MONASH University
Business and Economics

Centre for Health Economics

Research Paper 85 (2013)

Cross-national comparison of twelve quality of life instruments

MIC Paper 7

Germany

Professor Jeff Richardson

Foundation Director, Centre for Health Economics
Monash University

Dr Munir A Khan

Research Fellow, Centre for Health Economics
Monash University

Angelo Iezzi

Research Fellow, Centre for Health Economics
Monash University

Aimee Maxwell

Research Fellow, Centre for Health Economics
Monash University

May 2013

Centre for Health Economics
ISSN 1833-1173
ISBN 1 921187 84 0

The Multi Instrument Comparison (MIC) survey is a project funded by a National Health and Medical Research Council (NHMRC) project grant (ID 1006334) 'A cross national comparison of eight generic quality of life instruments'. (Since its inception, three additional instruments have been added).

The German arm of the study was funded by **Deutsches Krebsforschungszentrum** (DKFZ), the German Cancer Research Centre in Heidelberg and conducted with the **Institute for Innovation & Valuation in Health Care** in Weisbaden.

The Norwegian arm was facilitated by a grant from the University of Tromso.

Chief Investigators on the project are:

Professor Jeff Richardson (Monash University, Australia)
Professor Robert Cummins (Deakin University, Australia)
Professor Robert Kaplan (University of California Los Angeles, USA)
Professor Jan Abel Olsen (University of Tromso, Norway)
Professor Michael Schlander (University of Heidelberg, Germany)
Professor Joanna Coast (University of Birmingham, England)

Professor Jeff Richardson
Centre for Health Economics
Faculty of Business and Economics
Monash University Vic 3800
Australia

Ph: +61 3 9905 0754 Fax: +61 3 9905 8344
Jeffrey.Richardson@monash.edu

ABSTRACT

The Multi Instrument Comparison (MIC) project is the largest comparative study of health and wellbeing instruments undertaken worldwide. To date 8022 individuals have completed twelve instruments relating to their health or wellbeing. Data were collected from a representative healthy cohort and from patients in eight clinical areas in each of five countries.

This and other country-specific research papers report data related to the project study questions. They do not seek to interpret data or comment on the study questions. This will be the subject of later publications.

Countries, diseases and questionnaires included in the MIC are summarised in Boxes 1 to 4 below. The background study questions questionnaires and utility weights used are outlined in detail in MIC Paper 1, Background, Questions, Instruments (Richardson, Iezzi et al. 2012). Choice of weights is also discussed in Section 8.

Box 1 Country and disease area: Respondent numbers after editing

Total sample (after editing)	
Australia	1430
UK	1356
USA	1460
Canada	1330
Norway	1177
Germany	1269
Total	8022

Health state (after editing)	
Arthritis	929
Asthma	856
Cancer	772
COPD	66
Depression	917
Diabetes	924
Chronic heart disease	943
Stroke	23
Hearing problems	852
Total disease	6282
Healthy	1760

Box 2 Main Questionnaire

Type	**Title**	**Questions**
Subjective Wellbeing (SWB)	Personal Wellbeing Index (PWI)	9
	Integrated Household Survey (IHS)	5
	Satisfaction with Life Survey (SWLS)	4
	Subtotal	**18**
Multi Attribute Utility (MAU) Instruments	EQ-5D	5
	AQoL-4D [(1)] and AQoL-8D	44
	HUI 3	8
	15D	15
	QWB-[SA (2)]	77
	SF-6D (derived from SF-36)	
Non-Utility	SF-36	36
	Self TTO	1
	ICECAP-A [(3)]	5
Demographics		18
	Total items in composite instrument	**227**

[(1),(2),(3)] not used in Norway; [(2)] not used in Germany

Box 3 Respondents by disease and country

Diseases	Australia (1)	UK (2)	USA (3)	Canada (4)	Norway (5)	Germany (6)	Total (1-6)
Asthma	141	150	150	138	130	147	856
Cancer	154	137	148	138	80	115	772
Depression	146	158	168	145	140	160	917
Diabetes	168	161	168	144	143	140	924
Hearing loss	155	126	156	144	115	136	833
Arthritis	163	159	179	139	130	159	929
Heart disease	149	167	170	154	151	152	943
COPD	66	x	x	x	x	x	66
Stroke	23	x	x	x	x	x	23
Disease sample	1165	1058	1139	1002	889	1009	6262
No disease- 'Healthy public'	265	298	321	328	288	260	1760
Total	1430	1356	1460	1330	1177	1269	8022

Box 4 Sources of utility weights[1]

Instrument	Country and Respondents	Method of Calibration	Reference
EQ-5D-5L	US Public n=3691	TTO	Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-crosswalk-value-sets.html
SF6D	US Public n=611	SG	Brazier, J, Roberts J, Deverill M: The estimation of a preference-based measure of health From the SF-36. J Health Econ. 2002 mar;21(2)271-92
HUI3	Canada Public n= 256	SG	Furlong W, Feeny D, Torrance GW, et al. Multiplicative Multi-Attribute Utility Function for the Health Utilities Index Mark 3 (HUI3) System: A Technical Report, McMaster University Centre for Health Economics and Policy Analysis Working Paper No. 98-11, December 1998.
15D	Finland Public n=1255	VAS	Brazier, J., Ratcliffe, J., Salomon, JA. and Tsuchiya, A. (2007):'Measuring and Valuing Health Benefits for Economic Evaluation' Oxford University Press, page 195. http://www.15d-instrument.net/15d
QWB	USA Public n=435	VAS	Sieber W, Groessl E, David K, Ganiats T, Kaplan R. (2008): Quality of Well Being Self-Administered (QWB-SA) Scale, User's Manual, Health Services Research Centre, University of California, San Diego. https://hoap.ucsd.edu/qwb-info/QWB-Manual.pdf
AQoL-4D	Australia Public n=350	TTO	Hawthorne, G., Richardson, J., Day, N., Osborne, R., McNeil, H.(2000) Construction and Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument. Monash University Centre for Health Economics Working paper 101. http://www.buseco.monash.edu.au/centres/che/pubs/wp101.pdf
AQoL-8D	Australia Public =347 Patient =323 n=670	TTO	Richardson J, Iezzi A: Psychometric validity and the AQoL-8D Multi Attribute Instrument. Research Paper 71 (2011). Centre for Health Economics, Monash University, Australia http://www.buseco.monash.edu.au/centres/che/pubs/researchpaper71.pdf

[1] Choice of weights is also discussed in Section 8.

Box 5 List of abbreviations

MA	Multi attribute
MAU	Multi attribute utility
MAUI	Multi attribute utility instrument
SWB	Subjective wellbeing (‘happiness’)
CUA	Cost Utility Analysis

TABLE OF CONTENTS

List of Boxes

List of Tables

List of Figures

Cross-national comparison of twelve quality of life instruments: MIC Paper 7 Germany

1 Introduction

Objectives

The background and objectives of the MIC project are described in MIC Paper 1 (Richardson, Iezzi et al. 2012). In sum, the project is a response to the evidence that different MAU instruments produce different values for 'utility' and (despite the common label 'utility') measure different constructs. The principle objectives of the project are, firstly, to document the differences in the values produced by the instruments for different groups of patients in different countries; and, secondly, to determine what the different instruments measure – which dimensions of wellbeing explain variation in instrument scores.

To achieve these objectives we sought respondents with a diverse range of health states and, specifically, health states associated with major disease areas. This implies that the total sample is not representative of the population as the focus of the study is the relationships between instruments in different health states and not the wellbeing of the overall population. Despite this, comparisons may be made with population or other instrument norms. 'Patients' complete a disease-specific questionnaire for which there are norms and the non-patient sample may be weighted to correct for any mismatch between them and independently obtained norms if population values are needed.

The primary objectives relate to the content and validity of existing instruments, ie those which are currently used for cost utility analysis (CUA). While the investigation of the psychometric properties of the instruments are a further area of inquiry the main research, including results reported in this paper, use unadjusted MAU instruments irrespective of their reliability as indicated by the present data. The instruments are currently used irrespective of their properties.

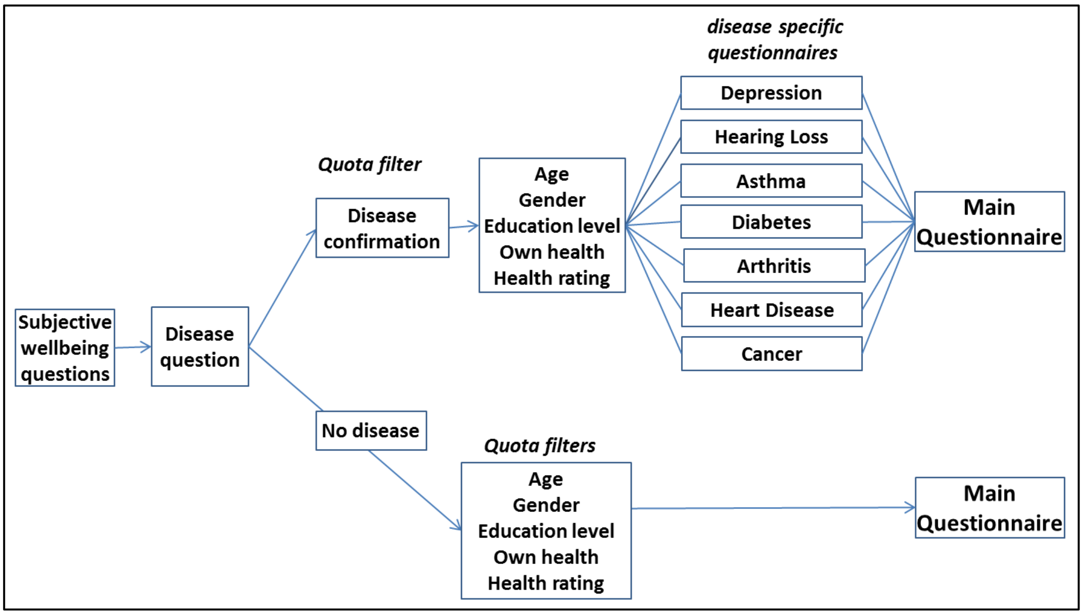
The administration of the MIC survey is illustrated in Figure 1. A survey company, CINT, invited individuals on their database to participate. A person accepting this invitation was first asked to complete the three subjective wellbeing questions: the Personal Wellbeing Index (PWI), the Integrated Household Survey (IHS) and the Satisfaction with Life Survey (SWLS). These questions were administered immediately as they seek to measure 'affect' – a person's 'undigested' feelings. Asking the questions after 'priming' respondents with questions about their health (do you have one of the eight diseases of interest?) would potentially create biased responses.

After completion of these questions the respondent was asked the following question: 'Have you got a current diagnosis of any of the following health problems? Please choose the most serious illness you have.'

Those nominating one of the survey diseases proceeded with the survey if and only if the quota – the target number of respondents – had not been reached. To confirm the patient's status the first question was a repetition of the question above. Patients then completed the core questionnaire which was administered to all respondents within the quota. This was followed by the disease-specific questionnaire which applied to their particular disease.

Those who did not report a disease were questioned about their age, gender and education. Additionally they were asked to indicate their overall health on a visual analogue scale (VAS) where *'Zero is the least desirable state of health you could imagine and 100 is the best possible health (physical, mental and social).'* The individual was invited to proceed to the core questions only if their VAS score exceeded 70 and their age, gender and education quota had not been filled. The VAS criterion was included to ensure that the 'healthy public' excluded those whose self rating was very poor. The web-based procedure employed here attracts a disproportionate number of distressed respondents and the procedure was adopted to reduce this effect and increase the sample size of respondents in good health. The number 70 was selected judgementally to achieve this goal but to permit variation in 'normal health'.

Figure 1 Administration of the MIC online questionnaires



Editing

Introductory comments from the panel company to their panellists were designed to deter unreliable respondents. The edit procedures used are detailed below.

(a) German Public

Table 1.1 Edit procedures German public

No completed		344
Deletions due to duplicated ID		0
Deleted	Edit 1	15
	Edit 2	0
	Edit 3	52
	Edit 4	0
	Edit 5	17
Total after edits		**260 (84 deleted)**

(337 respondents agreed to the 12 month follow up. After deletions, this left 253 agreeing to the 12 month follow up.)

Edit 1: Any responses that were done in less than 15 minutes were eliminated. The survey median completion time was 24.3 mins (range 7.4–135.6 minutes). Times between 15–20 minutes were marked to inspect at the final edit.

Edit 2: Inclusion into the survey was predicated on a VAS rating of 70 and above.

Edit 3: Respondents were removed from the dataset if they indicated that they suffered from any of the disease states that will be studied in the PATIENT rounds (asthma, cancer, COPD, depression, diabetes, hearing, osteoarthritis or stroke).

Edit 4: EQ5D question 4 (pain) and AQoL-8D question 22 (pain) answers were compared. Anyone with 2 response levels difference was eliminated.

(b) German Disease

Table 1.2 Edit procedures German disease

No completed		1216
Deletions due to mismatching disease choices		56
Deleted	Edit 1	53
	Edit 2	12
	Edit 3	8
	Edit 6	65
	Edit 7	11
Missing data		2*
Total after edits		**1009 (207 deleted)**

Two response sets were deleted from analysis as the SF36 and AQoL8D were missing from these sets, although participants completed the survey.

