10] as ''activity'' and ''participation'')—that is, the effect of a health state on a person's ability to function in a social environment. The items combine to describe between 945 and 8.7×10^{23} health states (QWB and AQoL-8D, respectively). (The recent revision of the EQ-5D, in which response categories were increased from three levels to five, increased the number of health states described from 243 to 3125. In the following, EQ-5D and EQ-5D-3L will refer to the 3-level version of the instrument, and EQ-5D-5L to the 5-level version.)

The present paper compares the six MAUIs listed in Table 1: the EQ-5D-5L, SF-6D, HUI 3, 15D, QWB, and AQoL-8D. The overall objective is to document differences in the utilities that are predicted by these instruments (objectives 1 and 2 below) and to investigate the reasons for the differences (objectives 3–5). The five specific objectives are as follows: 1) to determine the convergence between instruments: the extent to which each appears to measure a common construct and in common units; 2) to determine the predictive consistency of the instruments: the extent to which they predict the same loss of utility by patients in seven disease areas when they are compared with the healthy public; 3) to identify the role of measurement scale effects in explaining discrepancies between predicted utilities; 4) to determine differences in the content of each instrument as measured by independently constructed dimensions of the QoL; and 5) to carry out pairwise comparisons of the sensitivity of instruments: the extent to which one of the pair responds more to changes in particular dimensions of the HRQoL. The study draws on results from a large online survey that is described below.

METHODS AND DATA

Data

A multi-instrument comparison (MIC) survey was carried out in six countries: Australia, Canada, Germany, Norway, the UK, and the US. The online survey was administered by a global panel company, CINT Pty Ltd. The survey was approved by the Monash University Human Research Ethics Committee (reference number CF11/3192-2011001748).

Received 15 January 2014 from the Centre for Health Economics, Monash University, Melbourne, Australia (JR, MAK, AI, AM). Financial support for this study was provided by a grant from the National Health and Medical Research Council (NH&MRC) project grant (ID 1006334), ''A cross national comparison of eight generic quality of life instruments.'' The German arm of the study was funded by Deutsches Krebsforschungszentrum (DKFZ), the German Cancer Research Center in Heidelberg, and was conducted with the Institute for Innovation and Valuation in Health Care in Weisbaden. The Norwegian arm was facilitated by a grant from the University of Tromso. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. Authors include the researchers responsible for the creation of the AQoL-8D. Revision accepted for publication 12 June 2014.

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://mdm.sagepub.com/supplemental.

Address correspondence to Jeff Richardson, Centre for Health Economics, Monash University, Level 2 Building 75, Clayton VIC 3800 Australia; telephone: +61 3 9905 0754; +fax: 61 3 9905 8344; e-mail: Jeffrey.richardson@monash.edu.au.

Table 1 Comparison of the Dimensions and Content of Six MAUIs

	EQ-5D-5L	SF-6D	HUI 3	15D	QWB[a]	AQoL-8D
Dimensions of physical health[b]						
Physical ability/mobility/vitality/coping/control	1	1	2	2	8	3
Bodily function/self-care	1			3	13	1
Pain/discomfort	1	1	1	1	14	2
Senses			2	2	5	2
Usual activities/work	1	1		1	12	4
Communication			1	1	2	1
Dimensions of psychosocial health[b]						
Sleeping				1	1	1
Depression/anxiety/anger	1	1	1	3	4	7
General satisfaction						4
Self-esteem						2
Cognition/memory ability			1			
Social function/relationships		1				6
(Family) role		1				1
Intimacy/sexual relationships				1	1	1
Total no. of items (or symptoms, for QWB[a])	5	6	8	15	68[a]	35
No. of health states described	3125	18,000	972,000	3.1×10^{10}	945	24×10^{23}

a. QWB has 3 items plus 27 symptom/problem clusters.
b. For the physical and psychosocial dimensions of health, the values in the table are numbers of items (or symptoms, for QWB[a]).

Respondents were initially asked to indicate whether they had a chronic disease and to rate their overall health on a numerical scale where 0.00 represented death and 100 represented best possible health (physical, mental, and social). Quotas were then used to obtain a demographically representative sample of the healthy public, defined by the absence of a chronic disease and by a score above 70 on a numerical health scale. Quotas were also applied to obtain a target number of respondents in each of seven chronic disease areas: arthritis, asthma, cancer, depression, diabetes, hearing loss, and coronary heart disease. In the following text, respondents with a chronic disease and those considered healthy are referred to as ‘‘patient’’ and ‘‘public’’ respondents, respectively. To ensure statistical power in each subgroup, the survey sought a sample of 800 patients in each disease area and a sample of 1600 healthy respondents.

The survey included the six instruments described in Table 1. Utilities for five of the instruments were calculated using algorithms provided by the instruments' authors: SF-6D,[11] HUI 3,[12] 15D,[13] QWB,[14] and AQoL-8D.[15] The 5-level EQ-5D-5L utilities were obtained from the crosswalk published by the EuroQoL Group[16] using methods described by van Hout and colleagues.[17,18] In addition, respondents were administered the SF-36 (Short Form, 36 item, UK version), which is the most widely used and validated health profile instrument,[19] and a widely used instrument for measuring subjective well-being, the Satisfaction with Life Scale (SWLS).[20,21] Scores for the SF-36 dimensions and the SWLS were obtained conventionally by summing unweighted responses. These were rescaled to a 0.00–1.00 scale using the formula $S = (U_n - \text{Min})/(\text{Max} - \text{Min})$, where S is the rescaled score, U_n is the unweighted score, and Max and Min are the maximum and minimum possible unweighted scores, respectively.

Responses were subject to a set of stringent editing procedures based on a comparison of duplicated or similar questions in the questionnaire. Additionally, results were removed when an individual's (recorded) completion time fell below 20 minutes, which was judged to be the minimum time in which the 230 questions could be answered. Editing procedures, the questionnaire, and its administration are described by Richardson and colleagues.[22]

Convergence

Each of the MAUIs seeks to measure a common construct, namely utility. The conventional test for the existence of a common construct is the size of the Pearson correlation coefficient. However, the MAUIs seek to measure utility on the same scale, that is, one with a common interval property, as this is a necessary condition for equating the value of QALYs constructed from different MAUIs. Consequently, the absolute convergence between scales was also compared, using the intraclass correlation coefficient (ICC).

Prediction

Differences between the utilities of the public and patient groups were predicted using each of the six MAUIs. The term *prediction* is used here simply to denote the numerical values produced by the instruments that represent an estimate, or prediction, of the strength of preference for a health state.

