MONASH University
Business and Economics

Centre for Health Economics

Research Paper 2010 (50)

Summary Review of Multi-Attribute Utility (MAU) Instrument Comparison Studies: Methods, Techniques and Outcomes

Dr Munir A Khan

Research Fellow, Centre for Health Economics
Monash University

Dr. John McKie

Senior Research Fellow, Centre for Health Economics
Monash University

June 2010

Centre for Health Economics
ISSN 1833-1173
ISBN 1 921187 49 2

Correspondence:

Dr Munir A Khan
Centre for Health Economics
Faculty of Business and Economics
Monash University Vic 3800
Australia

Ph: +61 3 9905 0739 Fax: +61 3 9905 8344
Email address: Munir.Khan@buseco.monash.edu.au

ABSTRACT

Multi-attribute utility instruments (MAUs) permit the fairly rapid calculation of utility, or quality of life, scores for health states for use in economic evaluation. However, the descriptive system of a generic instrument is necessarily limited. Instruments may describe health states with greater or lesser accuracy and, generally, the fewer the number of items and dimensions the less sensitive it will be to differences in health states. The aims of this study are to review a number of studies that have compared MAU instruments, and to identify the methods and techniques that have been used in validating the instruments or comparing their effectiveness, ease of use, reliability and sensitivity. Different statistical econometric techniques, including descriptive statistics, ANOVA, Correlations, Regression, OLS, CLAD, T-Tests, IRT analysis, reliability and sensitivity tests were used in the studies surveyed. Significant differences in terms of utility were found between some MAU instruments: e.g. the EQ-5D and HUI3, SF36 and EQ-5D, SF-6D and EQ-5D, AQoL and all other instruments. There were also significant differences between instruments when comparing the HR-QoL of the general public and patients. In general the HUI3 and EQ-5D generate higher utility scores for both public and patients. When comparing the generic measures with disease specific measures it is found that the disease specific measures are better able to discriminate across different patient groups. It is evident from the outcome of the different studies surveyed that no single instrument or statistical technique is universally suitable for the general population and particular patient groups. Although choice of statistical technique – e.g. descriptive statistics, Correlations or simple Regression – had no impact on the results, the inclusion of t-tests, multivariate linear regression, factor analysis, multivariate analysis of variance, logistic regression, OLS, CLAD, IRT analysis or other advanced sensitivity techniques including partial derivatives (PD), sensitivity index (SI), Man-Whitney test (MW) and Smirnov test (S) may produce more accurate outcomes. But in general, the choice of instrument should depend on the study objectives.

TABLE OF CONTENTS

List of tables

Summary Review of Multi-Attribute Utility (MAU) Instrument Comparison Studies: Methods, Techniques and Outcomes

Introduction

Medical research has for many years been using disease-specific instruments to measure health-related quality of life (HR-QoL). There are now a large number of these, which are commonly used for economic evaluation. They are designed to assess quality of life, and changes in quality of life, in specific diagnostic groups or patient populations. In addition to the plethora of disease-specific instruments, a smaller number of generic instruments have also been designed. Generic measures purport to be broadly applicable across types and severities of diseases and conditions, and can be used to assess the quality of life of the general public. However, breadth of scope is gained at the expense of sensitivity and, consequently, researchers often employ generic and disease-specific instruments simultaneously.

It is sometimes objected that the different dimensions used in multi-attribute utility (MAU) instruments cannot be meaningfully combined into a single health-related quality of life score because the attributes are too heterogeneous. For example, it may be claimed that pain, independent living and coping ability are so dissimilar that their combination into a single index makes no sense. However, different items are combined into a single index when the GDP of a country is calculated, even though the products differ from theatre tickets to cars. Comparison is possible because it is the values of the products that are combined. The products per se are not combined. Similarly, the values of dissimilar health states as measured by people's preferences can, in principle, be compared or combined if the measurement is carried out properly.

Once constructed, MAUs permit the fairly rapid calculation of utility scores for health states, but they have limitations. The descriptive system of a generic instrument is necessarily limited. Instruments may describe health states with greater or lesser accuracy and, generally, the fewer the number of items and dimensions the less sensitive it will be to differences in health states. In general, there are substantial differences between the instruments with respect to their conceptual models of health-related quality of life, the content of the descriptive systems, the methods of weighting the different levels of health status, the algorithms for combing the different items and dimensions into utility scores, and the range of theoretical utility scores available. The aims of this study are to review a number of studies that have compared MAU instruments, and to identify the methods and techniques that have been used in validating the instruments or comparing their effectiveness, ease of use, reliability and sensitivity.

Methods

A large number of comparative studies have been published using existing multi-attribute utility instruments to measure and examine health related quality of life (HR-QoL). In order to identify the most recent research papers, an electronic library search was made using combinations of the key words 'comparison' 'MAU', 'instrument' and 'QoL' . The electronic database was used to search the literature from 2000 to 2010. The database included Econlit, ProQuest, Cinahl, Ovid Medline, Web of Science and Web of Knowledge. Hundreds of research publications were listed from the electronic library searches.

From the published research, the two basic categories of instruments mentioned above were identified: the generic instruments and disease-specific instruments. The generic instruments include the Rosser-Kind Index, the Quality of Well-Being (QWB), AQoL, EQ-5D (formerly the EuroQoL), the Canadian Health Utility Index (HUI), SF-6D (derived from the SF-36), SF-12, SF-36, and the Finish 15D. The disease specific measures include the Rheumatoid Arthritis Quality of Life Questionnaire (RAQoL) and Health Assessment Questionnaire (HAQ). These instruments may generate both psychometric and health utility scores. Health utilities can be measured either directly (using techniques such as the Standard Gamble (SG) or Time Trade-Off (TTO)) or indirectly (using multidimensional HR-QoL questionnaires developed using MAU instruments such as the Health Utilities Index 2 and 3 [1, 2], the SF-6D [3] and the EQ-5D [4]. Due to their ease of administration, these indirect measures are commonly used as the source of quality weightings in economic evaluation. Not all the studies uncovered in the literature search used these MAU instruments for validation purposes or to test instruments for reliability and sensitivity. Only papers reporting comparisons of utilities obtained directly or indirectly from patients and the general public are included in this summary review.

Description of MAU Measures

A good number of generic, preference-based and disease-specific measures have been identified from the retrieved comparison studies. Attention is given to eight generic measures and two disease-specific measures: the AQoL, the EQ-5D, SF-6D, SF-12, SF-36, HUI, 15D, QWB, RAQoL and HAQ. The latter two are disease specific measures. A brief description of each of the instruments is given below:

AQoL-8D

The AQoL-8D instrument has been recently developed in the Centre for Health Economics (CHE) at Monash University. The instrument consists of eight dimensions and 35 items. The eight dimensions include: independent living, life satisfaction, mental health, coping, relationships, pain and senses. The number of items and the number of responses per item vary. The full instrument may be obtained from the CHE website (http://www.buseco.monash.edu.au/centres/che/).

EQ-5D

The EQ-5D (EuroQoL) is a standardised instrument which was developed by a multi-disciplinary group of researchers from seven centres across five countries for use as a measure of health outcome. It is a five-dimensional preference-based measure, covering mobility, self-care, usual activities, pain and anxiety/depression. The five dimensions are measured on a three point response level: no, some or severe problems.

SF-6D

The SF-6D is derived from the SF-12 and SF-36 measures. The SF-36 is a self-administered questionnaire containing 36 questions that measures health across eight dimensions: physical functioning; role limitations- physical; social functioning; vitality or energy; bodily pain; mental health; role limitation– emotional; and general health. The SF-36 has become one of the most widely used measure of general health in clinical studies throughout the world. The SF-6D focuses on six domains of HR-QoL: physical functioning, role limitation, social functioning, pain, mental health and vitality. Only a selected number of items are used in the construction of the SF-6D.

SF-12

The SF-12 is derived from the SF-36 measures and includes 12 questions from the SF-36. These include: 2 questions concerning physical functioning; 2 questions on role limitations because of physical health problems; 1 question on bodily pain; 1 question on general health perceptions; 1 question on vitality (energy/fatigue); 1 question on social functioning; 2 questions on role limitations because of emotional problems; and 2 questions on general mental health (psychological distress and psychological well-being). Scoring of individual items is identical to the SF-36 measures. Scoring algorithms may then be applied to produce the Physical Component Summary (PCS) and Mental Component Summary (MCS) scores.

HUI3

The Health Utilities Index Mark 3 (HUI3) is a prominent measure of HR-QoL and widely used in population health surveys, clinical studies and cost-utility analyses, especially in Canada, where it originated. The HUI3 consists of eight dimensions or attributes: vision, hearing, speech, ambulation, dexterity, emotion and pain. The eight dimensions are measured on a scale of multiple levels ranging from 5 to 6 response levels. The HUI3 has been used to assess health status in a number of chronic conditions.