Edit 1: Any responses that were done in less than 15 minutes were eliminated. The survey median completion time was 30.6 minutes, mean completion time was 34 minutes (range 8.8–168.6 minutes). Times between 15–20 minutes were marked to inspect at the final edit.

Edit 2: The EQ-5D mobility question was duplicated in the survey. Anyone with a response that varied by more than+/-1 difference was eliminated. Those differing by only +/- 1 were examined with other criteria to determine their eligibility.

Edit 3: The SF-36 question 1 and question concerning Own Health were identical. Those with responses greater than +/- 1 were eliminated. Those without identical answers but within +/- 1 were earmarked.

Edits 4 and 5: were not possible as the QWB was not included in the German study.

Edit 6: EQ5D question 4 (pain) and AQoL-8D question 22 (pain) answers were compared. Anyone with 2 response levels difference was eliminated.

Edit 7: The number of inconsistencies from edits 2, 3, 4, 5 and 6 was coded. Anyone with 2 or more inconsistencies and a time less than 25 minutes was eliminated.

Edit 8: Anyone with 3 inconsistencies from edits 2, 3, 4, 5 or 6 was eliminated.

Table 1.3 German disease sample size

German disease	Sample size	Percent
Asthma	147	14.6
Cancer	115	11.4
Depression	160	15.9
Diabetes	140	13.9
Hearing loss	136	13.5
Arthritis	159	15.8
Heart disease	152	15.1
Total	1009	100.0

Utility weights

Utility weights for all instruments are not available for all countries. Box 4 reports the weights used in the initial analysis with the project. In principle the use of alternative weights for different countries may alter results. This is discussed further in Section 8 which presents a comparison of US and UK weights for the EQ-5D data from the MIC project. It does not suggest that the explanatory power of the EQ-5D could alter with a choice between these weights.

2 Respondent characteristics

The healthy public

After conclusion of the edit procedures outlined above 1269 respondents were retained, 1009 patients' and 260 representing the 'healthy public'. Table 2.1 shows the distribution of respondents by age and gender. The highest level of education of the public respondents by gender is reported in Table 2.2.

Patient samples

1009 patient surveys were retained. The focus of the study is upon the comparison of instruments and the purpose of the patient samples was primarily to maximise the diversity of health states in the sample. Consequently, no age-gender quotas were used. Table 2.3 disaggregates respondents by age, gender and disease group. It indicates that the overall sample is highly skewed with respect to age reflecting the increasing probability of a chronic disease with age.

Table 2.1 'Healthy Public': Age and gender

Age group	Germany Public		Total
	Male (%)	Female (%)	
18-24	4.6	8.5	17
25-34	16.8	23.3	52
35-44	18.3	18.6	48
45-54	21.4	25.6	61
55-64	19.1	15.5	45
65+	19.8	8.5	37
Total (%)	50.4	49.6	260

Table 2.2 Healthy public: Highest education by gender

Education	Germany Public		Total
	Male (%)	Female (%)	
High school	29	26.4	72
Some post-secondary, post-secondary certificate or diploma	55.7	59.7	150
University degree and higher	15.3	14	38
Total	50.4	49.6	260

Table 2.3 Distribution of total sample by age and gender

Disease	Distribution of diseases by age group and gender												Total		
	18-24		25-34		35-44		45-54		55-64		65+				
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	T
Asthma	10	10	7	14	28	20	15	17	8	9	7	2	75	72	147
Cancer	0	0	3	7	11	9	9	19	16	10	24	7	63	52	115
Depression	4	12	10	20	16	26	21	28	11	9	1	2	63	97	160
Diabetes	3	3	4	4	8	6	32	15	21	16	21	7	89	51	140
Hearing problems	3	3	1	5	12	10	28	23	21	13	17	0	82	54	136
Arthritis	0	0	2	4	11	8	33	44	20	22	9	6	75	84	159
Heart problems	2	3	2	2	9	3	19	17	45	23	23	4	100	52	152
No disease-Healthy public	6	11	22	30	24	24	28	33	25	20	26	11	131	129	260
Total	28	42	51	86	119	106	185	196	167	122	128	39	678	591	1269

3 Summary statistics

Mean values

Summary statistics for the twelve instruments are reported in Tables 3.1 and 3.2. MAU instruments purport to measure the same construct – utility. Consequently, direct comparison of their scores is appropriate. Other instruments may not be directly compared. The PWI, SWLS and IHS all measure facets of subjective wellbeing (SWB), however, they do not purport to measure the same construct and their correlation reflects this (see Table 4.4).

Differences between patient groups are not the principle focus of the present report.

Frequency distributions for each of the instruments are reported in Appendices 1 and 2.

Table 3.1 Summary statistics for the MAU instruments (Public n=260)

	EQ-5D(1)	HUI3	SF-6D	15D	AQoL-4D(2)	AQoL-8D(3)
Mean	0.89	0.88	0.82	0.95	0.82	0.88
N	260	260	260	260	260	260
SE	0.01	0.01	0.01	0.00	0.01	0.01
SD	0.11	0.13	0.11	0.06	0.18	0.11
Minimum	0.49	0.24	0.54	0.58	0.16	0.40
Maximum	1.00	1.00	1.00	1.00	1.00	1.00

(1) Kind et al. (1999)

(2) Hawthorne et al. (2012)

(3) Richardson et al. (2012)

Table 3.2 Summary statistics for the MAU instruments (Total n=1269)

	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
Mean	0.73	0.7	0.71	0.85	0.63	0.73
N	1269	1269	1269	1269	1269	1269
SE	0.006	0.007	0.004	0.004	0.007	0.006
SD	0.225	0.267	0.143	0.128	0.266	0.217
Minimum	-0.24	-0.28	0.3	0.39	-0.04	0.07
Maximum	1.00	1.00	1.00	1.00	1.00	1.00
Score			%			
1	18.4	6.2	1.8	6.9	6	2.2
0.95+	18.4	11.1	3.4	25.5	9.1	13
<0.4	9.5	15.3	1.3	0.3	21.9	10.2
<0.1	1.6	4.1	0.0	0.0	4.3	0.2
<0.0	0.7	1.7	0.0	0.0	0.3	0.0

Internal reliability

A test of scale reliability was carried out with public data using the Cronbach's alpha (Cronbach 1951). This determines the internal consistency or average correlation of items in a survey instrument. The reliability of a scale can vary depending on the sample that it is used with. Table 3.3 reports the alpha coefficient. If this is above 0.7, the scale can be considered reliable with the sample (Pallant 2010). The result shows that all of the scales pass this test except for the IHS and SF-36.

Table 3.3 Reliability of instruments

Instrument	No of items	Cronbach's Alpha
AQoL-4D	12	0.82
AQoL-8D	35	0.96
HUI3	8	0.74
EQ-5D	5	0.82
15D	15	0.88
ICECAP	5	0.84
SF-36	36	**0.68***
IHS	4	**0.47***
SWLS	5	0.92

* These values are below those generally accepted as indicating the reliability of a scale.

Figure 3.1 Mean of MAU instruments (Total n= 1269)

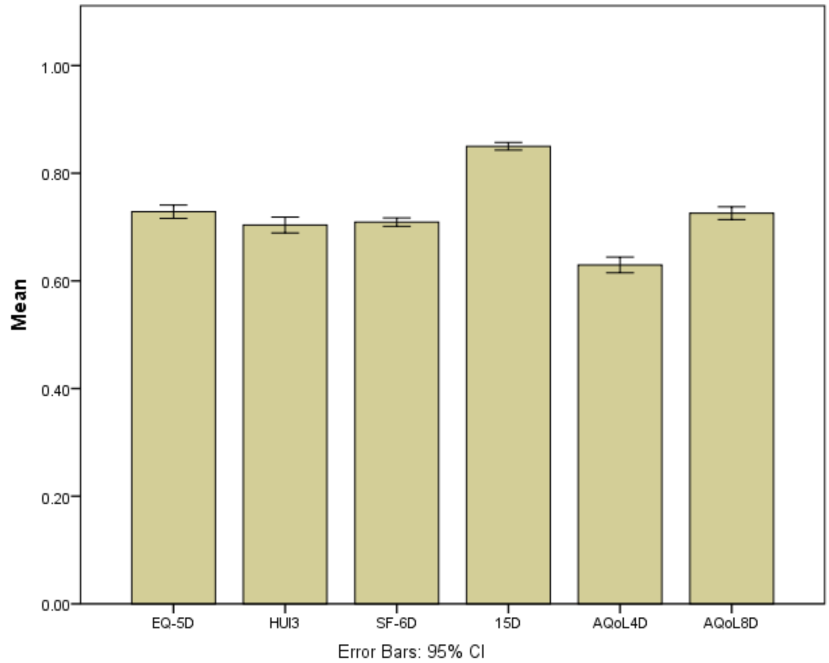
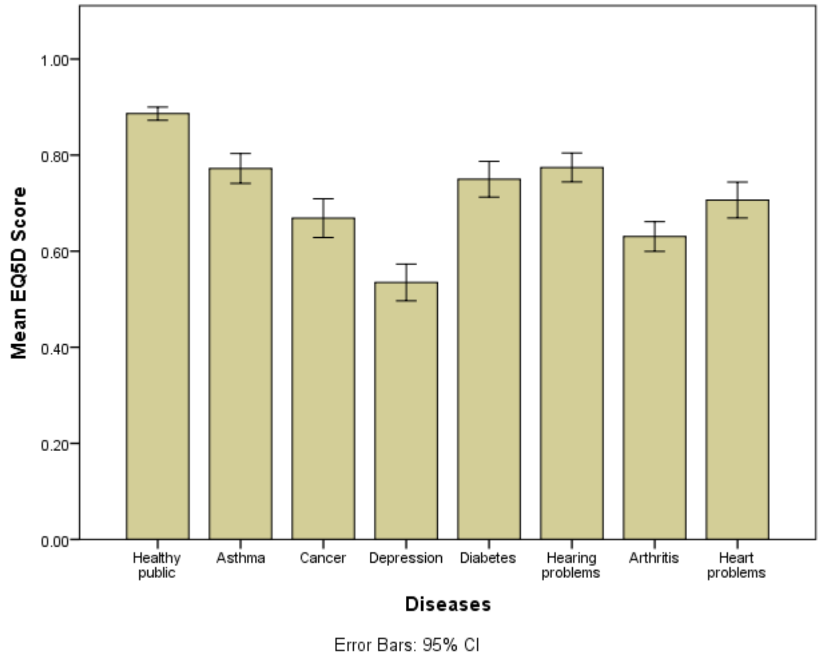


Figure 3.2 Mean EQ-5D by disease group (Total n=1269)



4 Correlation

Validation tests draw heavily upon correlation. In particular, convergent validity is established if an instrument correlates as predicted with other instruments or criteria scores which are believed to correlate with the construct. Higher correlation justifies greater confidence in overall validity. The MIC project collected several types of data to test convergent validity. These were:

1. *Other MAU scores*. As each MAU instrument is believed to reflect ‘utility’, the instruments can ‘cross validate’. Confidence in one MAU instrument increases when it correlates with the other MAU instruments.
2. *Subjective Wellbeing (SWB) score*. Utility is commonly equated with SWB. This is not strictly correct as people’s preferences do not always maximise happiness (Richardson, Maxwell et al. 2012). However the two constructs are undoubtedly related and high correlation with SWB is independently important if MAU instruments are to influence policy decisions. The three instruments used here – PWI, SWLS and IHS – are outlined in MIC Research Paper 1 (Richardson, Iezzi et al. 2012).
3. *Self TTO*. The concept and measurement of self TTO are also explained in MIC Research Paper 1. It is conceptually the same as a conventional TTO except that the health state evaluated is not ‘external’ as described to the respondent, but the respondent’s own health state. The relationship between self and conventional TTO is the subject of ongoing research (Richardson, Iezzi et al. forthcoming).
4. *Disease-specific QoL instruments*. These are not utilised in the current report.

The Pearson correlation between MAU instruments, between MAU and non-MAU instruments and between non-MAU instruments are reported in Tables 4.1-4.5 and Fig 4.1-4.8. The Pearson correlation indicates the extent to which changes in one variable correspond with changes in another. It does not indicate that two variables are the same or even the same order of magnitude. The low correlation between measures of utility and PWI and SWLS is in need of explanation. While preferences may differ from subjective wellbeing (SWB) their correlation might be expected to be higher than observed here.

The better measure of this is the intraclass correlation (ICC). This is reported in Table 4.6 and Figure 4.2. The difference is parenthesised by the relative score for the 15D. This has the highest average Pearson correlation but (reflecting significant differences in its predicted utility scores) it has the lowest ICC.

Overall the ICC reflects a poorer correspondence between instruments than the Pearson correlation. The imperfect correspondence is also illustrated by the use of R^2 coefficients in Figure 4.3 rather than Pearson correlation coefficients ($R^2 = \rho^2$). This is because a complete explanation of variation would imply $R^2 = 1$. The extent to which the R^2 falls short of 1.00 indicates the extent to which variance is explained by some unknown variable or variables.