Scale Effects

The relationship between instruments may be expressed as Equation 1:

$$U_i = a + bU_j + res, \quad \text{Equation 1}$$

where U_i and U_j are the utilities predicted by $MAUI_i$ and $MAUI_j$, respectively, a is a constant, and *res* is a residual resulting from differences in the (context specific) sensitivity of the two instruments and to measurement error. If both instruments measure utility on a common scale and incremental utilities share the same interval property, then $a = 0.00$ and $b = 1.00$. Deviation from these parameters indicates that differences in predicted utilities are, in part, attributable to scale effects. To test the magnitude of these effects, Equation 1 was estimated for each instrument pair using geometric mean square (GMS) regressions. The parameters of these are the geometric means of the parameters from the OLS regressions $U_i = a_1 + b_1U_j + e_1$ and $U_j = a_2 + b_2U_i + e_2$. The technique permits errors in both variables, and results are independent of choice of dependent and independent variable.[23]

Content

The relative importance of different dimensions of health was analyzed using scale invariant beta coefficients derived from the ordinary least squares (OLS) regression in Equation 2, in which U_i is the numerical score from $MAUI_i$, and D_n represents dimension scores.

$$U_i = a + \sum_{n=1}^{8} b_n(D_n) + e_i. \quad \text{Equation 2}$$

Dimensions were separately measured using the dimension scores of the SF-36 and the AQoL-8D. These are described in Table 2. Because these two instruments conceptualize and measure HRQoL in related but different ways, the two sets of results give related but different descriptions of the content of each MAUI. The construction of the AQoL-8D dimensions is outlined by Richardson and colleagues.[24] Psychometric methods used were similar to those used to construct the SF-36, and each of the resulting dimensions achieved satisfactory scores on tests of construct validity.

Sensitivity

Sensitivity refers to ''the degree to which gradations in the metric of a measurement scale allow detection of important or meaningful differences.''[25] The residual, *res*, in Equation 1, includes both an error term and, potentially, a systematic component if $MAUI_i$ and $MAUI_j$ differ in their sensitivity to dimension D_n. A positive correlation, $\rho\ (res, D_n) > 0$, implies that U_i varies more with D_n than U_j: That is, $MAUI_i$ is more sensitive to D_n than is $MAUI_j$. Conversely, a negative correlation implies that $MAUI_j$ is relatively more sensitive to D_n. The analysis again used the dimension scores from the SF-36 and AQoL-8D. With two exceptions, these are independent of the residual and error terms: Residuals from regressions including SF-6D are not independent of the SF-36 dimensions, and regressions including AQoL-8D are not independent of the AQoL-8D dimensions. These residuals were excluded from the analysis.

RESULTS

Table 3 summarizes the response to the online survey. Approximately equal numbers of respondents were obtained in each country. Editing procedures resulted in the exclusion of 16.6% of respondent records, leaving a sample of 8022: 1760 ''public'' respondents and 6262 ''patients.'' Due to the use of quotas, the age and gender distributions are very similar. The numbers in the three educational categories were approximately equal in Australia and the UK. In Canada, Norway, and Germany, the numbers in the middle educational categories were somewhat higher, and in the US the largest number was in the lowest educational category.

Description

Summary statistics are reported in Table 4. The mean utility for each instrument is very similar across countries. The largest difference between the Australian, US, UK, and German means across all of the

Table 2 Dimensions of the Health-Related Quality of Life (QoL) Instruments Used in the Content Analyses

	SF-36[a]	AQoL-8D[b]
Physical QoL	Physical function (Phys), 10 items[c] [21 levels[d]] • Vigorous/moderate activities • Lifting • Climbing stairs • Bending • Walking • Bathing Role physical (Role P), 4 items [5 levels] • Time spent on work • Difficulty performing work Bodily pain (B Pain), 2 items [10 levels] • Degree of pain • Interference with normal work due to pain level General health (Gen H), 6 items [21 levels] • Perceptions of general health rating • Excellent health	Independent living (Ind Liv), 4 items [56 levels] • Household tasks • Mobility • Walking and self-care Senses (Sense), 3 items [34 levels] • Vision; hearing and communication Pain (Pain), 3 items [38 levels] • Frequency of pain • Degree of pain • Interference with usual activities caused by pain
Mental QoL	Vitality (Vital), 4 items [21 levels] • Energy/tiredness Social functioning (Social), 2 items [9 levels] • Interference with normal and social activities Role limit emotional (Role E), 3 items [4 levels] • Work time • Work accomplished • Work less carefully than usual Mental health (MH), 5 items [26 levels] • Nervousness • Feel down • Felt calm/happiness	Happiness (Happy), 4 items [55 levels] • Contentment • Enthusiasm • Happiness, pleasure Coping (Cope), 3 items [41 levels] • Energy • Control • Coping Relationships (Relation), 7 items [50 levels] • Relationship with family, friends • Social isolation • Intimate relationships • Community role Self-worth (Worth), 3 items [48 levels] • Worthlessness/confidence Mental health (Mental), 8 items [72 levels] • Depression/sleep • Anger • Self-harm • Despair • Worry • Sadness • Tranquility
SWB	Satisfaction with Life Scale,[e] 5 items 5 items: satisfaction with conditions and way of life, past and present	

Note: QWB = Quality of Wellbeing index; SWB = subjective well-being.
a. Ware and Sherbourne.[19]
b. Richardson and colleagues.[15]
c. Some dot points contain more than one item.
d. *Levels* refers to the number of separate states used by at least one respondent. The number is less than the number theoretically possible when none of the 8022 respondents used a health state.
e. Diener and colleagues.[20]

Table 3 Respondents' Characteristics

Composition of Final Sample for Age, Gend1er, Education, and Total Number

	Excluded, %		Public, %							Patient, %							Education, %			
Country	Public	Patient	18–24	25–34	35–44	45–54	55–64	65+	Male	18–24	25–34	35–44	45–54	55–64	65+	Male	High School	Diploma, Certificate, or Trade	University	Total, *n*
Australia	36.5	15.3	11.3	18.1	18.9	18.5	14.7	18.5	46.4	2.1	8.0	10.3	19.5	32.6	27.5	50.4	35.8	35.1	29.1	1430
USA	17.1	11.2	10.3	17.8	18.1	20.2	16.2	17.4	45.2	4.8	8.8	13.1	25.0	25.5	22.8	36.4	36.1	29.3	34.6	1460
UK	18.8	13.2	11.4	15.4	20.1	18.1	14.4	20.5	47.7	7.1	12.7	9.7	16.4	29.0	25.1	51.4	38.1	30.2	31.7	1356
Canada	9.4	19.2	12.8	18.3	16.2	20.1	16.8	15.9	47.3	5.8	15.1	18.0	19.1	27.3	14.8	34.8	29.2	47.6	23.2	1330
Norway	19.1	19.1	12.8	16.0	16.7	18.4	15.6	20.5	50.3	6.2	8.2	10.2	16.8	26.0	32.6	63.6	28.0	48.5	23.5	1177
Germany	24.4	17	6.5	20.0	18.5	23.1	17.7	14.2	50.4	5.2	8.3	17.5	31.4	24.4	13.2	54.2	19.6	55.0	25.4	1269
Total	21.2	15.7	11.0	17.6	18.0	19.7	15.9	17.8	47.8	5.1	10.1	13.1	21.4	27.6	22.6	48.0	31.4	40.4	28.2	8022

instruments is 0.03. Norwegian means are between 0.03 and 0.08 above the average. The difference occurs in both the public and the patient samples.