15D

The 15D is a comprehensive, self-administered instrument for measuring HRQoL among adults (age 16+ years) using 15 dimensions: mobility, vision, hearing, breathing, sleeping, eating, speech, elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality and sexual functions. Five ordinary levels are used for all dimensions. The respondent chooses from each dimension the level which best describes her/his present health status. The 15D combines the advantages of a profile and a preference-based, single index measure. A set of utility or preference weights is used to generate the 15D score (single index number) on a

0-1 scale. In most of the important properties the 15D compares favourably with other preference-based generic instruments.

QWB

The Quality of Well-being (QWB) is a self-administered questionnaire containing five parts: acute and chronic symptoms, self care, mobility, physical activity and usual activity. The QWB combines preference-weighted measures of symptoms and functioning to provide a numerical point-in-time expression of well-being that ranges from zero (0) for death to one (1.0) for asymptomatic optimal functioning - i.e. higher scores represent better health. The QWB has three function scales: Mobility (MOB), Physical Activity (PAC), and Social Activity (SAC) with three to five response levels. Each step of these scales has its own associated preference weight.

K-10

The Kessler Psychological Distress Scale (K-10) dates from 1992. It has been widely used in the USA as well as in Australia. The K-10 scale is based on 10 questions (items) related to negative emotional states experienced by individuals during the past four week period. There are five response levels for each item based on the amount of time the respondent reports experiencing the particular problem.

PWI

The Personal Wellbeing Index (PWI) was developed from the Comprehensive Quality of Life Scale (ComQol). The PWI scale contains nine items relating to life satisfaction, each one corresponding to a quality of life domain. It comprises: standard of living, health, achieving in life, relationships, safety, community-connectedness, future security, spirituality/religion and the level of satisfaction as a whole.

SWLS

The Satisfaction with Life Scale (SWLS) uses five key statements associated with the level of satisfaction relating to the quality of life: in most ways life is close to ideal; the conditions of life are excellent; satisfied with life; so far gotten the things wanted in life; and if I could live my life over, I would change almost nothing.

RAQoL

The Rheumatoid Arthritis Quality of Life Questionnaire (RAQoL) was developed in the UK and the Netherlands as a disease-specific tool. It is widely used in routine clinical practice to assess the specific quality-of-life needs of individual patients and to monitor the outcome of care for patient groups. The Australian version of the RAQoL is a valid and reliable tool for the assessment of quality of life. It is claimed to be practical, easy to administer and has good potential for use in clinical settings and trials in Australia.

HAQ

The Health Assessment Questionnaire (HAQ) was originally developed in 1978 at Stanford University. It was one of the first self-report functional status (disability) measures and has become

the dominant instrument in many disease areas, including arthritis. The HAQ is a comprehensive outcome measure that assesses patient outcomes in four domains: disability, discomfort and pain, drug side effects (toxicity) and dollar cost. It is widely used throughout the world and has become a mandated outcome measure for clinical trials in rheumatoid arthritis and some other diseases. Its focus is on self-reported patient-oriented outcome measures, rather than process measures.

Results

More than 60 multi-instrument comparison studies were found in time period from 2000 to 2010 and only the most relevant papers have been included in this working paper. These studies have validated and compared specific instruments using either the general public or specific patient groups in order to judge the effectiveness and sensitivity of the instruments. Different methods of data collection and data analysis techniques have been used in the studies, depending on the aims and objectives of the researcher(s), to establish the comparative advantage of using a specific instrument or to establish the effectiveness of a particular instrument. In most cases self-administered joint questionnaires containing chosen MAU instruments have been used. The summary findings, statistical techniques and methods of some selected studies that used different sets of MAU instruments in measuring and comparing the HR-QoL of specific patients group and the general public, are reported in the Appendix.

The Appendix shows that some of the MAU instruments have been used more frequently by researchers in measuring HR-QoL than others. This is possibly due to the simplicity and easy to use of the most popular instruments. For example, both the EQ-5D and SF-6D have many similarities and are easy to administer. However, frequency of use does not necessarily indicate the completeness or sensitivity of the instruments.

The MAU instrument comparison by disease type/patient groups, and study references, are summarised in Table 1 and Table 2. It appears from Table 1 that the EQ-5D, SF-6D and HUI are the most commonly compared instruments using either the general public or patient groups. These instruments have also been compared with the two disease specific instruments, the HAQ and RAQoL. The AQoL has also been compared with the most widely used generic MAU instruments, including the 15D and QWB. A number of instruments included in the review involve cross-country comparisons between Australia, the USA, the UK, Canada, Germany and Norway.

Table 2 summarises the groups and instruments included in the review. It reveals that the EQ-5D has been used most often, followed by HUI3, and the SF-6D, whereas approximately an equal number of studies used the AQoL, SF36 and 15D. These latter three instruments were mainly used to measure the HRQoL of the general public (Table 2).

Most of the studies using the EQ-5D, SF-6D, HUI3 and AQoL compared the HRQoL of the general public and patients with multiple diseases: rheumatoid arthritis, heart disease, pain, spine conditions etc. Table 3 shows the frequencies of studies by MAU instrument comparison. The highest number of studies was found to involve the comparison of the EQ-5D with HUI3 (26), followed by EQ-5D and SF-6D (17), AQoL and EQ-5D (7) and the comparison of EQ-5D with SF12 or 15D or QWB or SF36 (6).

All the generic MAU and disease specific instruments are correlated except SF-36 and EQ-5D. In general there were low to moderate negative correlations between the dimensions of SF-36 and EQ-5D. For example, these two instruments did not provide equivalent information on the HRQoL

of patients with chronic low back pain. A significant difference in terms of utility was found between some MAU instruments: e.g. the EQ-5D and HUI3, SF36 and EQ-5D, SF-6D and EQ-5D, AQoL and all other instruments. It can be seen from the Appendix that there is significant disagreement between instruments comparing the HR-QoL of the general public and patients. Different instruments gave different utility values, although many of the measurement properties were similar. In general the HUI3 and EQ-5D generate higher utility scores for both public and patients. However, in the case of patients with HIV/AIDs, the 15D produced higher utility scores (.86) than SF6D (.73) and EQ-5D (.77). When comparing the generic measures with disease specific measures it is found that the disease specific measures are better able to discriminate across different patient groups.

Different statistical econometric techniques, including descriptive statistics, ANOVA, Correlations, Regression, OLS, CLAD, T-Tests, IRT analysis, reliability and sensitivity tests were used in the studies to show the relationships and comparative sensitivity among the MAU instruments. Table 4 summarises the statistical techniques that have been used for MAU instrument comparison studies. It appears that descriptive statistics, correlations, regression, ANOVA and ICC were the most commonly used techniques that produced relevant statistics for the comparison. The choice of statistical approach appears to have no influence on the results.

Conclusion

It is evident from the outcome of different studies that no single instrument is universally suitable for the general population and particular patient groups. Instruments are not interchangeable for assessing HR-QoL. The choice of instrument should depend on the study objectives. Different preference based MAU instruments may yield different utility scores, which could have a significant impact on the calculation of quality-adjusted life years (QALY's) and may therefore have considerable effects in health evaluation studies. No single MAU instrument can claim to be the gold standard for measuring HR-QoL. However, there is evidence that the AQoL-8D has greater sensitivity than other instruments and its psychometric properties, which were a particular focus during the course of its development, are superior to many other instruments. In summary, researchers should select the MAU instrument that is sensitive to the health states which they are investigating. The combination of an MAU instrument and a disease-specific instrument may offer broad coverage of important health domains, as well as sensitivity to the specific condition or health state under investigation.