Table 4.1 Pearson correlation between MAUI (Public n=260)

	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
EQ-5D	1	.649**	.595**	.654**	.530**	.514**
HUI3	.649**	1	.515**	.649**	.540**	.522**
SF-6D	.595**	.515**	1	.569**	.450**	.648**
15D	.654**	.649**	.569**	1	.558**	.597**
AQoL-4D	.530**	.540**	.450**	.558**	1	.623**
AQoL-8D	.514**	.522**	.648**	.597**	.623**	1
Ave	0.59	0.58	0.56	0.61	0.54	0.58

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.2 Pearson correlation between MAUI (Total n=1269)

	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
EQ-5D	1	.805**	.774**	.817**	.767**	.789**
HUI3	.805**	1	.720**	.837**	.784**	.816**
SF-6D	.774**	.720**	1	.783**	.749**	.806**
15D	.817**	.837**	.783**	1	.788**	.846**
AQoL-4D	.767**	.784**	.749**	.788**	1	.842**
AQoL-8D	.789**	.816**	.806**	.846**	.842**	1
Ave	0.79	0.79	0.77	0.81	0.79	0.82

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.1 Average Pearson correlation with other MAU Instruments (Total n=1269)

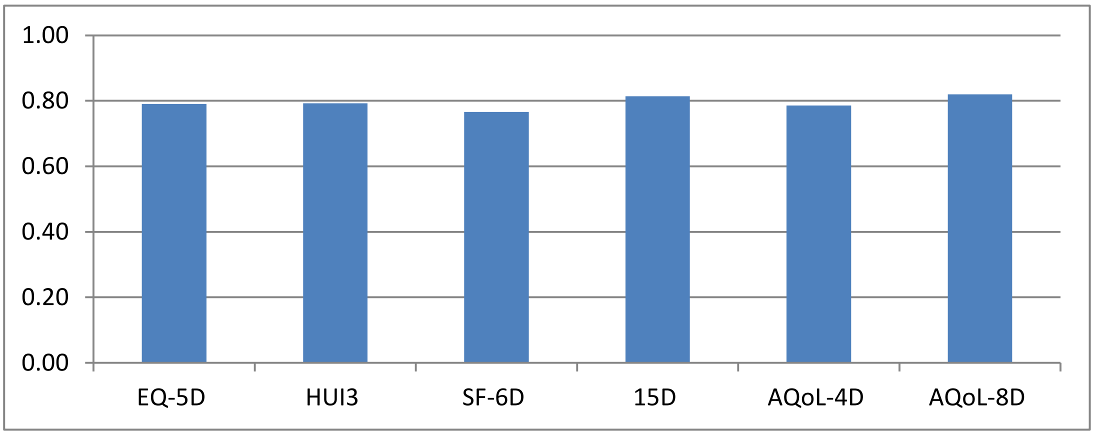


Table 4.3a Pearson correlations between MAUI and non-MAU instruments (Public n=260)

Instrument	PWI Sum	PWI	SWLS	IHS	ICECAP	Self-TTO	SF-36
EQ-5D	.223**	.233**	.298**	.214**	.291**	.202**	.626**
HUI3	.361**	.314**	.363**	.320**	.451**	.216**	.586**
SF-6D	.280**	.241**	.329**	.266**	.449**	.171**	.874**
15D	.180**	.253**	.286**	.203**	.400**	.285**	.629**
AQoL-4D	.328**	.357**	.455**	.379**	.542**	.193**	.508**
AQoL-8D	.471**	.457**	.577**	.519**	.686**	.163**	.641**

*. Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 4.3b Pearson correlations between MAUI and non-MAU instruments (Total n=1269)

Instrument	PWI Sum	PWI	SWLS	IHS	ICECAP	Self-TTO	SF-36
EQ-5D	.452**	.457**	.448**	.356**	.604**	.363**	.810**
HUI3	.517**	.518**	.502**	.417**	.664**	.355**	.763**
SF-6D	.489**	.495**	.504**	.393**	.637**	.371**	.934**
15D	.484**	.501**	.506**	.405**	.678**	.361**	.835**
AQoL-4D	.521**	.546**	.541**	.453**	.720**	.365**	.777**
AQoL-8D	.603**	.624**	.639**	.540**	.803**	.418**	.838**

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.2 Pearson correlation of MAU instrument with PWI (Total n=1269)

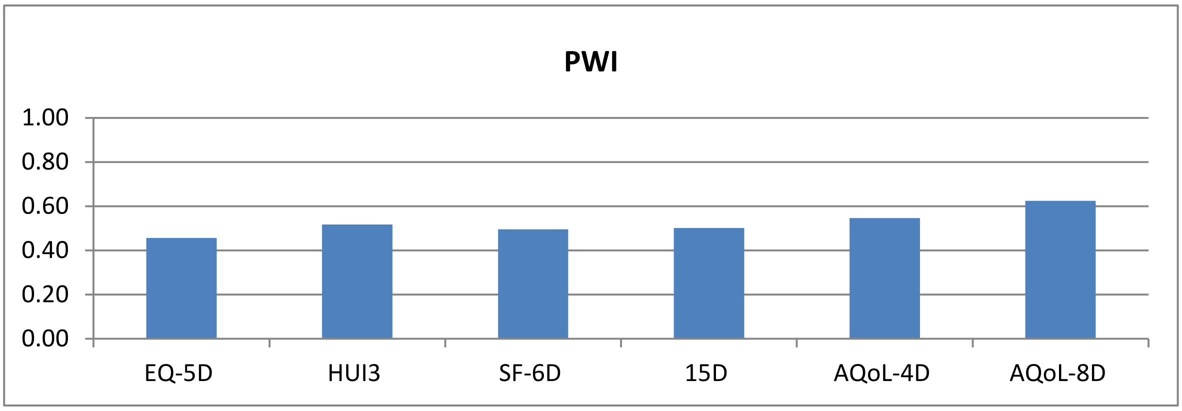


Figure 4.3 Pearson correlation of MAU instrument with SWLS (Total n=1269)

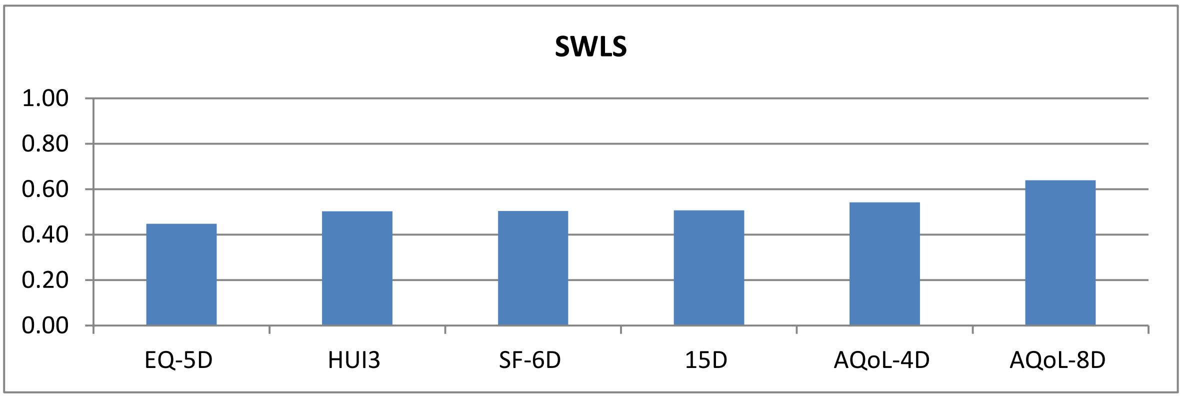


Figure 4.4 Pearson correlation of MAU instrument with Self-TTO (Total n=1269)

Figure 4.5 Pearson correlation of MAU instrument with SF-36 (Public n=260)

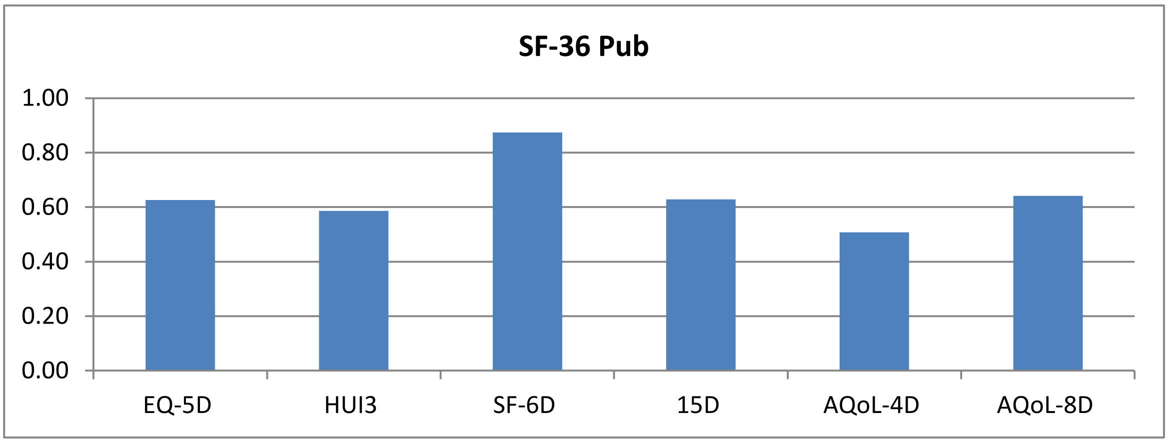


Figure 4.6 Pearson correlation of MAU instrument with SF-36 (Total n=1269)

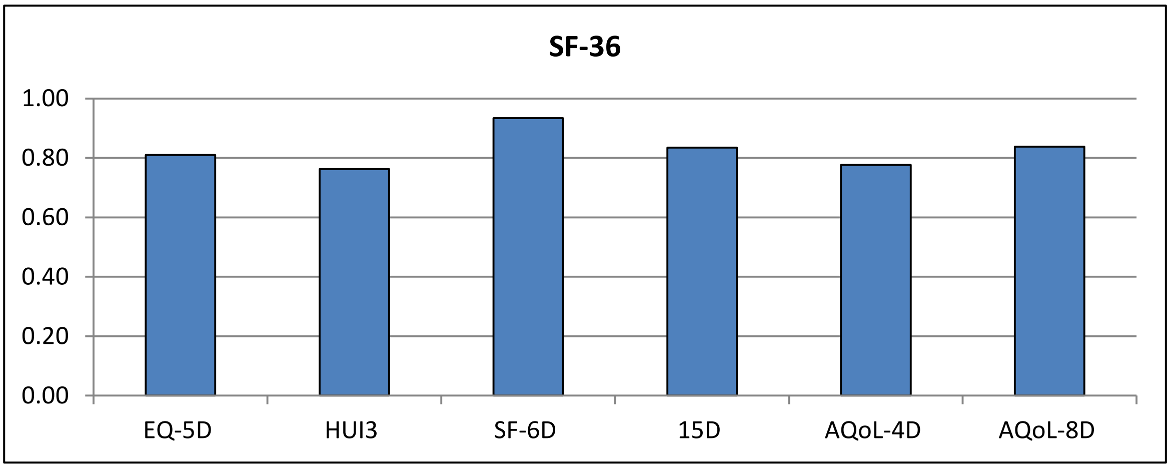


Table 4.4 Pearson correlations between non-MAU instruments (Public n=260)

Instrument	PWI Sum	PWI	SWLS	IHS	ICECAP	Self-TTO	SF-36
PWI Sum	1	.788**	.658**	.651**	.479**	0.067	.252**
PWI	.788**	1	.628**	.617**	.473**	0.117	.264**
SWLS	.658**	.628**	1	.799**	.615**	0.109	.347**
IHS	.651**	.617**	.799**	1	.583**	0.077	.254**
ICECAP	.479**	.473**	.615**	.583**	1	.131*	.423**
Self-TTO	0.067	0.117	0.109	0.077	.131*	1	.155*
SF-36	.252**	.264**	.347**	.254**	.423**	.155*	1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.5 Pearson correlations between non-MAU instruments (Total n=1269)

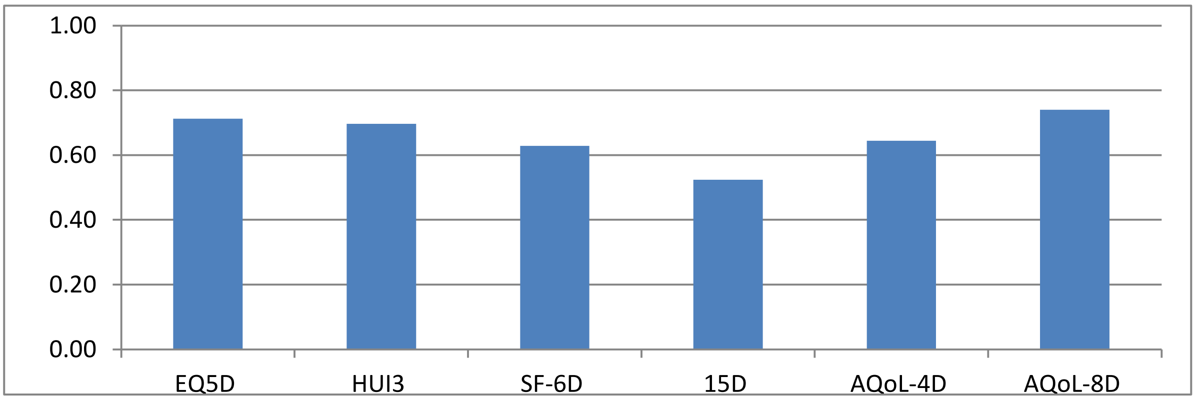
Non-MAUI	PWI Sum	PWI	SWLS	IHS	ICECAP	Self-TTO	SF-36
PWI Sum	1	.832**	.728**	.708**	.592**	.277**	.504**
PWI	.832**	1	.740**	.729**	.621**	.287**	.518**
SWLS	.728**	.740**	1	.788**	.658**	.292**	.529**
IHS	.708**	.729**	.788**	1	.584**	.276**	.408**
ICECAP	.592**	.621**	.658**	.584**	1	.346**	.664**
Self-TTO	.277**	.287**	.292**	.276**	.346**	1	.394**
SF-36	.504**	.518**	.529**	.408**	.664**	.394**	1

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4.6 Intraclass correlation between MAU instrument (Total n=1269)

Instrument	EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
EQ5D		0.79	0.7	0.58	0.7	0.79
HUI3	0.79		0.6	0.53	0.76	0.8
SF-6D	0.7	0.6		0.51	0.59	0.74
15D	0.58	0.53	0.51		0.4	0.6
AQoL-4D	0.7	0.76	0.59	0.4		0.77
AQoL-8D	0.79	0.8	0.74	0.6	0.77	
Ave	**0.71**	**0.7**	**0.63**	**0.52**	**0.64**	**0.74**

Figure 4.7 Average Intraclass correlation with other MAU Instruments (Total n=1269)



5 Linear relationships

The MAU instruments were designed for use in cost utility analyses (CUA) in which, typically, utilities are measured before and after an intervention. This implies that it is the change in measured utilities, not their absolute values, which are important for validity. The comparative performance of the different instruments in this respect is not identified by either Pearson or intraclass correlations. It is however, easily measured with linear regression.