In contrast, means for the different MAUIs vary significantly. For the public sample, as reported in Table 4, the QWB mean of 0.74 implies a potential for improvement of 0.26 (i.e., 1.00–0.74). The potential improvement measured by the EQ-5D and HUI 3 is 0.12, and that measured by the 15D is only 0.06. The standard deviation for the 15D in the total sample is one half the standard deviation of the HUI 3. The percentage of respondents reporting maximum scores ($U = 1.00$) varies from 19.1 for the EQ-5D to 0.3 for the AQoL-8D. The existence of floor effects limits the variation that will be detected by an instrument. Respondents with scores below 0.4 vary from 1.3% for the SF-6D to 14.7% for AQoL-8D. Ceiling and floor effects are analyzed in more detail in the supplementary data (Table S1).

Figure 1 plots the average mean utility predicted by each MAUI in each percentile of the sample when the sample is ranked from highest to lowest utility. The ranking of individuals therefore varies with the instrument, and the plots do not necessarily show comparative utilities for the same individuals. However, the figure visually illustrates the differences summarized above. Relative to other MAUIs, the 15D compresses utilities. QWB utilities decline very rapidly but then fall relatively slowly. SF-6D has the same pattern but the effects are less pronounced. Below the 75th percentile, incremental utilities predicted by the HUI 3 and EQ-5D decline very significantly, with both instruments predicting negative values for some health states.

Convergence

Convergence between MAUI scores is reported in Table 5. Pearson correlation coefficients are displayed in the top right of the diagonal in Table 5 and ICCs in the bottom left of the diagonal. Averages are reported for the correlations that include a particular instrument. The maximum and minimum averages obtained in the country-specific analyses are also reported. The highest Pearson correlation is between the 15D and the AQoL-8D (0.84) and the lowest is between the EQ-5D and QWB (0.65). However, with the exception of the QWB, there is relatively little difference between the averages. Correlations from the public sample shown in parentheses are substantially lower than correlations found in the total sample, reflecting the range dependence of correlation. Table 5 also includes a measure of subjective

Table 4 Mean and Standard Deviation by Country and Instrument

MIC Country	EQ-5D	SF-6D	HUI 3	15D	QWB[a]	AQoL-8D	SWB[b]
Australia	0.73 (0.22)	0.7 (0.13)	0.69 (0.28)	0.84 (0.13)	0.62 (0.15)	0.67 (0.22)	0.62 (0.21)
US	0.73 (0.23)	0.7 (0.14)	0.7 (0.27)	0.84 (0.13)	0.63 (0.16)	0.68 (0.23)	0.61 (0.21)
UK	0.71 (0.25)	0.7 (0.14)	0.67 (0.29)	0.83 (0.13)	0.62 (0.15)	0.64 (0.23)	0.59 (0.22)
Canada	0.75 (0.22)	0.72 (0.13)	0.72 (0.26)	0.85 (0.13)	0.64 (0.15)	0.69 (0.22)	0.64 (0.21)
Norway	0.79 (0.19)	0.74 (0.13)	0.79 (0.2)	0.89 (0.1)	NA	0.74 (0.2)	0.68 (0.2)
Germany	0.73 (0.23)	0.71 (0.14)	0.7 (0.27)	0.85 (0.13)	NA	0.67 (0.22)	0.62 (0.21)
Average total[a]	0.74 (0.23)	0.71 (0.14)	0.71 (0.27)	0.85 (0.13)	0.63 (0.15)	0.68 (0.22)	0.63 (0.21)
Average public[a]	0.88 (0.13)	0.80 (0.11)	0.88 (0.14)	0.94 (0.06)	0.74 (0.14)	0.83 (0.14)	0.70 (0.18)
U = 1.00, %	19.09	1.33	7.13	6.92	2.39	0.30	1.87
$U \leq 0.4$, %	8.90	1.31	13.94	0.30	6.53	14.66	19.31
Minimum utility	–0.51	0.30	–0.34	0.25	0.15	0.10	
Minimum possible[c]	–0.59	0.30	–0.36	0.11	0.09	0.09	

Note: Standard deviations are shown in parentheses. MIC = multi-instrument comparison; NA = not available. Total sample N = 8022.
a. QWB = Quality of Wellbeing index (QWB): n = 5576 (total), 1212 (public).
b. Subjective well-being (SWB) is measured by the Satisfaction with Life Scale (SWLS) described in Table 2.
c. The lowest score theoretically possible with the utility formula used.[1]