Table 1 Instruments compared and disease type/patient group

Study	Instruments compared	Instruments from our study (EQ-5D, HUI 3, SF-12, 15-D, QWB, PWI, SWB, SWLS)	Disease Type/Patient Group
5	AQoL, EQ-5D and SF-6D	EQ-5D and SF-6D	General public
6	AQoL, EQ-5D, HUI3, 15D,and QWB	EQ-5, HUI3, 15Dand QWB	General public
7	AQoL, EQ-5D, HUI3, SF-6D,SWLS,and PWI	EQ-5, HUI3, SF-6D,SWLS,and PWI	General public
8	EQ-5D, HUI2 and HUI3	EQ-5D, HUI2 and HUI3	General public
9	EQ-5D and SF-6D	EQ-5D and SF-6D	Health service researchers
10	SF-12 and EQ-5D	SF-12 and EQ-5D	General public
11	HUI, QWB, EQ5D and SF-6D	HUI, QWB, EQ5D and SF-6D	Literature based
12	SF-6D and EQ-5D	SF-6D and EQ-5D	General public and patients with diabetes
13	EQ-5D and HUI3	EQ-5D and HUI3	General public
14	EQ-5D and SF-6D	EQ-5D and SF-6D	Multiple diseases (10)
15	EQ-5D and SF-6D	EQ-5D and SF-6D	Multiple diseases (7)
16	AQoL and SF-36		Public and patients
17	SF-36		Chronic pain
18	HUI2, HUI3, SF-6D, EQ5D, RAQoL and HAQ	HUI, SF-6D and EQ-5D	Rheumatoid arthritis
19	EQ-5D and SF-6D	EQ-5D and SF-6D	Knee osteoarthritis
20	SF-36 and EQ-5D	EQ-5D	Low back pain
21	SF-36 and EQ-5D	SF-36 and EQ-5D	Symptomatic coronary heart disease
22	SF-6D and HUI3	SF-6D and HUI3	Cardiac
23	EQ-5D and SF-6D	EQ-5D and SF-6D	Mood or anxiety disorder
24	15D, EQ-5D and SF-6D	15D, EQ-5D and SF-6D	HIV/AIDS
25	AQoL and EQ-5D	EQ-5D	Hospitalised elderly
26	EQ-5D and SF-6D	EQ-5D and SF-6D	Liver transplant patients
27	EQ-5D and HUI2	EQ-5D and HUI	Benign prostatic hyperplasia
28	EQ-5D and HUI3	EQ-5D and HUI	Muscoskeletal
29	EQ-5D and HUI3	EQ-5D and HUI3	Intermittent claudication
30	EQ-5D, Rosser, QWB and HUI	EQ-5D, QWB and HUI	Road traffic injury
31	EQ-5D, AQoL	EQ-5D, AQoL	Hospitalised elderly
32	HUI2, HUI3, EQ-5D and SF-6D	HUI 3, EQ-5D	Rheumatoid arthritis
33	EQ-5D, HUI3 and SF-6D	EQ-5D, HUI3	Hearing-impaired people
34	EQ-5D, HUI 3	EQ-5D, HUI 3	Multiple sclerosis

Study	Instruments compared	Instruments from our study (EQ-5D, HUI 3, SF-12, 15-D, QWB, PWI, SWB, SWLS)	Disease Type/Patient Group
[35]	EQ-5D, SF-6D, HUI 2, HUI 3	EQ-5D, HUI 3	Stroke patients
[36]	HUI2, HUI3, SF-6D and EQ-5D, HAQ, RAQoL	HUI3, EQ-5D	Rheumatoid arthritis
[37]	EQ-5D, 15D	EQ-5D, 15D	Epilepsy
[38]	15D, EQ-5D and SF-6D	15D, EQ-5D	HIV/AIDS
[39]	HUI 3, EQ-5D, SF-6D	HUI 3, EQ-5D	Patients with musculoskeletal disease
[40]	EQ-5D, HUI, and SF-36	EQ-5D, HUI3	Spine patients
[41]	EQ-5D, 15D, HUI 2, HUI 3, SF-6D, QWB	EQ-5D, 15D, HUI 3, QWB	Rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders
[42]	HUI 3, EQ-5D, QWB	HUI 3, EQ-5D, QWB	Alzheimer's disease
[43]	HUI 3, EQ-5D	HUI 3, EQ-5D	Rheumatic disease
[44]	HUI 3, EQ-5D	HUI 3, EQ-5D	Stroke
[45]	SF36, HUI-2, HUI-3, EQ-5D, and HAQ	HUI-3, EQ-5D	Rheumatoid arthritis
[46]	HUI 3, EQ-5D	HUI 3, EQ-5D	Stroke
[47]	EQ-5D, SF-12	EQ-5D, SF-12	General public
[48]	EQ-5D, HUI 2, HUI3	HUI 3, EQ-5D	General public
[49]	HUI 3, EQ-5D	HUI 3, EQ-5D	General public
[50]	EQ-5D, SF-12	EQ-5D, SF-12	Back pain patients
[51]	EQ-5D, SF-12	EQ-5D, SF-12	General public
[52]	EQ-5D, SF-12	EQ-5D, SF-12	Patients with acute chest pain
[53]	EQ-5D, SF-12, HUI 3	EQ-5D, SF-12, HUI 3	General public
[54]	SF-36, 15D, EQ-5D, HAQ	EQ-5D, 15D	Rheumatoid Arthritis
[55]	EQ-5D, 15D	EQ-5D, 15D	General public
[56]	SF-36, HUI, EQ-5D, QWB	EQ-5D, QWB	General public
[57]	HUI, EQ-5D, QWB	EQ-5D, QWB	Cochlear implant patients
[58]	PANAS, OTH	SWLS, PWI	General public
[59]	AQoL, EQ-5D, HUI 3, 15D, QWB	AQoL, EQ-5D, HUI 3, 15D, QWB	General public
[60]	AQoL, EQ-5D, HUI 3, 15D, QWB	AQoL, EQ-5D, HUI 3, 15D, QWB	General public

Notes: **AQoL**: Australian Quality of Life; **EQ-5D**: EuroQoL 5D; **HUI**: Health Utilities Index; **SF-6D**: Short Form 6D; **SF12**: Short Form 12D; SF-36: Short Form 36D; **15D**: 15 Dimension; **HAQ**: Health Assessment Questionnaire; **QWB**: Quality of Well-Being; **PWI**: Personal Wellbeing Index; **SWB**: Subjective Well-Being; **SWLS**: Satisfaction With Life Survey; **PANAS**: Positive and Negative Affect Schedule; **OTH**: Orientations to Happiness; **RAQoL**: Rheumatoid Arthritis Quality of Life

Table 2 Study group, MAU instruments and study references

Study group/ Instru-ment	AQoL	EQ-5D	SF-6D	SF-12	SF-36	HUI	15D	QWB	Rosser -Kind Index	RAQoL	HAQ	SWLS	PWI	K-10	PANAS	OTH
General Public	[5-7,59-60]	[6-16,47-49, 51,55-56,59-60]	[6, 7, 11, 14, 17, 18]	[10, 13,47,51]	[56]	[6-8, 12, 15, 16,48,49, 56,59-60]	[15,55,59-60]	[12,56,59-60]				[7]	[7]	[7]	[58]	[58]
Multiple diseases		[18, 19,41]	[18, 19,41]			[41]	[41]	[41]								
Public and pa-tient	[20]				[20]											
Chronic Pain					[21]											
Rheuma-toid arthri-tis		[22-25,45,54]	[24, 25]		[22, 23,45,54]	[22, 24, 25,45]	[23,54]			[23, 25]	[22, 23, 25,45,54]					
Diabetes		[14]	[14]													
Knee os-teoarthri-tis (OA)		[26]	[26]													
Low back pain		[27,50]		[50]	[27]											
Heart disease		[28,44,46]	[28, 29]			[29,44, 46]										
Mood and/or anxiety disorders		[30]	[30]													
HIV/AIDS		[31]	[31]				[31]									
Hearing-impaired		[17]	[17]			[17]										

Study group/ Instru-ment	AQoL	EQ-5D	SF-6D	SF-12	SF-36	HUI	15D	QWB	Rosser -Kind Index	RAQoL	HAQ	SWLS	PWI	K-10	PANAS	OTH
people																
Spine		[32, 40]			[32]	[32,40]										
Elderly	[33]	[33]														
Liver transplant patients		[34]	[34]													
Benign prostatic hyperpla-sia		[35]				[35]										
Mus-coskeletal disease		[36]				[36]										
Epilepsy		[37]					[37]									
Intermit-tent clau-dication		[38]				[38]										
Road traffic injury		[39]				[39]		[39]	[39]							
Alz-heimer's disease		[42]				[42]		[42]								
Rheu-matic disease		[43]				[43]										
Acute chest pain		[52]		[52]												
Cochlear implant patients		[57]				[57]		[57]								

Table 3 Number of studies (from 2001 to present) by MAU instruments comparison

	EQ-5D	HUI3	SF-6D	SF-12	15D	QWB	PWI	SWB	SWLS	SF36	Rosser	HAQ	RAQoL
AQoL	7	4	2		3	3	1		1	1			
EQ-5D		26	17	6	6	6				6	1	4	2

Table 4 MAU instrument comparison studies – statistical tests, frequency of use and study reference