If instrument X is the criterion variable then the validity of the change predicted by instrument Y may be tested by the magnitude of the b coefficient in the linear relationship Y = a + bX. The absence of bias implies that b = 1.00. In the present case there is no criterion variable. However as with correlation, 'cross validation' may increase confidence: confidence rises if the b coefficients of an instrument are close to 1.00 in the linear relationships with the other MAU instruments. A technical problem which arises with this test is that, because both measured variables in the comparison are subject to error, the parameters will be sensitive to the choice of dependent and independent variable in OLS regressions. One solution to the problem is to use Geometric Mean Squares (GMS) regression. This is obtained by regressing Y on X then X on Y and deriving parameters from the geometric mean of the two regressions. Results are independent of the choice of dependent and independent variable. This technique was used in the present study.

Figure 5.1 reproduces the 15 pairwise GMS regressions, their scattergrams and the two GMS equations (Y on X; X on Y) using public data. Figure 5.2 gives the same results using the total sample.

Table 5.2 employs the corresponding results for the total sample (n=1269) to derive an average deviation away from b = 1 for each of the 6 regressions which include a particular MAUI. Depending upon the choice of left and right hand scale variable, 'b' may be greater than or less than 1.00. For consistency, the GMS regression was selected where b > 1. Thus from Figure 5.2 the linear relationship between the EQ-5D and HUI 3 for all respondents may be expressed either as (1) EQ-5D = 0.135 + 0.843 HUI 3 or as (2) HUI 3 = 0.158 + 1.186 EQ-5D. Table 5.1 reports the b coefficient which is greater than 1.00 which, in this case, is 1.186 (1.19).

Table 5.1 reports the b coefficient from pairwise regressions using abbreviations (eg H = 1.19 EQ). From the bottom row in Table 5.1 the deviation for the MAUI vary from 34.0 percent (AQoL-8D) to 74.6 percent (15D). If these linear relationships were generally true (and not just for the present sample) the results would imply that the choice of AQoL-8D rather than one of the other six instruments would result in a 34 percent discrepancy in measured change. The choice of the 15D rather than one of the other six instruments would result in a 74.6 percent discrepancy.

Table 5.2 presents a different comparison using b coefficients. The bottom left of the table reports the b coefficients when instrument B is the left hand variable in the regression and instrument A is the right hand variable. The first figure is derived from the public regression and the second figure from the total sample. (Thus, in the public regression EQ-5D = 0.135 + 0.852 HUI 3 (Figure 5.1), the reported b coefficient is 0.852 rounded to 0.85. The b coefficient for the total sample, Figure 5.2 is 0.84. The difference between these coefficients is given in parentheses in Table 5.2. This is an indicator of the stability of the linear relationships involving an instrument when the severity of the health state changes. Thus for example, between the two samples the average of the 6 coefficients in equations with the EQ-5D as the dependent variable change by 19 percent.

Figure 5.1 Geometric regression results (Public n=260)

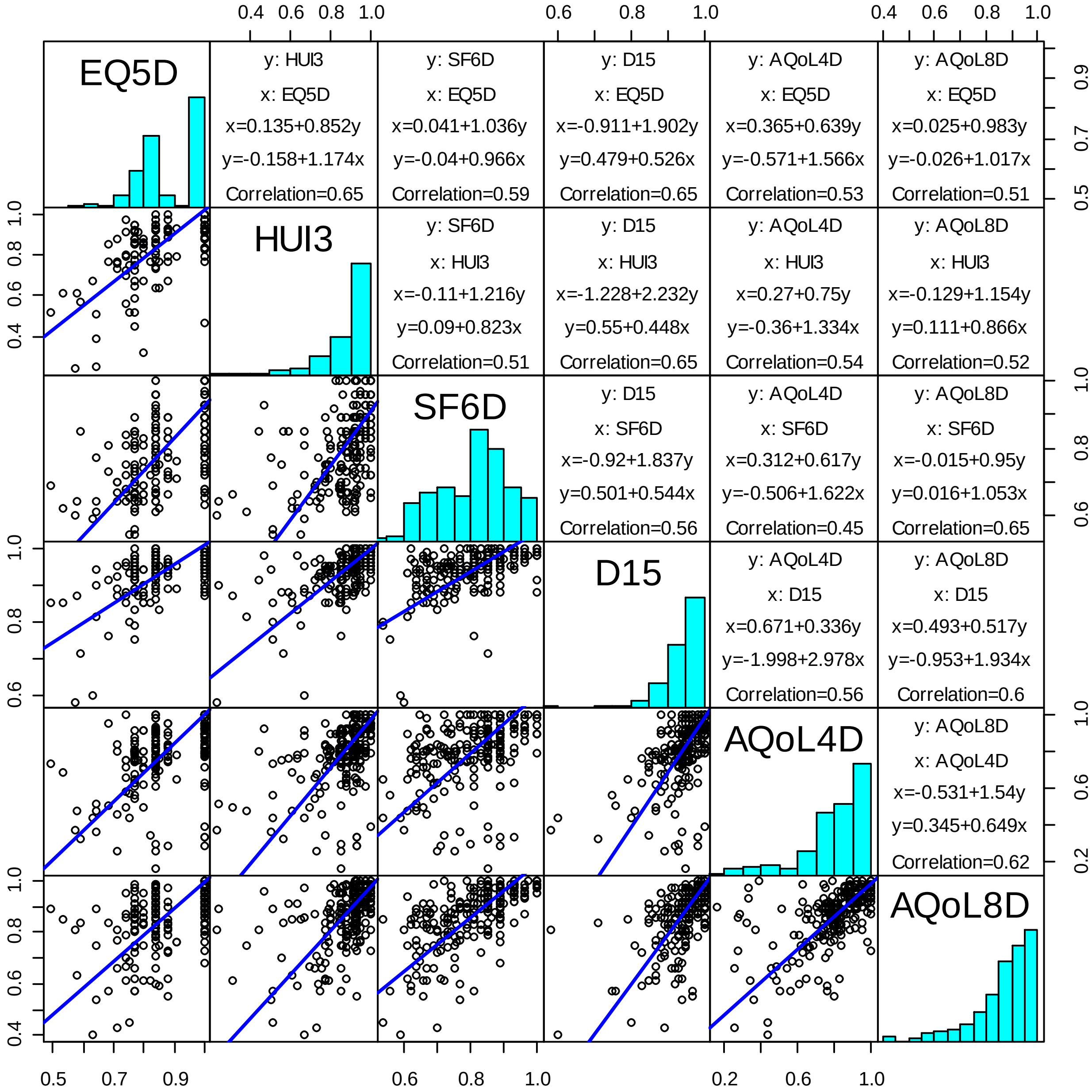


Figure 5.2 Geometric regression results (Total n=1269)

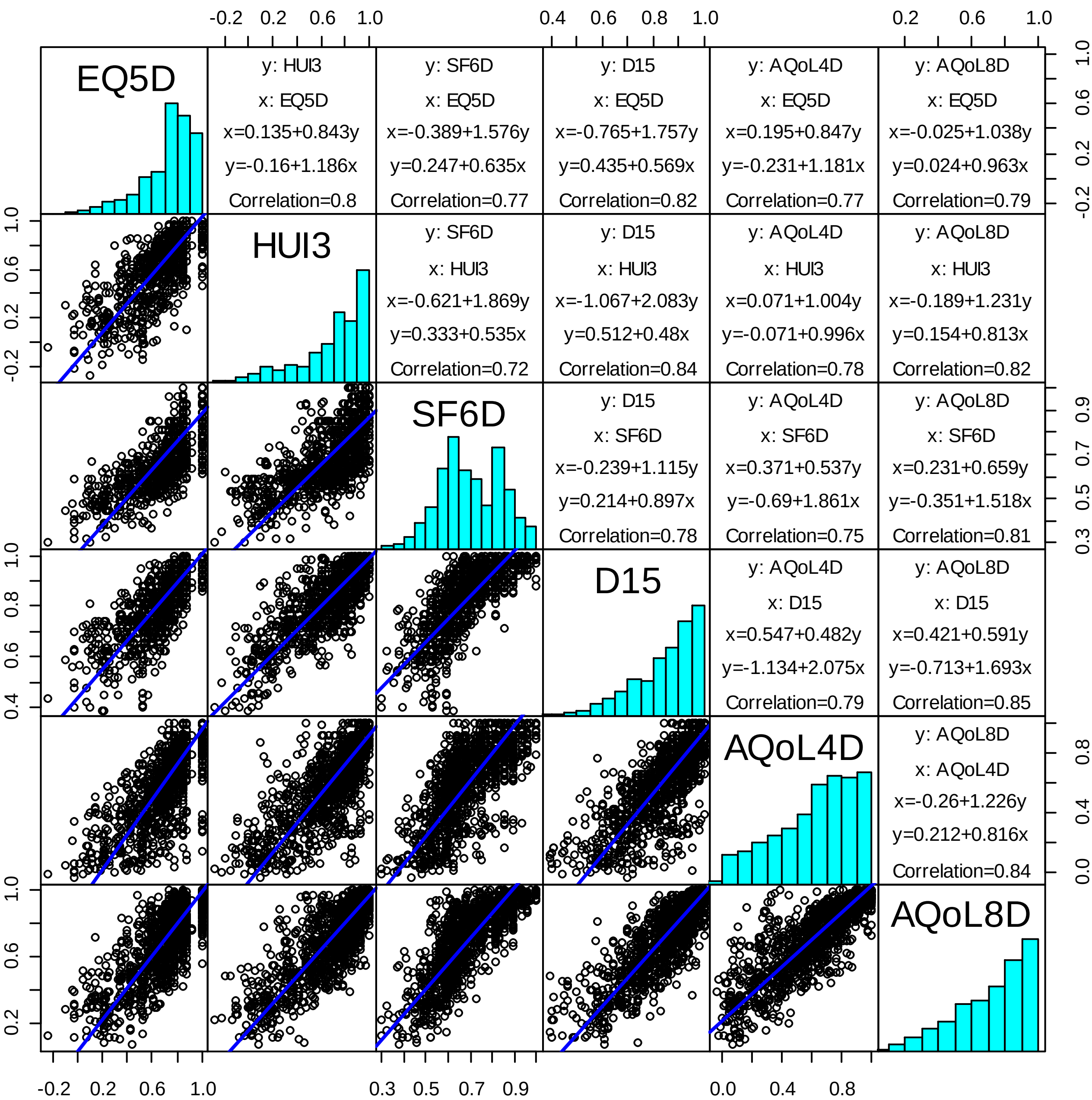


Table 5.1 Discrepancies in marginal change: slope, coefficient, b, in regression (Total n=1269)

(Instrument A=a+b instrument B)*

Instrument	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
EQ-5D (EQ)	1.00					
HUI3 (H)	H=1.19(EQ)	1.00				
SF-6D (SF)	EQ=1.58(SF)	H=1.87(SF)	1.00			
15D (D)	EQ=1.76(D)	H=2.08(D)	SF=1.12(D)	1.00		
AQoL-4D (A4)	A4=1.18(EQ)	H=1.00(A4)	A4=1.86(SF)	A4=2.08(D)	1.00	
AQoL-8D (A8)	EQ=1.04(A8)	H=1.23(A8)	A8=1.51(SF)	A8=1.69(D)	A4=1.23(A8)	1.00
Ave % Diff	35.0	47.4	58.8	74.6	47.0	34.0

(NB Constant terms in the equations have been dropped)

*Equations arranged to obtain b>1 as a consistent index of deviation (Geometric Mean Regressions permit this)

Table 5.2 Difference in marginal change: Public vs Total (instrument A=a+b instrument B)

Instrument B	EQ5D		HUI3		SF6D		15D		AQoL-4D		AQoL-8D	
	Pub	Tot ***(Diff)***	Pub	Tot ***(Diff)***	Pub	Tot ***(Diff)***	Pub	Tot ***(Diff)***	Pub	Tot	Pub	Tot
EQ-5D	1.00											
HUI3	.85	.84 ***(.01)***	1.00									
SF-6D	1.04	1.58 ***(.54)***	1.22	1.87 ***(.65)***	1.00							
15D	1.90	1.76 ***(.14)***	2.23	2.08 ***(.15)***	1.84	1.12 ***(.72)***	1.00					
AQoL-4D	.64	.85 ***(.21)***	.75	1.00 ***(.25)***	.62	.54 ***(.08)***	.34	.48 ***(.14)***	1.00			
AQoL-8D	.98	1.01 ***(.03)***	1.15	1.23 ***(.08)***	.95	.66 ***(.29)***	.52	.59 ***(.07)***	1.54	1.23 ***(.31)***	1.00	

6 Instrument content (sensitivity)

Each MAU defines a 'construct'. Results in this section seek to identify how clearly related dimensions of health/wellbeing are to the MAU constructs. Conversely the results seek to determine how sensitive the MAU constructs are to the dimensions. The dimensions used in the study are obtained from the SF-36 and AQoL-8D which have been independently shown to have construct validity (Richardson, Elsworth et al. 2011). Additionally, the widely used and validated SWB instruments, the PWI and SWLS are employed as is the yet unvalidated Self TTO. Similar results may be obtained for the IHS.