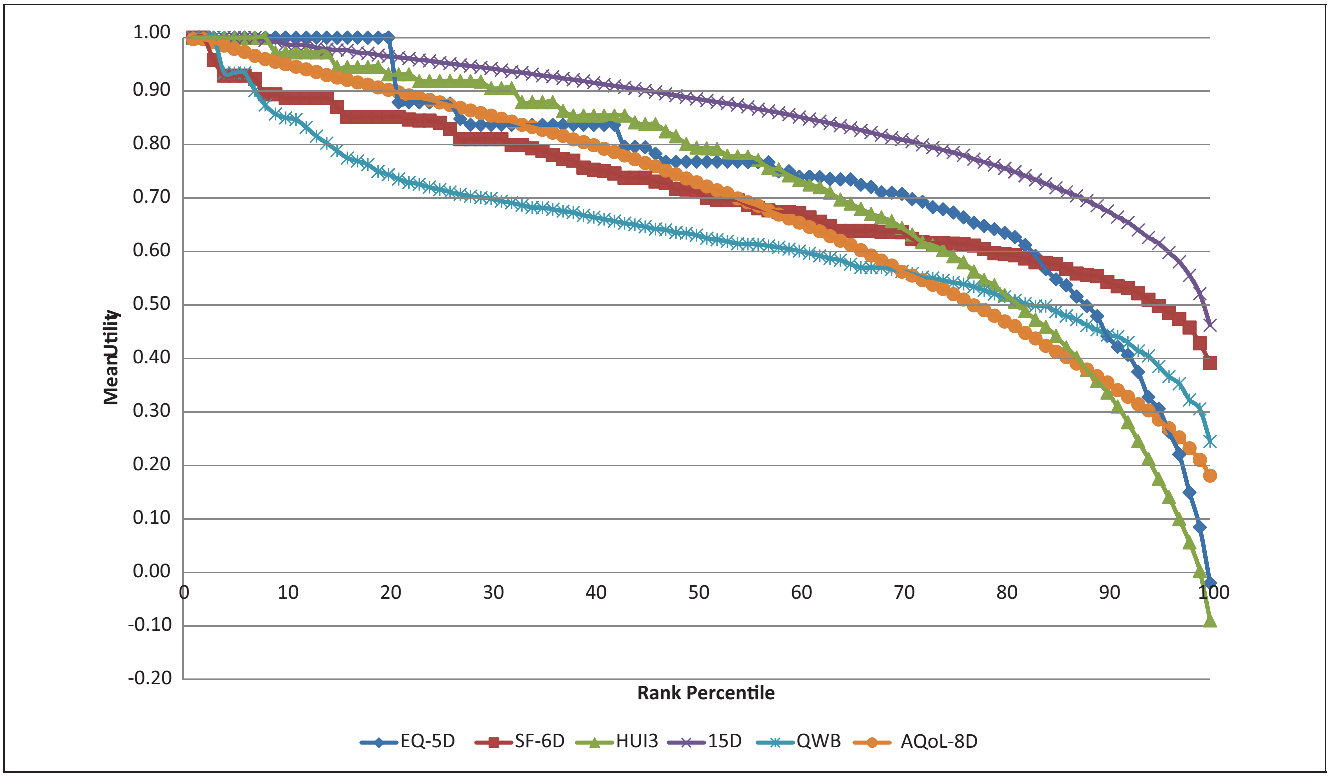

Figure 1 Mean instrument utility by ranked percentile.

well-being, the Satisfaction with Life Scale (SWLS). Its low correlation with all utilities indicates that the construct measured by the utility instruments has limited overlap with subjective well-being as measured by satisfaction scales. With the use of the ICC, correlations fall significantly, particularly for comparisons involving the 15D, where the average correlation drops from highest to lowest. Maximum and minimum country-specific results reported in the table indicate that correlations in

Table 5 Pearson and Intraclass Correlation Coefficients

	EQ-5D	SF-6D	HUI 3	15D	QWB[a]	AQoL-8D	Average,[b] Total Sample (Public Sample[c])	Maximum Average[d]	Minimum Average[d]	SWB[e]
EQ-5D	—	0.75	0.80	0.82	0.65	0.76	0.76 (0.57)	0.78 (UK)	0.75 (USA) (Can)	0.43
SF-6D	0.66	—	0.73	0.78	0.61	0.81	0.75 (0.53)	0.79 (Germ)	0.74 (Aus)	0.49
HUI 3	0.79	059	—	0.83	0.66	0.80	0.76 (0.56)	0.79 (UK)	0.73 (Can)	0.50
15D	0.59	0.50	0.53	—	0.73	0.84	0.80 (0.62)	0.80 (UK)	0.79 (Aus)	0.49
QWB[a]	0.54	0.60	0.54	0.34	—	0.69	0.68 (0.49)	0.70 (UK)	0.67 (Can)	0.39
AQoL-8D	0.73	0.71	0.78	0.50	0.62	—	0.78 (0.58)	0.82 (Germ)	0.77 (Aus)	0.66
Average,[b] total sample (public sample [c])	0.66 (0.47)	0.61 (0.40)	0.65 (0.46)	0.49 (0.32)	0.53 (0.34)	0.67 (0.48)	—			
Maximum average[d]	0.71 (Germ)	0.70 (Germ)	0.63 (Germ)	0.53 (Nor)	0.55 (UK)	0.74 (Germ)		—		
Minimum average[d]	0.66 (Aus)	0.65 (Aus)	0.59 (Aus) (Can)	0.47 (Aus)	0.51 (Aus) (Can)	0.68 (Aus)			—	

Note: All correlations were significant at the 0.01 level. Aus = Australia; Can = Canada; Germ = Germany; Nor = Norway.
a. Quality of Wellbeing index (QWB): $n = 1212$.
b. Average of the 5 correlations including the instrument.
c. Average for the Public sample: $n = 1760$.
d. Maximum and minimum average correlations in country-specific analyses.[26]
e. Subjective well-being (SWB) is measured by the Satisfaction with Life Scale (SWLS) described in Table 2.

Table 6 Difference in Mean Utilities between Public and Patient Respondents by Disease and MAUI

Group	EQ-5D	SF-6D	HUI3	15D	QWB	AQoL-8D		EQ-5D	SF-6D	HUI 3	15D	QWB	AQoL-8D	
Public: mean	0.88	0.8	0.88	0.94	0.74	0.83		0.88	0.80	0.88	0.94	0.74	0.83	
	Utility Difference: Public v. Patient						Max/Min	Percentage Reduction: Patient v. Public[a]						Max/Min
Arthritis	0.24	0.13	0.26	0.12	0.16	0.20	2.17	27.3	16.3	29.5	12.8	21.6	24.1	2.30
Asthma	0.12	0.09	0.12	0.09	0.11	0.14	1.55	13.6	11.3	13.6	9.6	14.9	16.9	1.76
Cancer	0.18	0.11	0.2	0.12	0.14	0.17	1.82	20.5	13.8	22.7	12.8	18.9	20.5	1.77
Depression	0.29	0.2	0.33	0.18	0.2	0.38	2.11	33.0	25.0	37.5	19.1	27.0	45.8	2.40
Diabetes	0.17	0.10	0.20	0.10	0.13	0.17	2.10	19.3	12.5	22.7	10.6	17.6	20.5	2.14
Hearing loss	0.09	0.05	0.18	0.06	0.1	0.11	3.60	10.2	6.3	20.5	6.4	13.5	13.3	3.25
Heart disease	0.16	0.10	0.18	0.11	0.13	0.15	1.80	18.2	12.5	20.5	11.7	17.6	18.1	1.75
Average difference	0.18	0.12	0.22	0.12	0.14	0.19		20.5	15.0	25.0	12.8	18.9	22.9	

a. {[*U*(Public) – *U*(Patient)] / *U*(Public)} × 100.

Table 7 Pairwise Linear Geometric Mean Square Regression Results

$MAUI_i = a + b\ MAUI_j$	R^2	$MAUI_i = a + b\ MAUI_j$	R^2	$MAUI_i = a + b\ MAUI_j$	R^2
EQ-5D = −0.43 + 1.64 SF-6D	0.56	SF-6D = 0.34 + 0.52 HUI 3	0.52	HUI 3 = −0.43 + 1.79 QWB	0.44
EQ-5D = 0.14 + 0.85 HUI 3	0.64	SF-6D = −0.21 + 1.08 15D	0.61	HUI 3 = −0.10 + 1.19 AQoL-8D	0.64
EQ-5D = −0.77 + 1.78 15D	0.67	SF-6D = 0.15 + 0.89 QWB	0.46	15D = 0.31 + 0.85 QWB	0.53
EQ-5D = −0.22 + 1.50 QWB	0.42	SF-6D = 0.29 + 0.61 AQoL-8D	0.64	15D = 0.46 + 0.57 AQoL-8D	0.69
EQ-5D = 0.05 + 1.01 AQoL-8D	0.58	HUI 3 = −1.07 + 2.09 15D	0.69	QWB = 0.18 + 0.68 AQoL-8D	0.47

Note: Total sample $N = 8022$.

different countries do not differ significantly from the average.