Statistical Tests	Number of Studies Used	Study Reference*
ANCOVA	2	17,50
ANOVA	7	7, 32, 29, 45, 50, 51, 54,
Bootstrapping methods	1	17
CLAD	1	10
Correlations	32	5, 6, 7, 15,16, 19, 20, 21, 22, 23, 27, 28, 29, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 46, 48, 49, 50, 52, 53, 54, 55,
Descriptive Statistics	44	5, 6, 7, 8, 9, 11, 12, 14, 15, 19, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53, 54, 55, 57, 58, 59, 60,
Factor analysis	3	16, 20, 32
F-statistic ratios	1	48
Interclass Correlation Coefficient (ICC)	5	15, 41, 42, 43, 44
IRT analysis	1	5
Kolmogorov-Smirnov Z test	1	14
Kruskall-Wallis	1	17
Log transformation with ANOVA	1	17
Logistic regression	1	7
Mann-Whitney U Test	3	14, 17, 54
Multinominal logit model	1	10
Paired t-test	4	22, 23, 33, 41
Partial Credit Rasch analysis	1	16
Regression	9	5, 10, 14, 15, 16, 32, 33, 46, 53
Standardised response means (SRM),	1	41
The Flesch-Kincaide (F.K) and Flesch Reading Ease (FRE) formulae	1	56
T-tests	4	17, 38, 51, 54
Two part model	1	10
Wilcoxson Rank	2	33, 40

Appendix

Table A. 1 Summary Findings, Statistical Techniques and Methods Used of Some Selected Multi-Instrument Comparison Studies

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Paz, S. H., H. H. Liu, et al. (2009)	SF-36, HUI, EQ-5D, QWB	General Public	Participants completed the questionnaire	The Flesch-Kincaid (F-K) and Flesch Reading Ease (FRE) formulae were used to estimate readability for every item in each measure.	The percentage of items deemed harder than "easy" according to FRE was 50 for the SF-36, 67 for the EQ-5D, 79 for the QWB-SA, 80 for the VFQ-25, 100 for the HUI, HALex, and the MLHFQ.	All seven surveys have a substantial number of items with high readability levels that may not be appropriate for the general population
Khan and Richardson (2009)	AQoL-8D, EQ-5D, SF-6D, HUI3, PWI, SWLS, K-10	General Public	Participants completed the questionnaire, (Psychometric)	Descriptive analysis, Correlation, ANOVA and logistic regressions.	All 7 instruments were highly correlated. AQoL-8D was most strongly correlated with K-10, SF-5D, EQ-5D and PWI. The HUI3 produced the highest number of individuals in full health and AQoL the fewest. Of the four instruments designed for economic evaluation studies, AQoL-8D and SF-6D outperformed the EQ-5D and HUI3.	AQoL-8D and SF-6D are more sensitive in measuring HRQoL of general public.
Torrance et al. (2009)	SF-36	Chronic Pain Physiopathology	Postal survey, Clinical Assessment Tools, Comparative Studies, Cross Sectional Studies	ANCOVA and t-tests; Kruskall-Wallis and Mann-Whitney U-tests; bootstrapping methods; and log transformation with ANOVA.	There were highly significant differences between the three groups, with lower scores in all SF-36 domains found those with chronic pain ($P < 0.001$).	Choice of statistical approach had no influence on the results.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Luo et al. (2009)	EQ-5D, HUI2, and HUI3	General public	Population health survey, index score	F-statistic ratios, Descriptive statistics, correlations	Poor agreement was found between instruments. The F-statistic values of EQ-5D and HUI2 index scores were similar, while those for EQ-5D versus HUI3 and for HUI2 versus HUI3 were significantly less than 1.0. The overall ceiling effects of the EQ-5D, HUI2, and HUI3 index scores were 48.9%, 15.4%, and 15.3%, respectively.	The EQ-5D seems to be marginally less informative. All three index scores were generally comparable in determining health burden of chronic medical conditions.
Kontodimopoulos et al. (2009)	SF-6D and EQ-5D	General public and patients with diabetes	Questionnaire survey	Cost Analysis, descriptive statistics	EQ-5D and SF-5D were in agreement and strongly correlated over the entire sample, but correlation varied according to socio-demographic factors and clinical conditions.	The hypothesis that EQ-5D generates higher scores in healthier populations and the SF-6D in less healthier groups were confirmed.
Chuang et al. (2009)	SF-12 and EQ-5D	General public	Panel survey 2003, index scores from SF-6D	Four econometric techniques (OLS, CLAD, Multinominal logit model, Two-part model).	Among four compared econometric techniques, OLS regression was the most accurate model in estimating the group mean. Models with item-based model specification performed better than those with summary score-based regardless of the chosen econometric technique. Nevertheless, the accuracy of OLS deteriorates in older and less healthy subgroups. The results also suggested that the two-part model, which addresses the heterogeneity issue, performs better in these vulnerable subgroups.	None of the mapping methods used in the study are suitable for estimating at the individual level.
Hawthorne, G.(2009)	AQoL, AQoL-8, EQ5D and SF6D	General Public	AQoL validation database and Australian Health Omnibus Survey	IRT analysis, descriptive statistics, Correlation, Regression, validity, reliability and sensitivity tests.	Similar psychometric properties found in AQoL-8 and AQoL. It correlated (intraclass correlation coefficient) 0.95 (or 90% of shared variance) with the AQoL. The AQoL-8 was as sensitive to six common health conditions as the AQoL, EQ5D, and SF6D	The utility scores fall on the same life-death scale as those of the AQoL. Where parsimony is imperative, researchers may consider use of the AQoL-8 to collect participant self-report HRQoL data.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Mitchell, J., R. Stanimirovic, et al. (2009)	PWI, SWLS, PANAS and OTH	General Public	Participants completed the questionnaire	Descriptive Stats	Significant changes were detected on the OTH subscales of engagement and pleasure. No changes in mental illness were detected by group or time. Attrition from the study was 83% at 3-month follow-up, with significant group differences in adherence to the intervention: strengths (34%), problem solving (15.5%) and placebo control (42.6%).	Although the results are mixed, it appears possible to enhance the cognitive component of well-being via a self-guided internet intervention.
Linde et al. (2008)	SF-36, EQ-15D, EQ-5D, RAQoL, and HAQ	Patients with Rheumatoid Arthritis	Patients completed the Questionnaires.	Descriptive Statistics, Correlations, ANOVA, T-Tests, Mann-Whitney U Test	All instruments discriminated between low, moderate, and high DAS28. RAQoL and HAQ displayed good repeatability (ICC > 0.95) and internal consistency (Cronbach's alpha > 0.90). All instruments were valid measures for HRQOL in RA. The RAQoL and HAQ displayed the best reliability, while the SF-36 bodily pain scale and VAS pain were the most responsive.	The choice of instrument should depend on the study objectives.
Barton et al. (2008)	EQ-5D(index), SF-6D, and EQ VAS	Patients with back pain, hip pain, knee pain, heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis, and osteoarthritis	Cross sectional survey. Patients completed the Questionnaires.	Descriptive statistics, Regression, Kolmogorov-Smirnov Z test, Mann-Whitney U test, t-test	There was a significant difference between HRQL of the majority of different groups according to each HRQL measure. However, not all of the measures could discriminate between groups of different ethnicity, gender, or smoking status, or those with and without asthma, smoke, cancer or diabetes.	The HRQL of the majority of different groups could be discriminated between by the EQ-5D index, SF-6D and EQ VAS
Horowitz et al. (2008)	HUI, QWB, EQ-5D, SF6D	Secondary data	Used questionnaires	Descriptive, statistical models,	Describes the method of creating and valuing preference based questionnaires and discusses the problems inherent in using the utilities they produce.	Specific preference based questionnaires. Using statistical models, a complete table of utilities for all profile may be constructed