Ceiling effects: From Table 6.1a ceiling effects differ greatly. In the public sample the maximum score (the 'ceiling') was obtained by 42.7 percent of the EQ-5D and 5.4 percent of SF-6D and AQoL-8D. Amongst the 111 respondents with an EQ-5D score of 1.00 the average scores on other instruments varied from 0.88 for SF-6D and AQoL-4D respectively to 0.98 for 15D.

Floor effects: Table 6.1c reveals similar differences in floor effects. For example, when EQ-5D < 0.4 its average score is 0.22. HUI 3, SF-6D, 15D and AQoL-8D have average scores of 0.23, 0.50, .64 and 0.36 respectively. When HUI 3 < 0.4 average values for EQ-5D,HUI3, SF-6D and AQoL-8D are 0.39, 0.19, 0.53 and 0.40 respectively.

Table 6.1a Ceiling effects Average value of other MAUI when a MAU=1.0 (Public n=260)

MAU =1	Average value							
	EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D	N	(%)
EQ5D	..	0.95	0.88	0.98	0.9	0.94	111	42.7
HUI3	0.99	..	0.88	0.99	0.95	0.95	42	16.2
SF-6D	0.99	0.95	..	0.98	0.95	0.98	14	5.4
15D	0.98	0.97	0.89	..	0.95	0.96	51	19.6
AQoL-4D	0.98	0.97	0.89	0.99	..	0.96	42	16.2
AQoL-8D	1	0.98	0.94	0.99	0.91	..	14	5.4

Table 6.1b Ceiling effects Average value of other MAUI when a MAU=1.0 (Total n=1269)

MAU =1	Average value							
	EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D	N	(%)
EQ5D	..	0.95	0.88	0.98	0.9	0.94	234	18.4
HUI3	0.98	..	0.87	0.98	0.93	0.95	79	6.2
SF-6D	0.99	0.96	..	0.98	0.95	0.98	23	1.8
15D	0.98	0.95	0.88	..	0.94	0.95	87	6.9
AQoL-4D	0.98	0.96	0.88	0.98	..	0.96	76	6
AQoL-8D	0.98	0.97	0.92	0.99	0.93	..	25	2

Table 6.1c Floor effects Average value of other MAUI when a MAU=<0.40 (Total n=1269)

MAU <.40	Average value							
	EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D	N	(%)
EQ5D	0.22	0.23	0.5	0.64	0.21	0.36	120	9.5
HUI3	0.39	0.19	0.53	0.65	0.25	0.4	194	15.3
SF-6D	0.23	0.14	0.36	0.59	0.1	0.26	17	1.3
15D	0.26	-0.11	0.51	0.4	0.09	0.28	4	0.3
AQoL-4D	0.47	0.38	0.57	0.7	0.22	0.45	278	21.9
AQoL-8D	0.31	0.19	0.5	0.62	0.18	0.29	130	10.2

Correlation with summary measures: Table 6.2 and Figure 6.1 report the correlation between MAU scores and the physical and psycho-social summary scores derived from the SF-36 and AQoL-8D. With one exception the correlations with the AQoL-8D (non-utility) super-dimensions are greater than with the SF-36 summary scores. With two exceptions correlation between MAU instruments and the physical summary score is greater than with the psycho-social summary scale. The Table suggests three groups of instruments. First, EQ-5D, HUI, SF-6D and 15D are relatively very sensitive to physical health (particularly EQ-5D). AQoL-8D is relatively very sensitive to psycho-social health. AQoL-4D is between these polar cases.

Table 6.2 Correlation of instruments with SF-36, AQoL-8D physical and psycho-social scales (Total 1269)

SF-36/AQoL-8D dimension	EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
SF-36	.810**	.763**	.934**	.835**	.777**	.838**
PCS	.705**	.593**	.658**	.650**	.565**	.513**
MCS	.536**	.570**	.748**	.621**	.620**	.763**
AQoL-8D	.789**	.816**	.806**	.846**	.842**	1
PSD	.829**	.791**	.752**	.804**	.782**	.793**
MSD	.618**	.644**	.752**	.704**	.741**	.893**

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 6.1 Correlation with summary scores of SF-36 (PCS and MCS) and AQoL-8D (PSD and MSD)

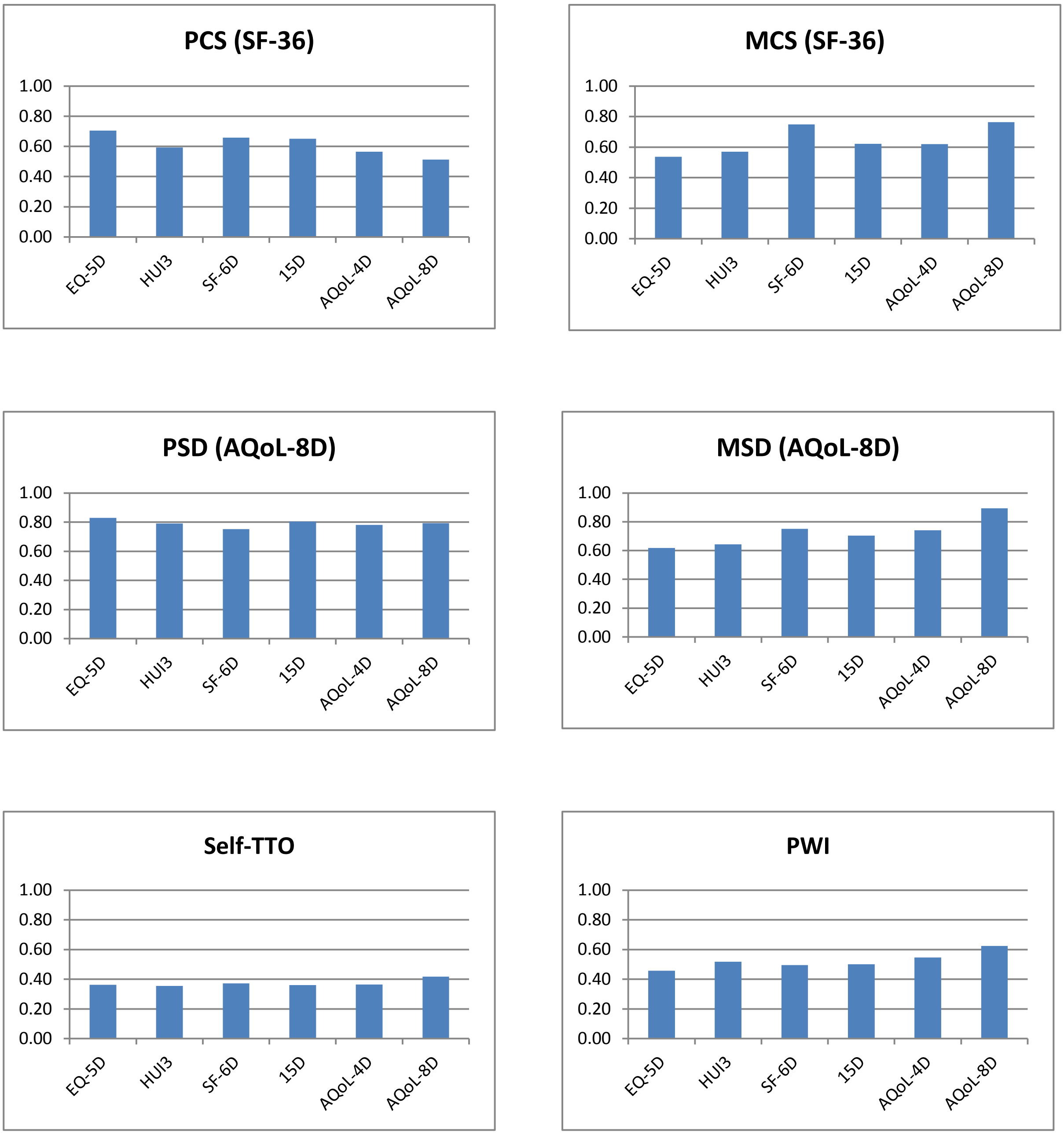
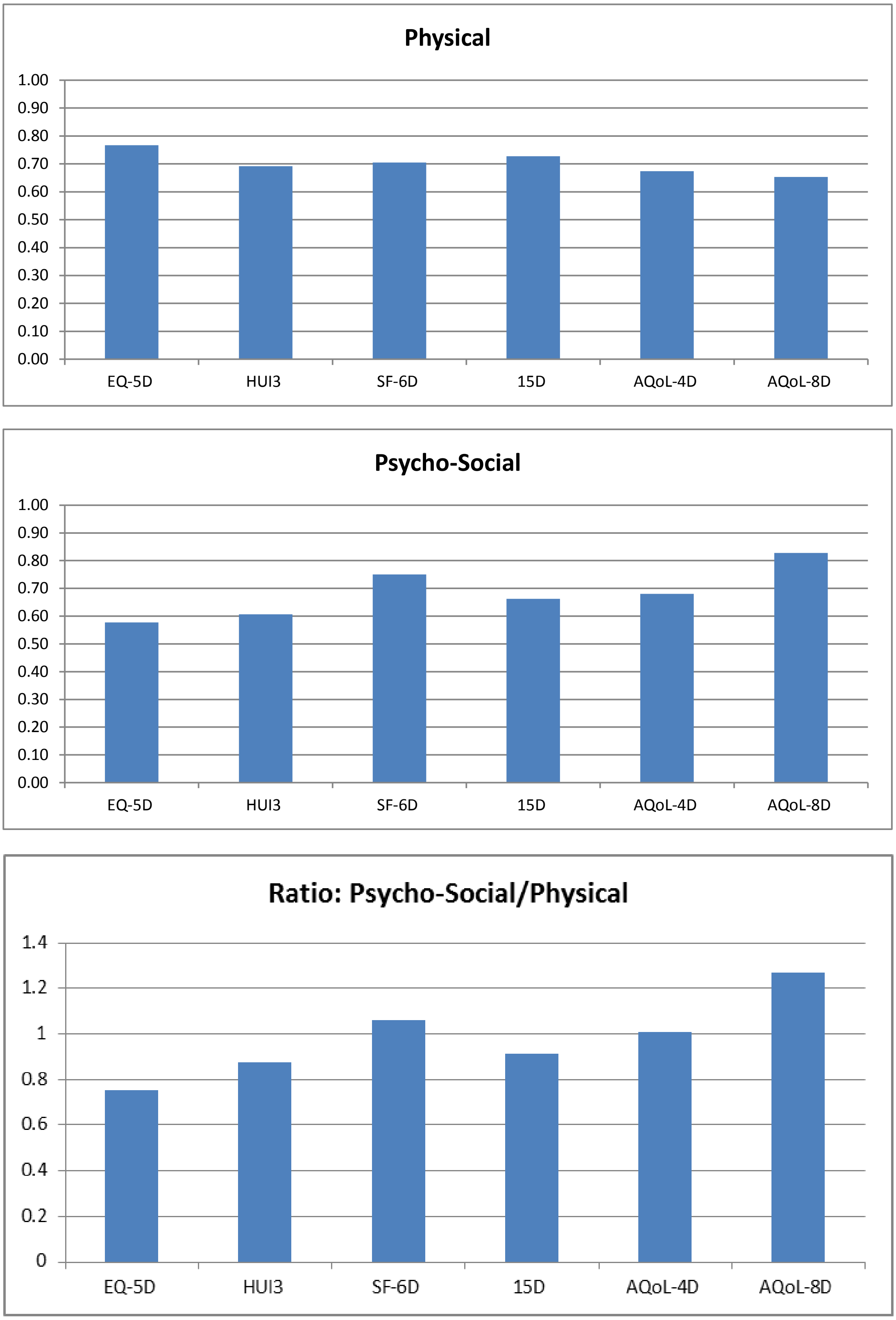


Figure 6.2 Comparison of Summary Physical and Psycho-Social Dimensions (Average SF-36 and AQoL-8D summary scores)



Split half analysis: Table 6.3 reports results from a comparison of two split halves of the full sample. Each MAU was used, in turn, to rank observations on the basis of which they were divided into a top and bottom half. Dimension and SWB scores were calculated for both halves. The table reports the ratio of these scores. Higher ratios indicate greater sensitivity of an instrument to a dimension or SWB.