Prediction

Table 6 reports mean differences between the utilities of the healthy public and patients in each of the disease groups as predicted by each of the MAUIs. To facilitate comparability, unadjusted samples were used to calculate public utilities. The table was also constructed with a public sample weighted to match the demographic profile of each patient group. The outcome, which is reported in Supplementary Table S2, did not differ significantly from Table 6.

The predicted reductions in utilities differ significantly with the choice of MAUI. The ratio of the maximum to minimum reduction varies from 1.55 for asthma to 3.60 for hearing loss. For the calculation of the burden of disease, the more useful statistic is the loss of utility as a percentage of the utility of the public. With this statistic, some differences are attenuated but overall the differences remain large. The percentage reduction differs by a factor of 1.75 for heart disease and 3.25 for hearing loss. The ratios are large because they compare maximum and minimum changes. However, a comparison of the three most widely used instruments—EQ-5D, HUI 3, and SF-6D—reveals differences which are of a similar magnitude.

Scale Effects

Results of pairwise linear GMS regressions are presented in Table 7. They indicate that with one exception, scale effects play a significant role in explaining differences between utilities. Results are consistent with the pattern shown in Figure 1. On average, the scale of the SF-6D and the 15D compresses utilities. From the first column of Table 7, incremental utilities for the two instruments must be inflated on average by 64% and 78%, respectively, to be aligned with the EQ-5D. Conversely, from column 3, incremental utilities predicted by HUI 3 and AQoL-8D must be compressed by factors of 0.52 and 0.61 to be compatible with SF-6D utilities. The single pairwise comparison suggesting a common scale is between the EQ-5D and AQoL-8D (column 1). As a test of the stability of the relationships in Table 7, results were re-estimated for each of the six countries. Supplementary Table S3 reports results for the countries that deviate most from the overall result. It indicates that the relationships in Table 7 are very stable. Separate reports for each of the six countries are available online.[26]

Content

The regressions of MAUI utilities on the SF-36 and AQoL-8D dimension scores are reported in Supplementary Table S4. The content of the MAUI was calculated from these results by assuming a 1 standard deviation increase in every dimension and using beta coefficients to apportion the total increase in utility to the different dimensions.

The results from the two sets of analyses, displayed in Figure 2, are very similar. The content of the EQ-5D is dominated by pain and physical function (63% of the total using SF-36 dimensions) and by pain and independent living (67.8%) using AQoL-8D dimensions. The psychosocial dimensions of vitality, self-worth, role-emotion, and mental health account for 30.7% of content in the first set of results, and happiness, mental health, coping, relationships, and self-worth account for 32.2% in the second set. In contrast, physical and psychosocial dimensions account for 34.6% and 65.4% of the content of the AQoL-8D, respectively, using SF-36 dimensions, and 29.3% and 70.7% of content using the AQoL-8D's own dimensions. Results for the SF-6D are between these polar cases. Physical and psychosocial dimensions account for 41.8% and 58.2% of content using SF-36 dimensions and 48.1% and 51.9% of content using AQoL-8D dimensions.

Sensitivity

The correlations between residuals from pairwise regressions and the dimension scores from the SF-36, AQoL-8D, and SWLS are reported in Supplementary Tables S5 and S6. Table 8 reports dimensions where the correlation exceeded ±0.1. Dimensions in italic type indicate (positive or negative) correlations above 0.2. Columns indicate the instrument residuals with the greater sensitivity to the dimension; rows indicate the instrument residuals with lesser sensitivity. For example, from the first column, the EQ-5D is more sensitive to pain than the SF-6D, HUI 3, 15D, and AQoL-8D. From the first row, the EQ-5D is less sensitive to all of the psychosocial dimensions than the SF-6D, 15D, and AQoL-8D. It is less sensitive than the HUI 3 to senses and happiness. Countries with different results are recorded in Supplementary Table S7. The relatively small number of exceptions indicates that results are very similar in the six countries.

As noted, the SF-6D and AQoL-8D utility instruments are not independent of the dimensions of the SF-36 and AQoL-8D. Consequently, Table 8 does not record results where the SF-6D is more sensitive to SF-36 dimensions or where AQoL-8D is more sensitive to its own dimensions as the outcome is not the result of an independent test.

DISCUSSION

Because of the importance of MAUIs for the evaluation of health benefits, numerous studies have compared the MAUIs used to assess quality of life.[1] The present paper presents results from one of the broadest comparative surveys to date in terms of the range of diseases and the number of instruments included. It confirms previous conclusions that the utilities predicted from the major instruments differ significantly. The paper extends the analysis to identify reasons for the differences in the predicted utilities and, specifically, the importance of the measurement scales used by each instrument and the extent to which instruments are sensitive to different dimensions of health. Pairwise comparisons of instruments are presented to assist researchers with the selection of an instrument that is sensitive to the disease states they are investigating.

A number of general conclusions may be drawn from the study. First, the six MAUIs compared here measure related but different constructs. Second, instruments differ significantly in their relationship to different health dimensions, and the differences are primarily the result of the instruments' descriptive systems. Third, the differing magnitudes of the utilities predicted for the same health states are attributable not only to differences in instrument content but also to differences in the scales used to measure utility. Fourth, the differences in the utilities predicted by the instruments are so large that the choice of MAUI will have an important effect, and possibly a decisive one, on the outcome of an economic evaluation of a health service.

Convergence

The correlation between instrument utilities found here indicates a relatively high degree of convergence, which suggests the existence of a common