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Moock, J. and T. Kohlmann (2008)	EQ-5D, 15D, HUI 2, HUI 3, SF-6D, and QWB	Musculoskeletal, cardiovascular, or psychosomatic disorders.		Descriptive Statistics, Correlations(ICC), paired t-tests, and standardized response means (SRM) were computed	Mean index scores at baseline ranged from 0.48 (HUI 3; psychosomatic) to 0.86 (15D; cardiovascular). At baseline, ceiling effects across all patient groups ranged from zero (SF-6D; cardiovascular and psychosomatic) to 21.6% (EQ-5D; cardiovascular). ICCs ranged from 0.26 (EQ-5D-QWB-SA; cardiovascular) to 0.80 (HUI 2-HUI 3; musculoskeletal). Substantial differences in responsiveness were observed between measures.	Differences between measures may have considerable effects in health economic evaluation studies, careful selection of instruments for a given study is essential
Hawthorne et al. (2008)	AQoL and SF-36	Patients and Public	AQoL validation data. Weighted score	Regression, partial credit Rasch analysis, Correlations, Factor analysis	Significant disagreement found. Many SF-36 items were limited predictors of AQoL items; some items from both instruments obtained disordered thresholds. All imputed scores were consistent with the AQoL model and fell within AQoL score boundaries. The explained variance between imputed and true AQoL scores was 61%.	Produced results are consistent with the axioms of utility measurement, while the proportion of explained variance was similar to regression-based modelling. Item properties on both instruments implied that some items should be revised using Rasch analysis.
Saarni, S., E. Saarni, et al. (2008)	EQ-5D and 15D	General public	Public completed the questionnaire	Descriptive statistics, Correlations	There were no differences in QoL or HRQoL between the entrepreneurs and salary earners. Farmers scored lowest on all measures; this finding remained even after adjusting for age, sex, marital status, education, and chronic conditions. The low WAI score of farmers was mainly explained by poor subjective work ability, while their low 15D score was mainly the result of poor functioning in the psychosocial domains of HRQoL. The low EQ-5D score of farmers was explained by problems with mobility, usual activities, and with pain or discomfort	Poorer work ability, QoL, and HRQoL do not appear to be caused by physical health problems.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Xie et al. (2007)	EQ-5D and SF-6D	Patients with knee osteoarthritis (OA)	Patients completed questionnaires. (Weighted score)	Descriptive statistics, Correlations	The mean +/- SD EQ-5D utility score was 0.49 +/- 0.31 (range -0.25-1.00) and the mean SF-6D utility score was 0.63 +/- 0.12 (range 0.32-0.89). This poor agreement was also demonstrated by the Bland-Altman plot and the low ICC (range 0.18-0.54). Correlations of the WOMAC and Lequesne index with the EQ-5D were higher than with the SF-6D.	Using different preference-based health-related quality of life instruments may yield different utility scores, which could have a great impact on QALY estimates. This highlights the importance of selecting appropriate instruments for economic evaluation.
Bas et al. (2007)	EQ-5D, HUI2 and HUI3	General public	Public completed the questionnaire	Descriptive statistics	Significant differences found between the instruments. Absolute informativity by instrument was consistently highest for HUI3 and lowest for EQ-5D, and relative informativity was highest for EQ-5D and lowest for HUI3.	Performance in terms of absolute and relative informativity of the common dimensions of the three instruments varies over dimensions.
Eker et al. (2007)	SF-36 and EQ-5D	Patients with low back pain	Patients completed both instruments during their enrolment procedure.	Pearson's Correlation Coefficient, Spearman's Rank Factor analysis	The partial correlation coefficients showed that there were generally low to moderate ($r<0.49$) negative correlation between the dimensions of the two instruments. Factor analysis revealed that although there are some similarities, the two instruments did not provide equivalent information on the health-related quality of life of patients with chronic low back pain.	The results of this study indicate that the instruments are not interchangeable for assessing health-related quality of life of patients with low back pain.
Van Stel et al. (2006)	SF-6D and EQ-5D	Patients with symptomatic coronary heart disease	Completed questionnaire prior to intervention and 1, 3, 6 and 12 months post-intervention.	Descriptive statistics, Correlations	Utility scores differed substantially. SF-6D focused more on social functioning, while EQ-5D gave more weight to physical functioning. Pain and mental health had similar contributions. The scoring range of the EQ-5D was twice the range of the SF-6D. Agreement was low, with an intra-class correlation of 0.45	EQ-5D and SF-6D are quite different. The low agreement and the differences in median values, scoring range and sensitivity to change after intervention show that the EQ-5D and SF-6D yield incomparable scores in patients with coronary heart disease.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Lamers et al. (2006)	EQ-5D and SF-6D	patients with mood and/or anxiety disorders	Data extracted from a randomised trial in Mental Health Care Centres.	Pared T-Tests, Spearman's Rank Correlation Descriptive statistics	Both EQ-5D and SF-6D utilities differed significantly between patients of adjacent severity groups. Mean utilities increased from 0.51 at baseline to 0.68 at 1.5 years follow-up for EQ-5D and from 0.58 to 0.70 for SF-6D.	Both EQ-5D and SF-6D discriminated between severity subgroups and captured improvements in health over time. Both instruments are to some extent interchangeable in cost-utility analysis.
Lee, H. Y., E. C. Park, et al. (2006)	HUI, EQ-5D and QWB	cochlear implants patients	Patients completed the questionnaire	Descriptive	Recipients used implants for an average of 5.6 years. The mean VAS, HUI, EQ-5D, and QWB score increased by 0.33 (from 0.27 before implantation to 0.60 at survey), 0.36 (0.29 to 0.65), 0.26 (0.52 to 0.78), and 0.16 (0.45 to 0.61), respectively. The discounted direct cost was $22 320, which yielded a cost-utility ratio of $19 223 per QALY using VAS, $17 387 per QALY using HUI, $24 604 per QALY using EQ-5D, and $40 474 per QALY using QWB.	
Haacke, C., A. Althaus, et al. (2006)	HUI2, HUI3 and EQ-5D	stroke	Data were collected by physicians using the questionnaire	Descriptive statistics, multivariate, correlations and regression	Four years after stroke, besides physical functioning, neuropsychological sequelae such as depression and cognitive impairment contributed to a reduced HRQoL. In addition, the incidence of incontinence proved to be an important factor for HRQoL. Explained variances in regression analysis models were high (R2=0.802 for HUI and 0.633 for EQ-5D-visual analogue scale) and were based on a few important determinants, including physical state, depression, cognitive impairment, and incontinence.	Results underscore the importance of nonmotor symptoms on HRQoL in patients with stroke.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Asadi-Lari, M., C. Packham, et al. (2005)	EQ-5D and SF-12	Patients with acute chest pain	Patients completed the questionnaire	Descriptive statistics, correlations	Women expressed more dissatisfaction than men overall ($p<0.05$) and appeared to have more physical needs. Women were more likely to complain about transport, which influenced their access to healthcare facilities ($p<0.001$), to be concerned about getting help with cleaning ($p<0.01$), and to request information about rehabilitation services, potential limitations on their daily activities, and nutrition and diet ($p<0.05$). Women had lower health-related quality of life scores in all the HRQL variables, which was significant in EQ-5D (usual activities, and pain/discomfort), Seattle angina questionnaire (angina stability), and both components of the Short Form-12.	Recognition of gender disparities in health needs and HRQL would clarify areas for improvement in healthcare services, and these might allow a better quality of life for infarct survivors.
Naglie, G., G. Tomlinson, et al. (2006)	EQ-5D, QWB and HUI3	Alzheimer's disease	Patients and care givers completed questionnaires	Descriptive statistics, Test-retest reliability (ICC) and Spearman correlations	Completion time was shortest for the combined EQ-5D and VAS. For patients with mild dementia and for proxies, reliability was >= 0.70 for the EQ-5D, QWB and HUI3. The EQ-5D had a ceiling effect for patient ratings. Convergent validity was demonstrated for patient and proxy ratings, with the strongest validity for EQ-5D ratings and the weakest validity for HUI3 patient ratings. Mean patient utility scores were significantly higher than mean proxy scores for all measures ($p < 0.001$).	For patient and proxy ratings, the EQ-5D had the best combination of measurement properties, although it had a substantial ceiling effect for patient ratings. Proxy QOL ratings did not accurately reflect patients' ratings