Sensitivity to dimensions: Tables 6.4a, 6.4b; 6.5a, 6.5b and Figure 6.3a, 6.3b report beta coefficients from the regression of MAU scores on dimension scores. The coefficients show the change in the MAU score with a one standard deviation change in the dimension score. MAU scores are measured in standard deviations (of the MAU score) to allow comparison of sensitivity. This avoids the confusion of a large standard deviation with instrument sensitivity. Thus, for example, the 15D compresses scores. But this is offset in the calculation of beta coefficients by a correspondingly small standard deviation. A larger beta coefficient suggests greater sensitivity.

Tables 6.4a and 6.5a report results from regressions with a single explanatory variable. Because of its correlation with other explanatory variables (dimensions) interpretation of the beta score is ambiguous. Table 6.3b and 6.4b use multiple regressions to obtain the standardised beta. In principle this means that the beta coefficients represent the effect of the dimension after standardising for other dimensions in the regression. From the regressions employing the SF-36 dimensions (Table 6.4b) a one sd increase in each dimension would result in a 1.06 sd increase in the EQ-5D of which 70 percent would be attributable to physical function and pain. Mental health would contribute 18 percent and vitality 7 percent. The same increase in the dimension scores would increase AQoL-8D by 1.08 sd of which 37 percent would be attributable to mental health, 27 percent to vitality and only 12 percent to pain and 13 percent to physical function. This suggests that in the AQoL-8D the effects of pain and physical function may be largely mediated through psycho-social factors.

The percentage contribution to total change following a one sd increase in every dimension using data from Tables 6.4a,b and 6.5b is shown in the pie charts, Figure 6.4.

Table 6.3a Ratio of scores in top and bottom 50% of total sample, ranked by MAUI (SF-36 dimensions)

Ranking MAUI	SF-36 dimensions									
	GH	PF	RP	BP	VT	SF	RE	MH	PCS	MCS
EQ5D	1.64	1.52	3.16	1.78	1.68	1.46	2.41	1.41	1.38	1.31
HUI3	1.57	1.44	2.67	1.62	1.65	1.47	2.19	1.43	1.31	1.33
SF-6D	1.62	1.46	3.94	1.7	1.85	1.68	3.68	1.53	1.32	1.49
15D	1.69	1.5	3.21	1.66	1.8	1.5	2.53	1.45	1.35	1.36
AQoL-4D	1.63	1.44	2.77	1.61	1.7	1.51	2.35	1.45	1.31	1.36
AQoL-8D	1.63	1.38	2.59	1.56	1.86	1.55	2.63	1.56	1.25	1.46

Table 6.3b Ratio of scores in top and bottom 50% of total sample, ranked by MAUI (AQoL-8D dimensions, SWB and Self-TTO)

Ranking MAUI	AQoL-8D dimensions										Non-MAUI		
	IL	Hap	MH	Cop	Rel	SW	Pain	Sen	PSD	MSD	PWI	SWLS	Self-TTO
EQ5D	1.29	1.25	1.32	1.26	1.26	1.24	1.35	1.1	1.5	1.9	1.31	1.3	1.31
HUI3	1.26	1.27	1.33	1.27	1.27	1.25	1.31	1.13	1.47	1.95	1.34	1.34	1.3
SF-6D	1.26	1.28	1.37	1.28	1.3	1.28	1.29	1.1	1.43	2.09	1.32	1.32	1.35
15D	1.29	1.27	1.35	1.29	1.29	1.26	1.3	1.12	1.47	2.05	1.33	1.33	1.38
AQoL-4D	1.26	1.29	1.35	1.28	1.36	1.28	1.29	1.15	1.48	2.15	1.35	1.35	1.31
AQoL-8D	1.26	1.37	1.49	1.35	1.42	1.35	1.28	1.14	1.45	2.64	1.43	1.45	1.35

Key

GH=general health; PF = physical functioning; RP = role limit physical; BP = bodily pain; VT = vitality; SF = social functioning; RE = role limit emotional; MH = mental health; PCS =physical component summary; MCS = mental component summary; IL = independent living; Hap = happiness; Cop = coping; Rel = relationships; SW = self worth; Pain=pain; Sen=senses; MSD = mental super dimension; PSD = physical super dimension; PWI = Personal Wellbeing Index; SWLS = Satisfaction with Life Survey; TTO = Time- trade-off

Sensitivity to dimensions: Tables 6.4a, 6.4b; 6.5a, 6.5b and Figure 6.3a, 6.3b report beta coefficients from the regression of MAU scores on dimension scores. The coefficients show the change in the MAU score with a one standard deviation change in the dimension score. MAU scores are measured in standard deviations (of the MAU score) to allow comparison of sensitivity. This avoids the confusion of a large standard deviation with instrument sensitivity. Thus, for example, the 15D compresses scores. But this is offset in the calculation of beta coefficients by a correspondingly small standard deviation. A larger beta coefficient suggests greater sensitivity.

Tables 6.4a and 6.5a report results from regressions with a single explanatory variable. Because of its correlation with other explanatory variables (dimensions) interpretation of the beta score is ambiguous. Table 6.3b and 6.4b use multiple regressions to obtain the standardised beta. In principle this means that the beta coefficients represent the effect of the dimension after standardising for other dimensions in the regression. From the regressions employing the SF-36 dimensions (Table 6.4b) a one sd increase in each dimension would result in a 1.06 sd increase in the EQ-5D of which 67 percent would be attributable to physical function and pain. Mental health would contribute 17 percent and vitality 7 percent. The same increase in the dimension scores would increase AQoL-8D by 1.09 sd of which 39 percent would be attributable to mental health, 16 percent to vitality and only 13 percent to pain and 14 percent to physical function. This suggests that in the AQoL-8D the effects of pain and physical function may be largely mediated through psycho-social factors.

The percentage contribution to total change following a one sd increase in every dimension using data from Tables 6.4a,b and 6.5b is shown in the pie charts, Figure 6.4.

Table 6.4a Sensitivity to SF-36 dimensions: Beta coefficient and R^2 from the regression of MAU on single dimensions of the SF-36 (Total n=1269)

$$(MAU = a + b\ Dim)$$

(SF-36 dimension)		EQ-5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
(GH)	Beta	0.65	0.59	0.69	0.68	0.61	0.65
	R^2	0.42	0.35	0.47	0.46	0.38	0.42
(PF)	Beta	0.73	0.66	0.67	0.72	0.62	0.59
	R^2	0.53	0.44	0.45	0.52	0.39	0.35
(RP)	Beta	0.62	0.57	0.75	0.64	0.57	0.57
	R^2	0.39	0.33	0.56	0.41	0.33	0.33
(BP)	Beta	0.76	0.65	0.75	0.66	0.62	0.63
	R^2	0.58	0.42	0.56	0.43	0.38	0.4
(VT)	Beta	0.65	0.61	0.78	0.71	0.65	0.77
	R^2	0.43	0.37	0.61	0.5	0.43	0.6
(SF)	Beta	0.64	0.63	0.8	0.67	0.69	0.73
	R^2	0.41	0.4	0.64	0.46	0.47	0.54
(RE)	Beta	0.56	0.56	0.74	0.62	0.58	0.66
	R^2	0.31	0.31	0.55	0.39	0.33	0.44
(MH)	Beta	0.62	0.65	0.77	0.68	0.67	0.82
	R^2	0.39	0.42	0.59	0.47	0.45	0.66
(PCS)	Beta	0.71	0.59	0.66	0.65	0.57	0.51
	R^2	0.5	0.35	0.43	0.42	0.32	0.26
(MCS)	Beta	0.54	0.57	0.75	0.62	0.62	0.76
	R^2	0.29	0.33	0.56	0.39	0.38	0.58

Table 6.4b Sensitivity to SF-36 dimensions: Beta coefficient from the regression of MAU on all dimensions of the SF-36 (Total n=1269)

$$(\boldsymbol{MAU} = \boldsymbol{a} + \sum_{\boldsymbol{u-1}}^{\boldsymbol{8}} \boldsymbol{b1Dimi})$$

(SF-36 dimension)		EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
(GH)	Beta	.03 (ns)	.03 (ns)	.01 (ns)	0.11	0.08	0.08
	t				5.22	3.23	4.09
(PF)	Beta	0.33	0.33	0.11	0.35	0.24	0.15
	t	15.85	13.48	7.72	16.69	9.63	7.96
(RP)	Beta	-.03 (ns)	-.02 (ns)	.14	.01 (ns)	-.01 (ns)	-.06
	t			9.35			-2.8
(BP)	Beta	0.38	0.22	0.25	0.11	0.15	0.14
	t	17.87	8.74	17.61	4.96	5.97	7.33
(VT)	Beta	0.07	-.03 (ns)	.15	0.13	.05 (ns)	0.17
	t	2.82		9.06	5.14		7.25
(SF)	Beta	0.07	0.09	0.18	0.06	0.19	0.1
	t	2.94	3.06	11.57	2.7	6.95	4.55
(RE)	Beta	.03 (ns)	.04 (ns	0.17	0.09	.04 (ns)	0.09
	t			11.37	4.09		4.47
(MH)	Beta	0.18	0.34	0.18	0.22	0.26	0.42
	t	7	10.94	10.39	8.37	8.57	17.63
R^2		**0.75**	**0.65**	**0.89**	**0.75**	**0.65**	**0.79**
F		**473**	**291**	**1265**	**460**	**292**	**603**

ns = not significant

1 Same as Table a

2 Direct comparison of the overall fit with the fit of SF-6D is invalid as it is derived from the SF-36

Table 6.5a Sensitivity to AQoL-8D dimensions: Beta coefficient and R2 from the regression of MAU on single dimensions of the AQoL-8D

$$(MAU = a + b\ Dim)$$

(AQoL-8D dimension)		EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
(IL)	Beta	0.77	0.75	0.71	0.79	0.71	0.74
	R^2	0.59	0.56	0.51	0.63	0.50	0.55
(Hap)	Beta	0.60	0.65	0.65	0.65	0.69	0.87
	R^2	0.36	0.43	0.43	0.43	0.47	0.75
(MH)	Beta	0.60	0.62	0.70	..67	0.67	0.85
	R^2	0.36	0.38	0.49	0.45	0.45	0.72
(Cop)	Beta	0.66	0.68	0.72	0.74	0.71	0.88
	R^2	0.44	0.47	0.52	0.55	0.51	0.78
(Rel)	Beta	0.51	0.56	0.63	0.60	0.70	0.77
	R^2	0.26	0.31	0.40	0.36	0.48	0.60
(SW)	Beta	0.61	0.66	0.68	0.69	0.69	0.88
	R^2	0.37	0.43	0.46	0.48	0.48	0.77
(Pain)	Beta	0.80	0.67	0.67	0.66	0.64	0.66
	R^2	0.64	0.45	0.45	0.43	0.41	0.44
(Sen)	Beta	0.38	0.51	0.37	0.52	0.53	0.53
	R^2	0.14	0.26	0.14	0.27	0.28	0.28
(PSD)	Beta	0.83	0.79	0.75	0.80	0.78	0.79
	R^2	0.69	0.63	0.57	0.65	0.61	0.63
(MSD)	Beta	0.62	0.64	0.75	0.70	0.74	0.89
	R^2	0.38	0.42	0.57	0.50	0.55	0.80

Table 6.5b Sensitivity to AQoL-8D dimensions: Beta coefficient from the regression of MAU on all dimensions of the AQoL-8D

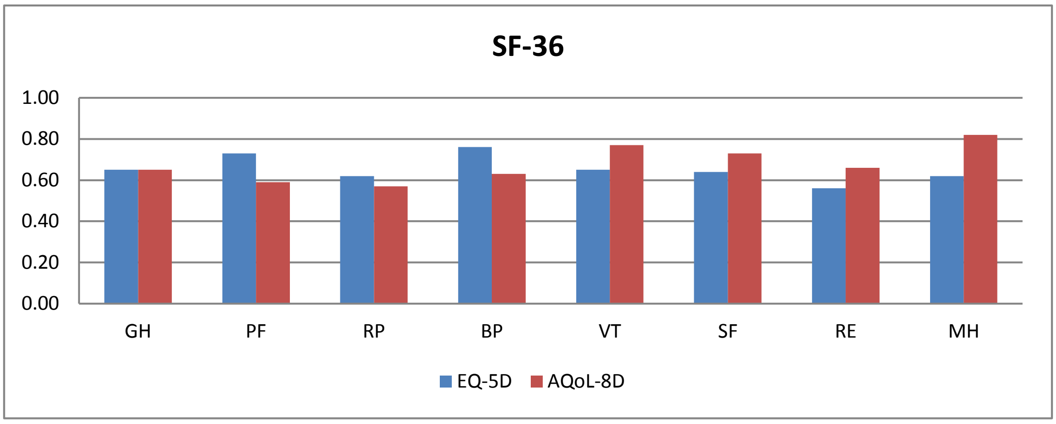
$$MAU = a + \sum_{u-1}^{8} b1Dimi)$$

(AQoL-8D dimension)		EQ5D	HUI3	SF-6D	15D	AQoL-4D	AQoL-8D
(IL)	Beta	0.29	0.28	0.24	0.36	0.18	0.09
	t	14.32	12.22	10.36	17.4	8.12	18.08
(Pain)	Beta	0.48	0.28	0.27	0.18	0.25	0.19
	t	27.62	14.02	13.39	10.19	13.27	45.46
(Sen)	Beta	.00 (ns)	0.17	-.01 (ns)	.15	0.19	0.14
	t		10.39		10.03	12.34	38.99
(Hap)	Beta	0.07	0.17	-.02 (ns)	-.01 (ns)	.10	0.18
	t	2.55	5.95			3.7	29.09
(MH)	Beta	0.05	.00 (ns)	0.2	0.12	.00 (ns)	0.15
	t	2.28		7.54	4.88		27.24
(Cop)	Beta	0.09	0.04 (ns)	0.14	0.17	0.06	0.18
	t	3.43		4.59	6.36	2.26	27.78
(Rel)	Beta	-0.05	-.01(ns)	.12	.01 (ns)	0.25	0.07
	t	-2.27		4.97		11.27	14.66
(SW)	Beta	0.12	0.15	0.1	0.14	0.11	0.25
	t	4.74	5.38	3.63	5.43	4.25	41.42
R^2		**0.79**	**0.73**	**0.72**	**0.78**	**0.76**	**0.99**
F		**597**	**422**	**413**	**566**	**494**	**12603**

ns = not significant
1 Beta coefficients are the change in the dependent variable, measured in standard deviations (of the dependent) when the independent variable changes by one standard deviation (after standardising for other variables in the regression). They allow direct comparison of the importance of independent variables.
2 Direct comparison of the overall fit with the fit of AQoL-8D is invalid as it is an (exponential) function of the dimensions