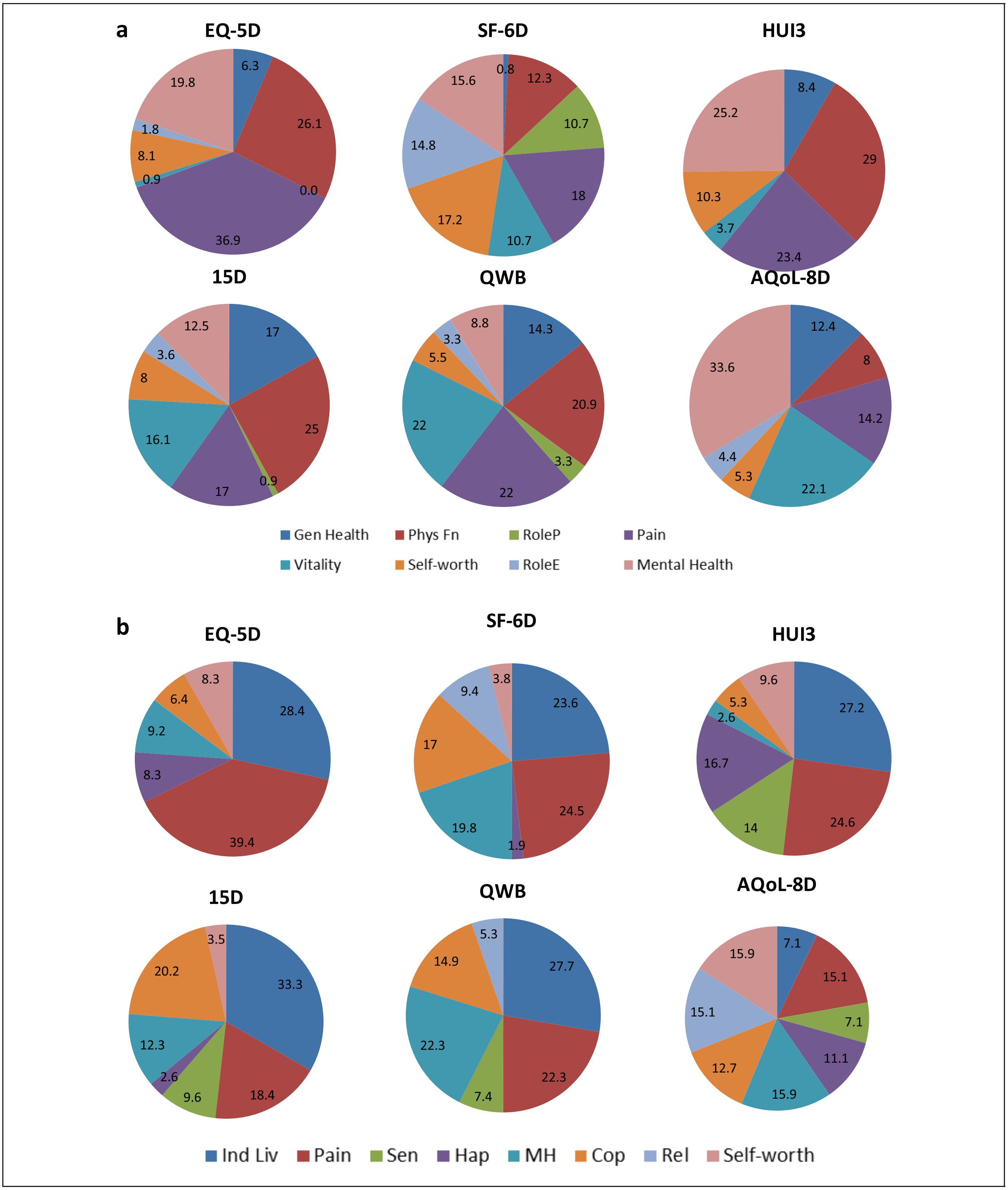

Figure 2 *Instrument content by dimension. (a) Relative increase in utility with a 1 standard deviation increase in each dimension of the SF-36. (b) Relative increase in utility with a 1 standard deviation increase in each dimension of the AQoL-8D. Numbers indicate the percentage of the explanatory power attributable to the dimension. Source: Supplementary Table S4.*

Table 8 Pairwise Comparison of Instrument Sensitivity to the Dimensions of the SF-36 and AQoL-8D

MAU with Less Sensitivity	MAU with Greater Sensitivity[a]				
	EQ-5D	**SF-6D[b]**	**HUI 3**	**15D**	**AQoL-8D[b]**
EQ-5D		Relation Mental	Sense Happy SWLS	Cope Vital Gen H Sense Relation Worth Role E Happy SWLS Mental	Vital Role E Social *MH* *SWLS*
SF-6D[b]	B Pain		Sense	Ind Liv Sense	
HUI 3	B Pain Pain	MH		Gen H Role P Vital Cope	Gen H Social Role E *Vital* *MH* *SWLS*
15D	B Pain Pain	Sense			Vital Role E *MH* *SWLS*
QWB	Phys Ind Liv B Pain *Pain* Social	Happy Mental Cope Relation Worth	Phys Social MH SWLS Worth Ind Liv Pain Sense Happy Cope	Every dimension of the SF-36, AQoL-8D, and SWLS	Gen H *Vital* *Social* Role E *MH* *SWLS*
AQoL-8D[b]	B Pain Phys		Phys	B Pain Role P *Phys*	

Note: Italic type indicates a correlation above 0.2; all other correlations are above 0.1 and below 0.2. Source—Supplementary Tables S5 and S6. Dimensions are described in Table 2. Countries with differing results are shown in Supplementary Table S7. SWLS = Satisfaction with Life Scale.
a. There is no column for QWB, as it was not the more sensitive instrument in any comparison.
b. Comparisons including SF-6D and AQoL-8D instrument scores with SF-36 and AQoL-8D dimensions are excluded.

construct. In part, the high correlation is attributable to the wide range of observations. When the data are restricted to public respondents, the average overall correlation reported in Table 5 falls from 0.76 to 0.56, suggesting that instruments are related to, but differ from, the common construct. Interpretation of the construct is further complicated by the fact that, despite the expectation that people should have a preference for subjective well-being—implying a high correlation between utility and the SWSL—this correlation is low, varying from 0.43 (EQ-5D) to 0.66 (AQoL-8D). The relationship between measured utility and subjective well-being requires further research.

Scale

For instruments in which change is interpreted relative to its own norms (such as many disease-specific and subjective well-being instruments), scale effects are not important: The linear transformation of the scale would have no significance if norms were adjusted accordingly. The utility used to construct QALYs differs. A 10% change in the numerical value of utility is treated as equally important as a 10% change in life expectancy (adjusted for time preference). Consequently, the differences in the linear relationships between utilities reported in Table 7 imply significant differences for the outcome of economic evaluations, not as a result of instrument sensitivity, but because of the differing scales imposed by the instrument's utility weights. Figure 1 indicates that the relationship between instrument utilities is not linear across the range of observations. Parameters reported in Table 7 provide a point comparison of the scales at the sample mean, but the relationship between scales varies across the range of observations.

Content

There is no gold standard for defining or measuring the appropriate dimensions of health that should influence utility. The analysis here has adopted the dimensions of the SF-36, which is the most widely used measure of HRQoL worldwide.[27] As described in Table 2, the dimensions of the AQoL-8D cover similar content but with differing emphasis and description. Results displayed in Figure 2 are also calculated from linear regressions. Consequently, they reflect average effects, which may vary across the range of observations as the meaning and importance of incremental dimension scores change.

Subject to these caveats, both sets of dimensions indicate that the content of the MAUIs varies significantly. In principle, generic items with appropriate weighting can compensate for the absence of an item dedicated to a particular dimension. For example, only two of the eight HUI 3 items relate to psychosocial health, but psychosocial items account for 39.2% of explained variance in the SF-36 equations (Figure 2). Nevertheless, the chief conclusion to be drawn from the present analysis is that MAUI content is a close reflection of the items in the descriptive system. Three of the five EQ-5D items concern pain and physical capacity, and 63% of the variance explained by the SF-36 dimensions is attributable to pain and physical function. In contrast, only two of the six SF-6D items relate to these dimensions, and only 41% of the explained variance is attributable to the corresponding SF-36 dimensions. Similarly, two of eight HUI 3 items and two of the 15D items relate to senses, and 14.0% and 9.6% of variance is explained, respectively, by the AQoL-8D dimension for ''senses.'' This dimension explains none of the variance of the EQ-5D or SF-6D, and these instruments have no items dedicated to senses. EQ-5D and AQoL-8D have 1 of 5 and 22 of 35 items, respectively, relating to psychosocial dimensions, and in the SF-36 regression analyses psychosocial dimensions account for 30.6% and 65.4% of the explained variance, respectively.