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Marra, C. A., J. C. Woolcott, et al. (2005)	HUI2, HUI3, SF-6D, EQ-5D , HAQ and RAQoL	Patients with Rheumatoid arthritis (RA)	Patients completed the questionnaires	Descriptive statistics, Correlations	Assessed the construct validity of utility scores from four generic preference-based measures and two disease specific measures in a sample of 313 RA patients in British Columbia, Canada. The disease-specific measures were better able to discriminate across groups with higher RA severity; however, utility scores from each of the scales also appeared to discriminate well across RA severity categories.	all of the preference-based utility measures that were evaluated appear to adequately discriminate across levels of RA severity.
Kaplan et al. (2005)	SF36, HUI-2, HUI-3, EQ-5D, and HAQ	Patients with Rheumatoid arthritis	Participants completed the questionnaires	Analysis of Variance, Descriptive Statistics, Confidence Interval	All 4 imputed scores were significantly correlated with HUI-2, HUI-3, EQ-5D, and the disease-specific HAQ scores at baseline and at the end of the clinical trial period ($P < 0.05$). Changes in the imputed scores from baseline to end of study also were significantly correlated with corresponding changes in the measured utility scores and the HAQ score ($P < 0.0001$).	Imputed utility-based score estimates are significantly correlated with measured utility outcome and had more constrained variability. Utility-based measures should continue to be favored for cost-effectiveness analysis.
Barton et al. (2005)	EQ-5D, HUI3 and SF-6D	hearing-impaired people	Participants completed the questionnaires	Correlation Descriptive statistics, Paired T-Tests, Linear Regression, Wilcoxson Rank Test	The mean utility scores on the EQ-5D (0.79), SF-6D (0.77), and HUI3 (0.56) were all significantly different from each other The agreement between measures was most commonly moderate according to the intra-class correlation coefficient (ICC =0.36 to 0.58).	The result suggests that different utility measures will provide different estimates of the effectiveness of hearing aid provision.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Bryan et al. (2005)	EQ-5D and SF-6D	Outcome from other researchers	Data collected from 35 health service researchers experienced in using health utility instruments.	Descriptive statistics	There remains considerable disagreement concerning the preferred generic utility-based measure of health-related quality of life for use in constructing quality-adjusted life years. The recent appearance (in a published form) of a new measure, the SF-6D, has highlighted this issue.	SF-6D and EQ-5D have many similarities but marked variation has been shown in the result. The findings suggest that the SF-6D can describe some poor health states, including states that (according to the EQ-5D scoring algorithm) are viewed as worse that the state of being 'dead'.
Fisk, J. D., M. G. Brown, et al. (2005)	HUI2, HUI3, EQ-5D and SF-6D	Rheumatoid arthritis	Patients completed the questionnaires	Descriptive statistics, Correlations	Correlations between assessment of clinical function and each health utility measure were strongest for the HUI Mark III (HUI Mark III EDSS p = -0.77, HUI Mark III ambulation index p = -0.76, HUI Mark III timed 25 foot walk p = -0.73, HUI Mark III nine hole peg test p = -0.65).	The HUI Mark III may be the most appropriate for cost effectiveness evaluations of MS therapies.
McDonough et al. (2005)	EQ-5D, HUI, and SF-36	Spine patient	Patients completed the questionnaires	Spearman rank correlations. Wilcoxson signed rank tests.	Mean values ranged from 0.39 to 0.63 among 2097 participants ages 18-93 (mean age 53, 47% female) with significant differences in pair-wise comparisons noted for all systems. Mean differences (95% CI) between those very dissatisfied and all others were 0.30 (0.269, 0.329) for EQ-5D, 0.22 (0.190, 0.241) for HUI(3), 0.18 (0.161, 0.201) for HUI(2), 0.11 (0.095, 0.117) for SF-6D, 0.04 (0.039, 0.049) for eQWB, and 0.07 (0.056, 0.077) for VAS (with transformation applied to group means).	Differences in preference-weighted health state classification systems are evident at baseline. Caution should be used when comparing health state values derived from various systems.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Stavem, K., S. S. FrÃ¸land, et al. (2005)	15D, EQ-5D and SF-6D	Patients with HIV/AIDS	Prospective observational study of 60 Norwegian patients with HIV/AIDS from two hospitals	T-Tests, Descriptive statistics, Correlations, Reliability test	On average, the 15D gave higher utility scores than the other two measures, the mean utility scores were: 15D--0.86, SF-6D--0.73, and EQ-5D Index--0.77. Test-retest reliability was acceptable for all measures, with intraclass correlation coefficients between 0.78 and 0.94. The correlation between scores of the 3 scales was substantial (p = 0.74-0.80). There was no major difference in responsiveness between the measures.	The different measures gave different utility values in this sample of patients with HIV/AIDS, although many of the measurement properties were similar.
Carr, J., J. Moffett, et al. (2005)	SF12, EQ5D	back pain	Patients completed the questionnaires	Descriptive statistics, Correlations, ANCOVA, Post-hoc	No statistically significant differences in change scores between groups on the primary outcome measure at three months (CI 72.24 to 0.49) and at 12 months (CI 71.68 to 1.39). Only minor improvements in disability scores were observed in the Back to Fitness group at three months and 12 months respectively (mean change scores; 70.89, 70.77) and in the individual physiotherapy arm (mean change scores; 70.02, 70.63).	Patients from the most severely deprived areas were marginally worse at three month follow-up whereas those from more affluent areas tended to improve (CI 0.43 to 3.15).
Pickard, A. S., J. A. Johnson, et al. (2005)	EQ-5D , SF-6D , HUI2 and HUI3	Stroke	Patients completed the questionnaires	Descriptive statistics, Correlations	The SF-6D, HUI3, and EQ-Index were generally more responsive than the HUI2 and EQ-5D Visual analogue scale (EQ-VAS). QALY estimates based on the EQ-5D index and HUI3 were twice as large as estimates based on the SF-6D and HUI2.	The results of this study may assist in informing the selection of a preference-based generic HRQOL measure, although choice will also depend on study goals and context.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Luo, N., J. Johnson, et al. (2005)	EQ-5D, HUI2, and HUI3	General public	Participants completed the questionnaire	Descriptive statistics, Correlations	Index scores (standard errors) for the general adult U.S. population as assessed by the EQ-5D, HUI2, and HUI3 were 0.87 (0.01), 0.86 (0.01), and 0.81 (0.01), respectively. The 3 overall preference indices were strongly correlated (Pearson's r: 0.67-0.87), but systematically different, with intraclass correlation coefficients between these indices ranging from 0.59 to 0.77.	This study provides U.S. population norms for self-reported health status on the EQ-5D, HUI2, and HUI3. Although these measures appeared to be valid and demonstrated similarities, health status assessed by these measures is not exactly the same.
McNamee, P. and J. Seymour (2005)	AQOL,EQ-5D, HUI3,15D, QWB	General public	Reviews	Descriptive	This paper reviews the methods that have been employed to examine the degree of interchangeability between the Assessment of Quality of Life, EuroQol-5D, Health Utilities Index Mark III, Short-Form-6D, Quality of Wellbeing and 15-dimension measures.	There is a need to develop alternative econometric strategies and to explore, more fully, economic concepts of validity.
Holland, R., R. D. Smith, et al. (2004)	EQ-5D, AQoL	Hospitalised elderly	Patients completed the questionnaires	Descriptive statistics	Poor agreement was found between both the absolute scores from each instrument and change in scores over time. Although the AQoL appeared to have more favourable construct validity, the EQ-5D was easier to administer, had a higher completion rate, and appeared more sensitive to change.	The AQoL is probably less well suited to measuring health status in a very elderly population than the EQ-5D
Pickard, A., J. Johnson, et al. (2004)	HUI3 and EQ_5D	Stroke	Patients and their family caregivers completed the questionnaires	Descriptive statistics, Intraclass Correlation coefficient (ICC)	Cross-sectional point estimates of agreement were generally acceptable (ICC >0.70) for the EQ-5D Index and HUI3 summary scores when assessed >or=1 month after baseline. Agreement between change scores was generally poor to fair (ICC <0.60), but systematic bias was not observed for the indirect preference-based summary scores between baseline and 6 months.	EQ-5D and HUI3 provided suboptimal agreement with patient assessment, limited systematic bias may support their consideration as alternatives to missing data or statistical imputation.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Brazier et al. (2004)	EQ-5D, SF-6D	Chronic obstructive air way disease, osteo-arthritis, irritable bowel syndrome, low back pain, leg ulcers, menopause, elderly	The overall data set has been formed by combining 'baseline' and follow-up' observations from five of the seven studies.	Spearman rank correlation, Descriptive statistics, Intraclass Correlation coefficient (ICC) reliability analysis, OLS regression.	Small difference in EQ-5D and SF-6D scores at the aggregate level but the SF-6D has a small er range and lower variance. Evidence of floor effects in the SF-6D and ceiling effects in the EQ-5D.	The discrepancies arise from differences in their full health state classifications and the methods used to value them.
Marra et al. (2004)	HUI2, HUI3, EQ-5D and SF-6D	Rheumatoid arthritis	Patients completed the questionnaires	AVOVA, Confidence intervals, Descriptive statistics, Factor analysis, Correlations, Regression, repeated measures	Mean (standard deviation) global utility scores were 0.63 (0.24) for the SF-6D, 0.66 (0.13) for the EQ-5D, 0.71 (0.19) for the HUI-2, and 0.53 (0.29) for the HUI-3 (P = 0.02 by repeated-measures analysis of variance). The intraclass correlation across all the indices was 0.67 (95% confidence interval 0.62-0.71). Bland-Altman plots revealed that agreement among instruments was poor at lower utility values. In this elderly RA sample, all of the global utilities mostly measured functional ability and pain.	There are significant differences in utilities obtained from different indirect methods. Agreement among the instruments was moderate but poorer at lower utilities. It is unlikely that these utility values, if used as the weightings for quality-adjusted life years, would result in comparable estimates.
Long worth and Bryan (2003)	EQ-5D, SF-6D	Liver transplant patients	Patients completed the questionnaire	Descriptive statistics	Important variations in the results from the EQ-5D and SF-6D. SF-6D does not describe health states at the lower end of the scale but is more sensitive in detecting small changes towards the top of the scale.	The two instruments generally provide similar results