Figure 6.3 Effect of SD change in dimension on standardised score (beta coefficient)

(a) Content of EQ-5D vs AQoL-8D (SF-36 Dimensions)



(b) Contrast of EQ-5D vs 15D (AQoL-8D Dimensions)

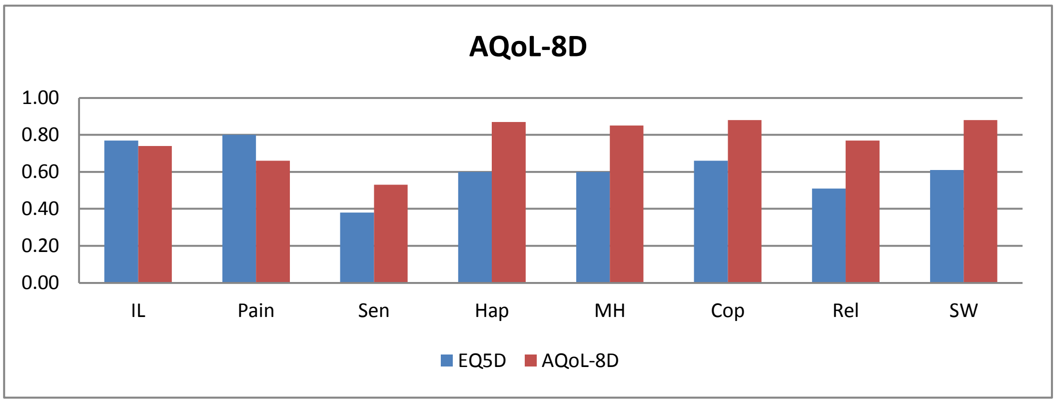


Table 6.6 Instrument content: regression of MAU on non-MAU instruments

Dependent	EQ5D	HUI3	SF-6D	15D	QWB	AQoL-4D	AQoL-8D	
MAU = a+b (PWI)								
a	0.42	0.29	0.5	0.66		0.2	0.32	PWI
b	0.51	0.68	0.35	0.32		0.72	0.67	
Beta	0.46	0.52	0.5	0.5		0.55	0.62	
R^2	0.21	0.27	0.25	0.25		0.3	0.39	
F	331	465	411	425		538	810	
MAU = a+b (SWLS)								
a	0.43	0.3	0.49	0.66		0.2	0.31	SWLS
b	0.49	0.65	0.35	0.32		0.7	0.67	
Beta	0.45	0.5	0.5	0.51		0.54	0.64	
R^2	0.2	0.25	0.25	0.26		0.29	0.41	
F	319	426	432	435		523	873	
MAU = a+b (Self-TTO)								
a	0.55	0.5	0.6	0.75		0.42	0.53	Self-TTO
b	0.23	0.27	0.15	0.13		0.27	0.26	
Beta	0.36	0.36	0.37	0.36		0.37	0.42	
R^2	0.13	0.13	0.14	0.13		0.13	0.17	
F	192	182	202	189		195	268	

Figure 6.4 Instrument content: Disaggregated by AQoL-8D dimensions

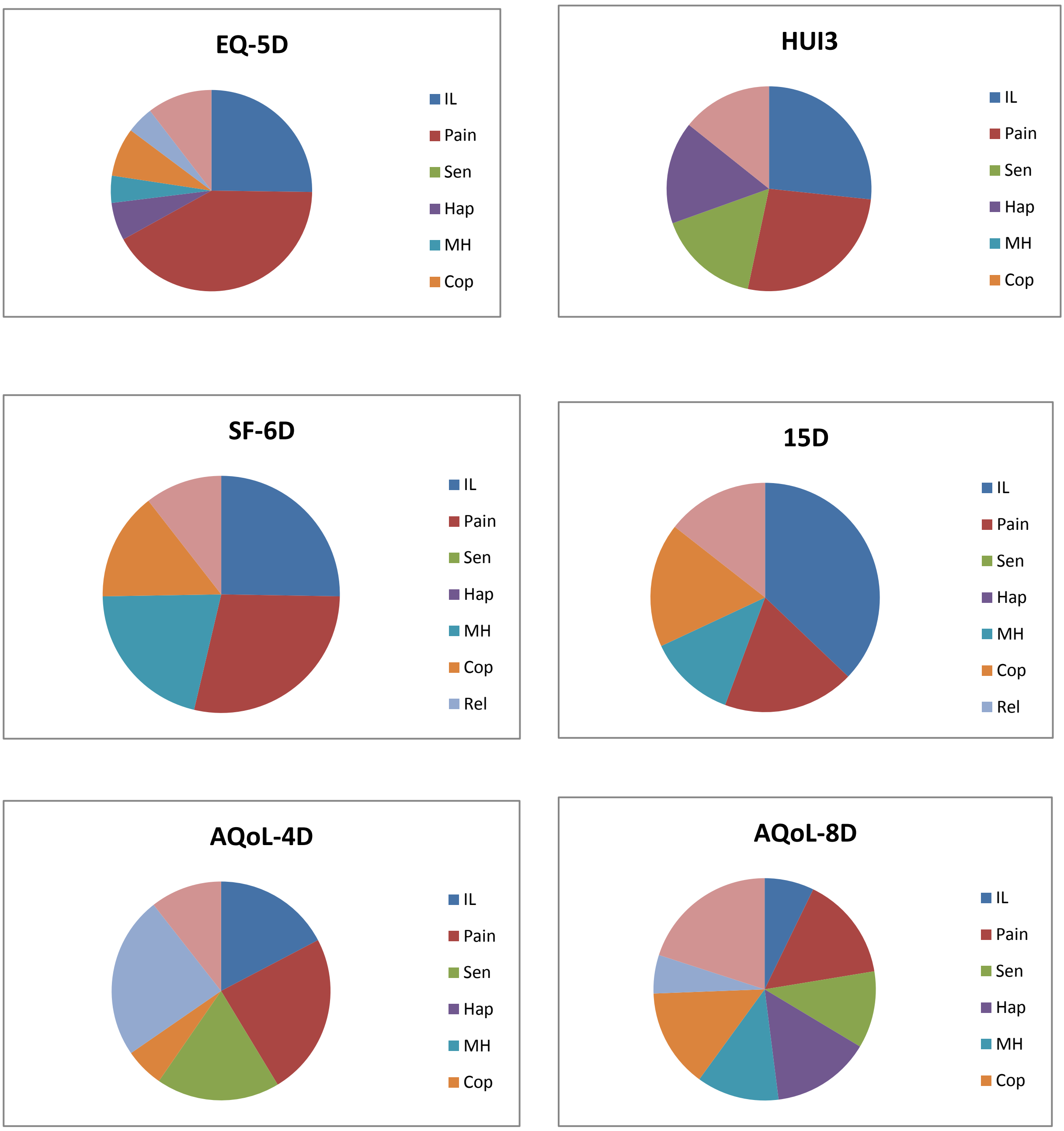
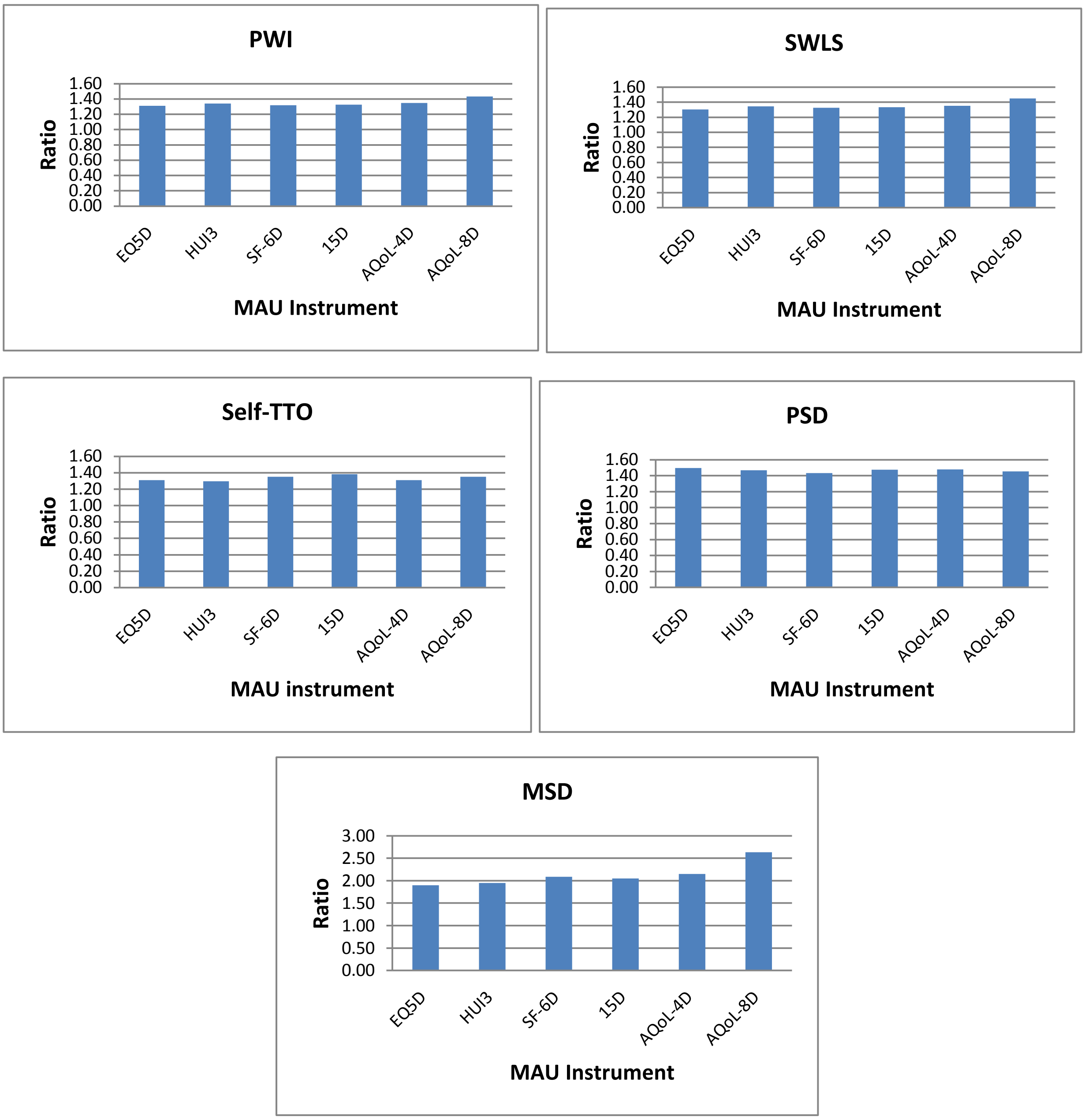


Figure 6.5 Split half analysis: Ratios of values in top/bottom half of population ranked by instrument



7 Pairwise comparison of instruments

The GMS regressions reported earlier were employed to help explain differences between the instruments' content. The residual from the regression of one instrument upon another was correlated with each of the major dimensions and non-MAU instruments. A positive correlation between the residual of Y regressed upon X and a dimension, D or index, I, indicates a greater sensitivity of the instrument Y to dimension D or index I.

Figure 7.1 presents the correlation results from Table 7.1 and 7.2. Table 7.3 summarises the results and therefore the implications of the data for the relative sensitivity of instruments.

A negative correlation implies the greater sensitivity of instrument X. Since regressions were calculated using geometric mean squares the results are independent of the choice of dependent and independent variable.

Results are given in Tables 7.1 and 7.2. The frequency distributions of the residuals are given in Appendix 3. To put the magnitude of the correlation coefficients in perspective, the average correlation between *unstandardised instruments* is 0.75; that is, a correlation between a *residual and a single dimension* of 0.25 is 0.25/0.75 or one third of this magnitude which is quantitatively large.