Sensitivity

The responsiveness of the MAUIs to different dimensions is closely related to instrument content. The pain component of the EQ-5D is larger than for any other instrument, and the EQ-5D is the most sensitive instrument for measuring pain. Using SF-36 regression results, the mental health content of the AQoL-8D is larger than for any other MAUI, and the AQoL-8D is the most sensitive instrument for measuring mental health (Table 8).

However, the correspondence between content and sensitivity is imperfect. As measured here, ''content'' refers to the composition of the dimensions which explain that part of the variation in a given MAUI that can be explained. Sensitivity measures the responsiveness of the MAUI to change in a dimension. The distinction is illustrated by a comparison of the QWB and 15D. Figure 2 indicates that the QWB has greater content relating to pain, physical function, and vitality using SF-36 dimensions and pain and mental health using AQoL-8D dimensions. However, from Table 8 we see that the 15D is more sensitive with respect to each of these dimensions.

A second distinction is between sensitivity and the numerical measure of change. In Table 6, the change in utility associated with every disease is greater when it is measured by the EQ-5D than by the SF-6D. In every case, it is greater when it is measured by the HUI 3 than by the 15D. However, these results do not reflect the sensitivity of the instruments. Table 7 indicates that with the exception of pain and independent living, the SF-6D is more, not less, sensitive to every AQoL-8D dimension than the EQ-5D. The 15D is more, not less, sensitive than the HUI 3 with respect to a variable number of dimensions, and it is not more sensitive than the 15D with respect to any dimension. These results are attributable to the scale effects discussed above. From Figure 1, the two apparently insensitive instruments based on the magnitude of the change in utility in Table 6—the SF-6D and 15D—have compressed utility scales. In comparison, HUI 3 and EQ-5D utility weights magnify the size of incremental differences particularly for poor health, and it is the scale effects of the utility weights, not the responsiveness to health dimensions, which account for the greater change in utility in Table 6.

Prediction

Differences in scale and content result in the discrepancies in predicted utilities reported in Table 6. These imply a corresponding discrepancy in the estimated cost per QALY when different MAUIs are used. Results in Table 6 imply that a cure for each disease that restored patients' quality of life to the same level as that of the healthy public would increase QALYs by an amount that varied by a factor of 1.55 for asthma and by a factor of 3.60 for hearing loss depending on the choice of instrument. The solution to this problem is not to mandate the use of a single instrument, as this would disadvantage services that improved health dimensions which the instrument measured poorly. In particular, the results indicate that the use of the EQ-5D, as promoted by the National Institute for Health and Clinical Excellence (NICE), will discriminate against services that primarily affect psychosocial health—a problem that has led the EuroQoL Group[28] to accept the need for the use of ''bolt-ons.''

A better solution is to evaluate services with the MAUI that is most sensitive to the relevant dimensions and to use a transformation to place results on a common scale.[5,29] The goal would be preservation of the instrument's sensitivity but standardization of pure scale effects. The key decision in applying this solution would be the choice of the ''numeraire'' scale. A crosswalk to the SF-6D scale would result in smaller QALY gains than a crosswalk to the HUI 3 scale. While the treatment of HRQoL would be standardized, the tradeoff between HRQoL and length of life would be significantly different.

Limitations

Both the data and analysis used in this study have limitations. First, data were obtained from respondents registered with a panel company, who may differ from the norm. A similar problem exists for the minority of the population that is willing to respond to a conventional survey. However, in both cases there are no strong grounds for believing that any atypical trait in these subgroups should affect the comparison of instruments: that answers provided on one scale will vary in a systematic way from answers provided on a second scale because of the unobserved trait. With online surveys, data are not subject to quality control at the point of collection. As described, this led to the application of rigorous editing procedures and the removal of results from inconsistent respondents. By the standards of previous studies, the correlation found between instruments was high, suggesting that frivolous answers did not have a significant effect. A strength of the survey was the inclusion of individuals with a wide range of chronic conditions. A limitation was the exclusion of individuals who were too ill to complete the survey. This implies the exclusion of the most severely ill.

Second, results were necessarily based on comparisons with imperfect instruments and using simplified relationships. Neither the SF-36 nor AQoL-8D dimensions represent a gold standard for the measurement of content and sensitivity: There are no gold standard instruments in this field. As reported in Table 2, several of the SF-36 dimensions have a limited number of levels, and ceiling effects are likely to result in insensitivity at the top of the scale. AQoL-8D dimensions have more levels, but almost 30% of the sample achieve the maximum score for two of the eight dimensions (pain and independent living). The analyses of both content and sensitivity are also based on a linear decomposition of instrument scores. These achieve high explanatory power: With three exceptions, the R^2 coefficients in every linear decomposition regression in Table S4 exceed the R^2 obtained from every pairwise regression between instrument utilities in Table 7. Nevertheless, analyses based on nonlinear functions might, in principle, alter the present results. Similarly, in the analysis of sensitivity, the residuals used in the correlations might be systematically altered by an omitted variable. There are no prior grounds for believing this, but further analysis with alternative statistical procedures is desirable to confirm the present results.

The third concern is that the same utility weights were used irrespective of the nationality of the respondents. This was unavoidable, as national weights do not exist for all of the instruments (and for this reason, the instruments are commonly used with the present weights in the six study countries). The problem is also general: The representative sample for a country is unlikely to produce the same results as a sample from the same country with different socioeconomic, cultural, and health characteristics; that is, within-country variation may be as large as or larger than between-country variation in preference structures. All weighting involves an averaging of disparate groups.

However, the present study suggests that differences in preferences between countries may be of secondary importance compared with the content of the instrument. The 15D, with Finnish weights, has a higher correlation with both the SF-6D and EQ-5D than the correlation between the latter two instruments, both of which have UK weights. Conversely, when the present EQ-5D-5L survey results were calibrated with US weights, the correlation between UK and US utilities, reported in Supplementary Figure S1, was 0.9898. Van Hout and colleagues[18] similarly reported that based on a common method (discrete choice modeling), results for the EQ-5D-3L in four countries are ''surprisingly consistent.'' The effect of alternative weights may, however, be tested by independent research teams using responses to the present survey when the database becomes freely available on the AQoL website in December 2014.