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Luo, N., L. Chew, et al. (2003)	HUI3 and EQ-5D	Rheumatic disease	Patients were interviewed twice within 2 weeks using a questionnaire containing the EQ-5D, HUI3, and SF-36	Descriptive statistics, test-retest, reliability, ICC correlations	ICC values for the EQ-5D and HUI3 were 0.64 and 0.75, respectively (n = 90, median interval: 7 days). EQ-5D and HUI3 utility scores were similar (mean +/- SD: 0.75 +/- 0.21 vs 0.76 +/- 0.17, p = 0.647, paired t test) and showed moderate correlation (Spearman's r: 0.45, p < 0.001). Differences were present in patients' responses to these 2 instruments: e.g., 12 patients reporting no problems with mobility (EQ-5D item) reported different levels of disability with ambulation (HUI3 item	The EQ-5D and HUI3 performed equally well in measuring utility-based HRQoL in patients with rheumatic disease, although they measured slightly different, though related, dimensions of health
O'Brien et al. (2003)	SF-6D and HUI3	Patients at increased risk of sudden cardiac death	Patients completed the questionnaires (Utility score)	Descriptive statistics, Paired t-test, Pearson correlation	Differences in utility score found in SF-6D and HUI3 instruments. Mean scores for HUI3 and SF-6D were 0.61 (95% CI 0.60–0.63) and 0.58 (95% CI 0.54–0.62) respectively; a difference of 0.03 (p50.03). Score intervals for HUI3 and SF-6D were (-0.21 to 1.0) and (0.30–0.95). Correlation between the instrument scores was 0.58 (95% CI 0.48–0.68) and agreement by ICC was 0.42 (95% CI 0.31-0.52). Correlations between dimensions of SF-6D were higher than for HUI3.	Utility differences may be due to differences in underlying concepts of health being measured or different measurement approaches, or both. No gold standard exists for utility measurement and the SF-6D is a valuable addition that permits SF-36 data to be transformed into utilities to estimate QALYs.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Franks, P., E. Lubetkin, et al. (2003)	EQ-5D, SF-12 and HUI3	General public	Participants completed the questionnaire	Descriptive statistics, correlation, regression	For the EQ-5D Index regression, the adjusted variance explained was 58% (bootstrap validation 95% confidence interval [CI], 46-66). For the HUI3 regression, the adjusted variance explained was 51% (bootstrap 95% CI, 39-59). The correlation coefficient between the 2 predicted measures was 0.96. The correlation of the predicted HUI3 with the EQ-5D Index (0.73) and the predicted EQ-5D Index with the HUI3 (0.70) exceeded that between the 2 original preference-based measures themselves (0.69).	The results show that SF-12 could be successfully mapped to both the EQ-5D Index and HUI3, yielding preference-based scores that demonstrate convergent validity in a low-income, minority sample.
Conner-Spady, B. and M. E. Suarez-Almazor (2003)	HUI3, EQ-5D and SF-6D	Musculoskeletal disease	Patients visiting a rheumatology clinic completed the questionnaire	Descriptive, Correlations	Correlations ranged from 0.66 to 0.79. An interaction effect showed that for the better group, the EQ-5D showed a significantly greater mean improvement (0.15) than the HUI3 (0.07) or the SF-6D (0.05). For the worse group, the EQ-5D showed a significantly greater mean decrease (0.19) than either the HUI3 (0.05) or the SF-6D (0.03). QALYs differences between the better and worse groups were significantly greater (0.23) with the EQ-5D than with the HUI3 (0.11) or the SF-6D (0.09)	The instruments are not interchangeable because they are scaled differently and produce varying results. Possible approaches are sensitivity analysis or standardization of scores before calculation of QALYs
Schulz et al. (2002)	EQ-5D, HUI-II	Benign prostatic hyperplasia	Patients completed the questionnaire (utility score)	Descriptive, Correlations	Both EQ-5D and HUI-II accept able to patients but EQ-5D was easier to administer. Both instruments failed to detect 'symptomatic' changes in patients' conditions.	

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Hawthorne et al. (2001)	AQoL, EQ-5D, SF-6D, HUI-III, 15D,	General sample	Community members and outpatients attended two of Melbourne's hospitals completed the questionnaire	Descriptive statistics, Spearman correlations,	Substantial differences found between the five instruments with respect to their conceptual models of HR-QoL, the content of the descriptive systems, methods of weighting the different levels of health status and the algorithms for utility score. The largest range of scores given by the AQoL, EQ-5D and HUI-III. The most highly correlated instruments were the AQoL and the 15D; the least correlated were the EQ-5D and the HUI-III. EQ-5D and HUI-III were the least sensitive to change at the top of the scale.	Researchers should select the utility instruments that is sensitive to the health states which they are investigating.
Lubetkin and Gold (2001)	EQ-5D, HUI-III	General sample	Participants completed the questionnaire	Descriptive stats, correlations	Moderate to strong correlations seen between levels/categories of do mains between the EQ-5D and HUI-III.	
Hawthorne, G. and J. Richardson (2001)	AQoL, EQ-5D, HUI-III, 15D, QWB	General sample	Reviewed other studies	Descriptive stats,	No current instrument satisfies all the requirements for multiattribute utility measurement. We recommend that users should choose instruments most relevant to their circumstances and that studies should include two instruments. Rigorous sensitivity analyses should be conducted and both results reported. Subject to these caveats, preference should be given to instruments best meeting multiattribute utility theoretical requirements, viz., the AQoL or HUI3. However, we recognise that other instruments may perform as well, or even better, under certain circumstances.	Some instruments may perform better under certain circumstances.

Study Reference (1)	MAU instruments compared (2)	Population group/ disease areas (3)	Methods of data collection and score used (4)	Technique Used (5)	Outcome or summary findings (6)	Comments (7)
Corner-Spady and Suarez-Almazor (2001)	EQ-5D, HUI-III	Muscoskeletal disease	Patients completed the questionnaire	Descriptive stats, correlations	Moderately high correlations between utility scores from EQ-5D and HUI-III. EQ-5D most sensitive to change over 12 months. Largest range of scores seen for EQ-5D.	
Stavem et al. (2001)	EQ-5D, 15D	Epilepsy	Patients completed the questionnaire (utility score)	Descriptive, correlations, Reliability tests	Test-retest reliability good for both the EQ-5D and the 15D. Construct validity results similar for both instruments. Poor agreement of utility scores between EQ-5D and 15D.	
Johnson and Pickard (2000)	EQ-5D and SF-12	General public	Mail survey: Participants completed the questionnaire (Index score)	Descriptive statistics, ANOVA, T-Tests	Moderate correlation was found between EQ-5D index scores and SF-12 summary scores. Significant differences were found when analysed EQ-5D responses by demographic variables and self-reported chronic medical conditions. For subjects reporting no problems on the instruments score were significantly lower for people reporting medical problems or feeling of depression.	The results generally supported the validity of the EQ-5D but important ceiling effect was observed for the EQ-5D. The combination of the EQ-5D and SF-12 provides relatively broad coverage of important health domains and scores for various purposes.
Bosch and Hunink (2000)	EQ-5D, HUI-III	Intermittent claudication	Patients completed the questionnaires	Descriptive statistics, Correlations, ANOVA	Both EQ-5D and HUI-III demonstrated some lack of discriminatory power in this clinical condition but both were reasonably sensitive to change.	Both the HUI3 and EQ-5D showed similar relationships with other (HRQoL) measures. However, clinicians and researchers should be aware of the differences in the mean HUI3 and EQ-5D scores
Elvik (1995)	EQ-5D, Rosser, QWB, HUI	Road traffic injury	Participants completed the questionnaires (Index score)	Descriptive statistics, Standard error of mean	Poor agreement across the four instruments was found. Whilst the evidence on external validity was not strong for any of the instruments, the best was thought to be EQ-5D.	EQ-5D instrument appears to be the most valid of the four indexes. None of the four indexes give values that are consistent with public policy objectives. No 'gold standard' for validity exists.

References

1. Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care* 2002;40:111-128.

2. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang YM, Wang QN. Multiattribute utility function for a comprehensive health status classification system - Health Utilities Index Mark 2. *Medical Care* 1996;34:702-722.

3. Brazier J, Roberts J, M. D. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* 2002;21:271-292.

4. The-EuroQoL-Group. Euroqol - a new facility for the measurement of health-related quality of life. *Health Policy*; 1990:199-208.

5. Hawthorne G. Assessing Utility Where Short Measures Are Required: Development of the Short Assessment of Quality of Life-8 (AQoL-8) Instrument. *Value in Health* 2009;12:948-957.

6. Hawthorne G, Richardson J, Day NA. A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Annals Of Medicine* 2001;33:358-370.

7. Khan M, Richardson J. Report on Health Related Quality of Life and Lifestyle of Bangladeshi Migrants in Melbourne: use of Multi-Attribute Instruments. Research Paper 42, Centre for Health Economics, Monash University; 2009.

8. Bas Janssen MF, Birnie E, Bonsel GJ. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a US general population survey using Shannon's indices. *Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation* 2007;16:895-904.

9. Bryan S, Longworth L. Measuring health-related utility: why the disparity between EQ-5D and SF-6D? *The European Journal Of Health Economics: HEPAC: Health Economics In Prevention And Care* 2005;6:253-260.