CONCLUSIONS

The utilities predicted by the six MAUIs in this study differ significantly and imply that the effects of a carefully controlled epidemiological study could be offset by the choice of the MAUI. The outcome is necessarily attributable to differences in the utility weights and the descriptive systems, and the analyses here found that both of these factors were of importance.

Despite seeking to measure utility on a common scale—which is a necessary condition for the comparability of QALYs—utility weights result in differing measurement scales. The pairwise regressions between instruments' utilities imply that, at the sample mean, there are significant differences in the interval properties of the utilities predicted by the six MAUIs. The differences vary over the full range of observations as the relationship between utilities varies.

The content of the descriptive systems and the sensitivity of instruments to independently measured health dimensions are a close reflection of the items in the descriptive system. EQ-5D utilities primarily reflect pain and physical function, and in all of the pairwise comparisons EQ-5D had greater sensitivity to these dimensions. AQoL-8D utilities have relatively greater psychosocial content, and in pairwise comparisons the AQoL-8D was the instrument most sensitive to these dimensions. (Because the dimensions used in this analysis are not independent of the SF-6D and AQoL-8D, these two instruments were not compared.)

It is desirable that the choice of MAUI in an economic evaluation be determined by true sensitivity and not by the scale on which utility is measured. The distinction between these implies the need for standardization to eliminate pure scale effects. Transformations may be used to achieve this so that in principle, the choice of instrument might be determined by the context-specific sensitivity of the instrument. In practice, the result will be imperfect as the ideal transformation might vary substantially by patient type and disease severity. However, compared with current practice, the results of such analyses would significantly improve the validity of comparisons between services.

Endnote

Access to the MIC database is described on the AQoL website: http://www.aqol.com.au/index.php/mic-data.

ACKNOWLEDGMENTS

The authors acknowledge helpful comments from Professors Jan Abel-Olsen, Julie Ratcliffe, Dr. Erik Nord, Dr. John McKie, and, in particular, Dr. Gang Chen.

REFERENCES

1. Richardson J, McKie J, Bariola E. Multi attribute utility instruments and their use. In: Culyer AJ, ed. Encyclopedia of Health Economics. San Diego (CA): Elsevier Science; 2014. p 341–57.

2. Nord E, Richardson J, Macarounas-Kirchmann K. Social evaluation of health care versus personal evaluation of health states. Int J Technol Assess. 1993;9(4):463–78.

3. Gold MR, Siegel JE, Russell LB, Weinstein M, eds. Cost-Effectiveness in Health and Medicine. Oxford (UK): Oxford University Press; 1996.

4. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. Ann Med. 2001;33:358–70.

5. Fryback DG, Palta M, Cherepanov D, Bolt D, Kim J. Comparison of 5 health related quality of life indexes using item response theory analysis. Med Decis Making. 2010;30(1):5–15.

6. McDonough CM, Grove MR, Tosteson TD, Lurie JD, Hilibrand AS. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among Spine Patient Outcomes Research Trial (SPORT) participants. Qual Life Res. 2005;14:1321–32.

7. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. Qual Life Res. 2008;17:485–95.

8. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E. Toward a consensus on the QALY. Value Health. 2009;12(1):S31–S5.

9. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford (UK): Oxford University Press; 2007.

10. World Health Organization. International classification of functioning, disability and health. http://apps.who.int/classifications/icfbrowser/

11. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002;21: 271–92.

12. Feeny D, Furlong W, Torrance G, et al. Multi attribute and single attribute utility functions for the Health Utilities Index Mark 3 System. Med Care. 2002;40(2):113–28.

13. Sintonen H, Pekurinen M. A fifteen-dimensional measure of health related quality of life (15D) and its applications. In: Walker S, Rosser R, eds. Quality of Life Assessment. Dordrecht (The Netherlands): Kluwer Academic Publishers; 1993.

14. Kaplan R, Bush J, Berry C. Health status: types of validity and the index of wellbeing. Health Serv Res. 1976;11(4):478–507.

15. Richardson J, Sinha K, Iezzi A, Khan MA. Modelling utility weights for the Assessment of Quality of Life (AQoL) 8D. Qual Life Res. http://link.springer.com/article/10.1007/s11136-014-0686-8#page-1

16. Rabin R, Oemar M, Oppe M, Janssen B, Herdman M. EQ-5D-5L user guide: basic information on how to use the EQ-5D-5L instrument. http://www.euroqol.org/fileadmin/user_upload/Documenten/PDF/Folders_Flyers/UserGuide_EQ-5D-5L_v2.0_October_2013.pdf

17. van Hout B, Janssen MF, Feng Y, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15:708–15.

18. van Hout B, Krabbe P, Kind P, Mulhern B. Innovative responses to methodological challenges (Slide 10). Presented at ISPOR 2011, Workshop Valuing Health in Economic Evaluations. Madrid: International Society for Pharmacoeconomics and Outcomes Research (ISPOR); 2011.

19. Ware J, Sherbourne D. The MOS 36-item Short-Form Health Survey (SF-36), I: conceptual framework and item selection. Med Care. 1992;30(6):473–83.

20. Diener E, Emmons RA, Larsen RJ, Griffin S. The Satisfaction With Life Scale. J Pers Assess. 1985;49(1):71–5.

21. OECD guidelines on measuring subjective well-being. http://dx.doi.org/10.1787/9789264191655-en

22. Richardson J, Iezzi A, Khan MA, Maxwell A. Cross-National Comparison of Twelve Quality of Life Instruments: MIC Paper 1: Background, Questions, Instruments, Research Paper 76. Melbourne (Australia): Centre for Health Economics, Monash University; 2012.

23. Tofallis C. Model fitting for multiple variables by minimising the geometric mean deviation. In: Van Huffel S, Lemmerling P, eds. Total Least Squares and Errors-In-Variables Modeling: Algorithms, Analysis and Applications. Kluwer Academic. http://ssrn.com/abstract=1077322

24. Richardson J, Elsworth G, Iezzi A, et al. Increasing the Sensitivity of the AQoL Inventory for Evaluation of Interventions Affecting Mental Health, Research Paper 61. Melbourne (Australia): Centre for Health Economics, Monash University; 2011.

25. Lohr K, Skillman S. Glossary: health outcomes methodology. Med Care. 2000;38(9 Suppl II):II-7–II-13.

26. Richardson J, Khan MA, Iezzi A, Maxwell A. Cross-national comparison of twelve quality of life instruments, Research Papers 78, 80–83, 85. MIC Report: 2: Australia; 3: UK; 4: USA; 5: Canada; 6: Norway; 7: Germany. Melbourne (Australia): Centre for Health Economics, Monash University; 2012.

27. Simmons CA, Lehmann P. Tools for Strengths-Based Assessment and Evaluation. New York: Springer; 2013.

28. EuroQoL Group. EQ-5D products: EuroQoL Group. http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html

29. Nord E. Health state values from multiattribute utility instruments need correction. Ann Med. 2001;33:371–4.