10. Chuang LH, Kind P. Converting the SF-12 into the EQ-5D: An Empirical Comparison of Methodologies. *PharmacoEconomics* 2009;27:491-505.

11. Horowitz E, Hassidim H, Abadi-Korek I, et al. [Assessment of health related quality of life--part 3--preference based measures]. *Harefuah* 2008;147:914-919.

12. Kontodimopoulos N, Pappa E, Papadopoulos A, Tountas Y, Niakas D. Comparing SF-6D and EQ-5D utilities across groups differing in health status. *Quality of Life Research* 2009;18:87-97.

13. Lubetkin EI, Gold MR. Areas of decrement in health-related quality of life (HRQL): Comparing the SF-12, EQ-5D, and HUI 3. *Quality of Life Research* 2003;12:1059-1067.

14. Barton GR, Sach TH, Avery AJ, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged >/= 45 years. *Health Economics* 2008;17:815-832.

15. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* 2004;13:873-884.

16. Hawthorne G, Densley K, Pallant JF, Mortimer D, Segal L. Deriving utility scores from the SF-36 health instrument using Rasch analysis. *Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation* 2008;17:1183-1193.

17. Torrance N, Smith BH, Lee AJ, Aucott L, Cardy A, Bennett MI. Analysing the SF-36 in population-based research. A comparison of methods of statistical approaches using chronic pain as an example. *Journal of Evaluation in Clinical Practice* 2009;15:328-334.

18. Marra CA, Woolcott JC, Kopec JA, 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. *Social Science & Medicine* 2005;60:1571-1582.

19. Xie F, Li S-C, Luo N, et al. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. *Arthritis And Rheumatism* 2007;57:1043-1049.

20. Eker L, Tuzun EH, Daskapan A, Bastug ZO, Yakut Y. The relationship between EQ-5D and SF-36 instruments in patients with low back pain. *Fizyoterapi Rehabilitasyon* 2007;18:3-10.

21. van Stel HF, Buskens E. Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. *Health And Quality Of Life Outcomes* 2006;4:20-20.

22. O'Brien BJ, Spath M, Blackhouse G, Severens JL, Dorian P, Brazier J. A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics* 2003;12:975-981.

23. Lamers LM, Bouwmans CAM, van Straten A, Donker MCH, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Economics* 2006;15:1229-1236.

24. Stavem K, FrÃ¸land SS, Hellum KB. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. *Quality Of Life Research: An International Journal Of Quality Of Life Aspects Of Treatment, Care And Rehabilitation* 2005;14:971-980.

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 Economics* 2004;13:793-805.

26. Longworth L, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics* 2003;12:1061-1067.

27. Schulz MW, J Chen, H H Woo, M Keech, M E Watson, Davey PJ. A comparison of techniques for eliciting patient preferences in patients with benign prostatic hyperplasia. *J Urol* 2002;168:155-159.

28. Conner-Spady B, ME. S-A. A comparison of preference-based health status tools in patients with musculoskeletal disease. . *18th Plenary Meeting of the EuroQol Group*; 2001:235–245.

29. Bosch JL, Hunink MGM. Comparison of the Health Utilities Index Mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. *Quality of Life Research* 2000;9:591-601.

30. Elvik R. THE VALIDITY OF USING HEALTH STATE INDEXES IN MEASURING THE CONSEQUENCES OF TRAFFIC INJURY FOR PUBLIC-HEALTH. *Social Science & Medicine* 1995;40:1385-1398.

31. 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;13:793-805.

32. Marra CA, Esdaile JM, Guh D, et al. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Medical Care* 2004;42:1125-1131.

33. Barton GR, Bankart J, Davis AC. A comparison of the quality of life of hearing-impaired people as estimated by three different utility measures. *International Journal of Audiology* 2005;44:157-163.

34. Fisk JD, Brown MG, Sketris IS, Metz LM, Murray TJ, Stadnyk KJ. A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of Neurology, Neurosurgery and Psychiatry* 2005;76:58-63.

35. Pickard AS, Johnson JA, Feeny DH. Responsiveness of generic health-related quality of life measures in stroke. *Quality of Life Research* 2005;14:207-219.

36. Marra C, Woolcott J, Kopec J, 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). *Social Science & Medicine* 2005;60:1571-1582

37. Stavem K, Bjørnæs H, Lossius MI. Properties of the 15D and EQ-5D utility measures in a community sample of people with epilepsy. *Epilepsy Research* 2001;44:179-189.

38. Stavem K, Frøland SS, Hellum KB. Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. *Quality of Life Research* 2005;14:971-980.

39. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care* 2003;41:791-801.

40. McDonough CM, Grove MR, Tosteson TD, Lurie JD, Hilibrand AS, Tosteson ANA. Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among Spine Patient Outcomes Research Trial (SPORT) participants. *Quality of Life Research* 2005;14:1321-1332.

41. Moock J, Kohlmann T. Comparing preference-based quality-of-life measures: results from rehabilitation patients with musculoskeletal, cardiovascular, or psychosomatic disorders. *Quality of Life Research* 2008;17:485-495.

42. Naglie G, Tomlinson G, Tansey C, et al. Utility-based quality of life measures in Alzheimer's disease. *Quality of Life Research* 2006;15:631-643.

43. Luo N, Chew L, Fong K, et al. A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease. *Journal of Rheumatology* 2003 30:2268-2274.

44. Pickard A, Johnson J, Feeny D, Shuaib A, Carriere K, Nasser A. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. *Stroke* 2004;35:607-612.

45. Kaplan RM, Groessl EJ, Sengupta N, Sieber WJ, Ganiats TG. Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. *Medical Care* 2005;43:79-87.

46. Haacke C, Althaus A, A S, Siebert U, Back T, Dodel R. Long-term outcome after stroke: evaluating health-related quality of life using utility measurements. *Stroke* 2006;37:193-198.

47. Franks P, Hanmer J, Fryback D. Relative disutilities of 47 risk factors and conditions assessed with seven preference-based health status measures in a national U.S. sample: toward consistency in cost-effectiveness analyses. *Medical Care* 2006;44:478-485.

48. Luo N, Johnson JA, Shaw JW, Coons SJ. Relative Efficiency of the EQ-5D, HUI2, and HUI3 Index Scores in Measuring Health Burden of Chronic Medical Conditions in a Population Health Survey in the United States. *Medical Care* 2009;47:53-60.

49. Luo N, Johnson J, Shaw J, Feeny D, Coons S. Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Medical Care* 2005;43:1078-1086.

50. Carr J, Moffett J, Howarth E, et al. A randomized trial comparing a group exercise programme for back pain patients with individual physiotherapy in a severely deprived area. *Disability & Rehabilitation* 2005;27:929-937.

51. Johnson JA, Pickard AS. Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Medical Care* 2000;38:115-121.

52. Asadi-Lari M, Packham C, Gray D. Gender difference in health-related needs and quality of life in patients with acute chest pain. *British Journal of Cardiology* 2005;12:459-464.

53. Franks P, Lubetkin E, Gold M, Tancredi D. Mapping the SF-12 to preference-based instruments: convergent validity in a low-income, minority population. *Medical Care* 2003;41:1277-1283.

54. Linde L, SÃ¸rensen J, Ostergaard M, HÃ¸rslev-Petersen K, Hetland ML. Health-related quality of life: validity, reliability, and responsiveness of SF-36, EQ-15D, EQ-5D, RAQoL, and HAQ in patients with rheumatoid arthritis. *Journal of Rheumatology* 2008;35:1528-1537.

55. Saarni S, Saarni E, Saarni H. Quality of life, work ability, and self employment: a population survey of entrepreneurs, farmers, and salary earners. *Occupational & Environmental Medicine* 2008;65:98-103.

56. Paz SH, Liu HH, Fongwa MN, Morales LS, Hays RD. Readability estimates for commonly used health-related quality of life surveys. *Quality of Life Research* 2009;18:889-900.

57. Lee HY, Park EC, Kim HJ, Choi JY, Kim HN. Cost-utility analysis of cochlear implants in Korea using different measures of utility. *Acta Oto-Laryngologica* 2006;126:817-823.

58. Mitchell J, Stanimirovic R, Klein B, Vella-Brodrick D. A randomised controlled trial of a self-guided internet intervention promoting well-being. *Computers in Human Behavior* 2009;25:749-760.

59. Hawthorne G, Richardson J. Measuring the value of program outcomes: A review of Multiattribute utility measures. *Expert Reviews of Pharmacoeconomics Outcomes Research* 2001;1:215-228.

60. McNamee P, Seymour J. Comparing generic preference-based health-related quality-of-life measures: Advancing the research agenda. *Expert Review of Pharmacoeconomics and Outcomes Research* 2005;5:567-581.