Table 7.1 Dimension and instrument correlations with MAU residuals (Total 1269) and SF-36 dimensions and SWB instruments

Residual	SF-36 Dimension										SWB and other				
	Gen	Phys	RoleP	Pain	SumP	Vital	Social	RoleE	MH	SumM	PWI	SWLS	IHS	Self-TTO	SF-36
EQ5D-1.576*SF6D	-.054	.080**	-.180**	.023	.071*	-.188**	-.231**	-.280**	-.216**	-.316**	-.057*	-.083**	-.056*	-.011	-.184**
EQ5D-.843*HUI3	.094**	.105**	.086**	.190**	.180**	.069*	.013	-.006	-.048	-.055	-.098**	-.085**	-.099**	.014	.076**
EQ5D- 1.757*15D	-.052	.013	-.021	.179**	.091**	-.088**	-.054	-.108**	-.101**	-.141**	-.073**	-.094**	-.082**	.005	-.041
EQ5D- .847*AQOL4D	.053	.153**	.074**	.213**	.206**	-.001	-.063*	-.033	-.077**	-.123**	-.131**	-.135**	-.143**	-.003	.049
EQ5D- 1.038*AQoL8D	-.004	.209**	.078**	.208**	.297**	-.186**	-.141**	-.161**	-.299**	-.350**	-.258**	-.292**	-.284**	-.083**	-.043
HUI3- 1.869*SF6D	-.127**	-.016	-.233**	-.138**	-.087**	-.226**	-.218**	-.246**	-.154**	-.238**	.031	-.004	.032	-.022	-.228**
HUI3- 2.083*15D	-.157**	-.100**	-.116**	-.018	-.100**	-.168**	-.070*	-.108**	-.054	-.089**	.030	-.007	.021	-.010	-.125**
HUI3- 1.004*AQOL4D	-.034	.060*	-.004	.042	.043	-.066*	-.076**	-.029	-.034	-.075**	-.042	-.059*	-.054	-.016	-.021
HUI3- 1.231*AQoL8D	-.101**	.116**	-.005	.027	.132**	-.270**	-.163**	-.167**	-.271**	-.319**	-.175**	-.226**	-.203**	-.104**	-.124**
SF6D- 1.115*15D	.008	-.069*	.164**	.141**	.011	.111**	.186**	.186**	.128**	.193**	-.009	-.002	-.019	.016	.150**
SF6D- .537*AQOL4D	.104**	.073**	.243**	.185**	.132**	.178**	.160**	.234**	.132**	.182**	-.072*	-.050	-.084**	.009	.223**
SF6D- .659*AQOL8D	.054	.131**	.276**	.192**	.233**	.007	.102**	.133**	-.080**	-.025	-.209**	-.216**	-.237**	-.075**	.153**
15D- .482*AQOL4D	.104**	.148**	.098**	.058*	.132**	.081**	-.016	.066*	.013	.002	-.069*	-.053	-.074**	-.007	.089**
15D- .591*AQOL8D	.051	.229**	.114**	.048	.247**	-.123**	-.106**	-.071*	-.241**	-.257**	-.223**	-.240**	-.244**	-.103**	-.006
AQOL4D-1.226*AQOL8D	-.070*	.055	.000	-.019	.092**	-.215**	-.087**	-.146**	-.253**	-.256**	-.140**	-.175**	-.155**	-.093**	-.109**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

(1) Source: Figure 5.2 (constants omitted as they do not affect correlation)

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 7.2 Dimension and instrument correlations with MAU residuals (Total 1269) and AQoL-8D dimensions

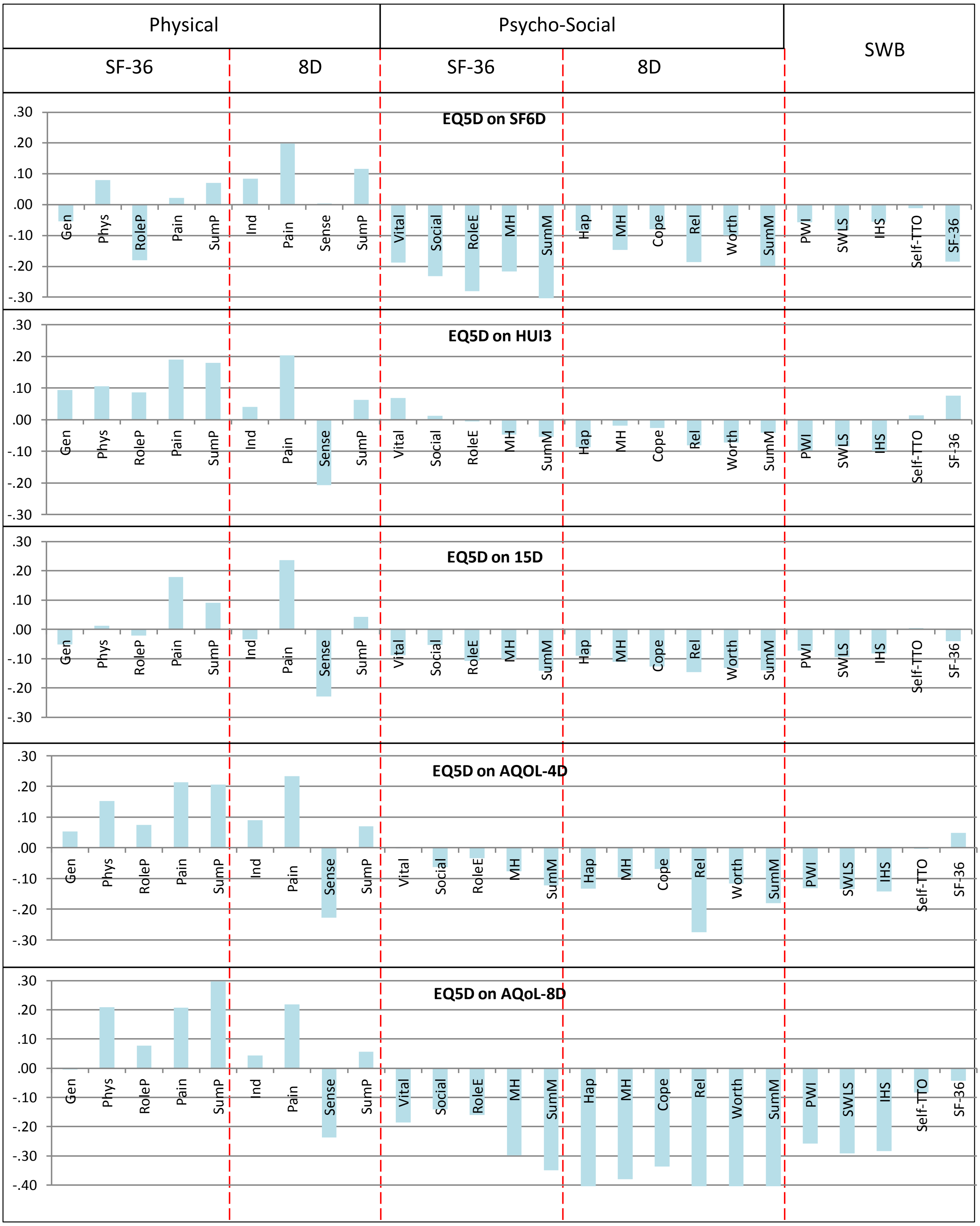
Residual	AQoL-8D Dimensions									
	Ind	**Pain**	**Sense**	**SumP**	**Hap**	**MH**	**Cope**	**Rel**	**Worth**	**SumM**
EQ5D-1.576*SF6D	.085**	.199**	.004	.116**	-.084**	-.146**	-.080**	-.186**	-.099**	-.199**
EQ5D-.843*HUI3	.041	.204**	-.207**	.063*	-.088**	-.019	-.027	-.082**	-.073**	-.041
EQ5D- 1.757*15D	-.034	.236**	-.229**	.042	-.090**	-.111**	-.126**	-.146**	-.131**	-.140**
EQ5D- .847*AQOL4D	.090**	.234**	-.228**	.070*	-.133**	-.096**	-.069*	-.275**	-.117**	-.180**
EQ5D- 1.038*AQoL8D	.044	.218**	-.238**	.056*	-.411**	-.380**	-.337**	-.406**	-.414**	-.422**
HUI3- 1.869*SF6D	.043	.009	.177**	.052	-.001	-.115**	-.049	-.098**	-.027	-.144**
HUI3- 2.083*15D	-.080**	.027	-.016	-.023	.002	-.097**	-.103**	-.064*	-.058*	-.103**
HUI3- 1.004*AQOL4D	.056*	.050	-.039	.013	-.054	-.081**	-.045	-.207**	-.051	-.147**
HUI3- 1.231*AQoL8D	.006	.023	-.041	-.004	-.350**	-.387**	-.333**	-.350**	-.368**	-.410**
SF6D- 1.115*15D	-.118**	.013	-.215**	-.080**	.003	.046	-.034	.056*	-.019	.074**
SF6D- .537*AQOL4D	.007	.037	-.223**	-.043	-.048	.047	.010	-.088**	-.018	.016
SF6D- .659*AQOL8D	-.047	.011	-.253**	-.068*	-.340**	-.240**	-.267**	-.224**	-.327**	-.227**
15D- .482*AQOL4D	.126**	.027	-.025	.034	-.056*	.003	.045	-.153**	-.001	-.058*
15D- .591*AQOL8D	.088**	-.003	-.029	.019	-.384**	-.324**	-.258**	-.317**	-.343**	-.342**
AQOL4D-1.226*AQOL8D	-.059*	-.033	.001	-.020	-.314**	-.323**	-.307**	-.137**	-.338**	-.271**
**. Correlation is significant at the 0.01 level (2-tailed).										
*. Correlation is significant at the 0.05 level (2-tailed).										

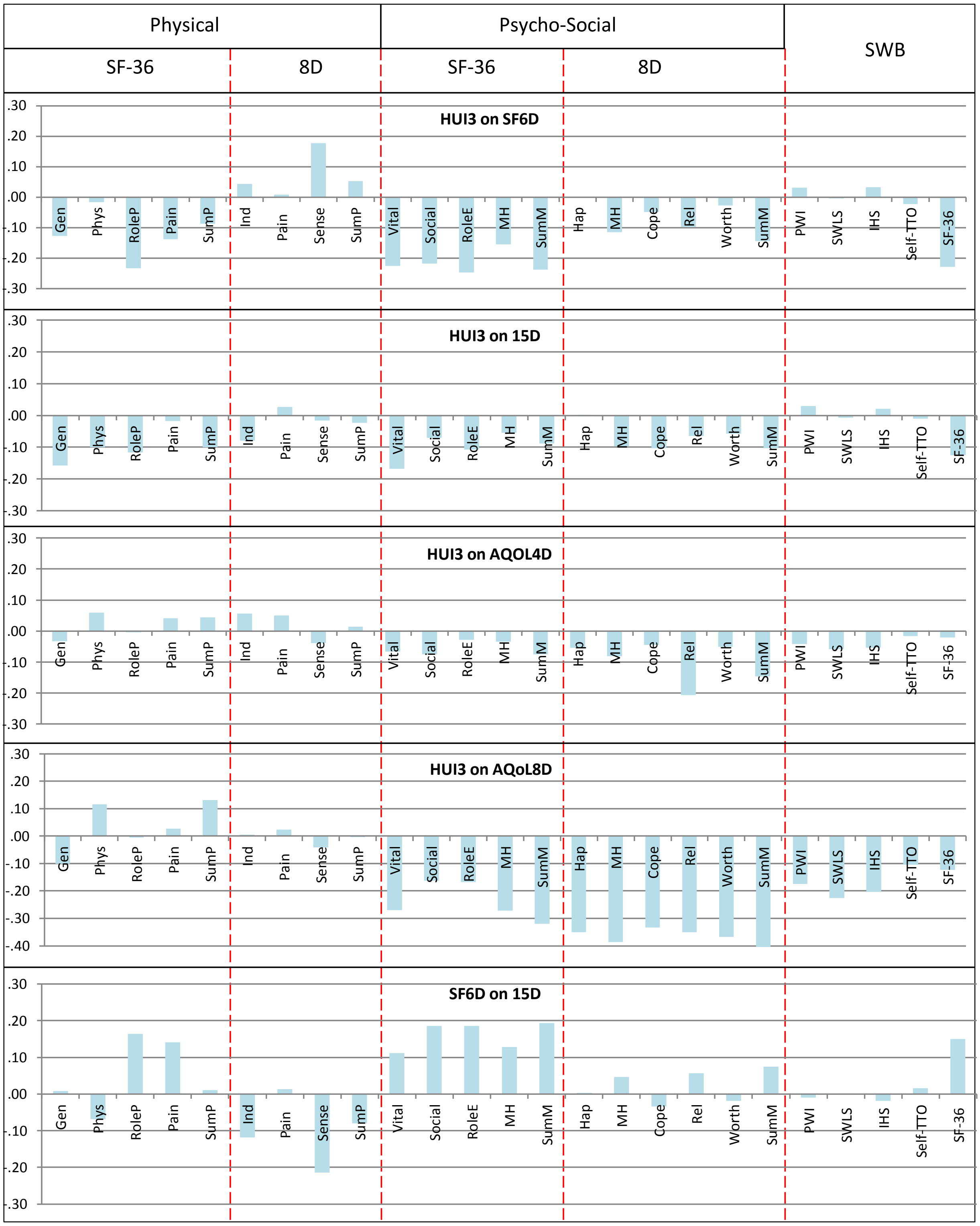
(1) Source: Figure 5.2 (constants omitted as they do not affect correlation)

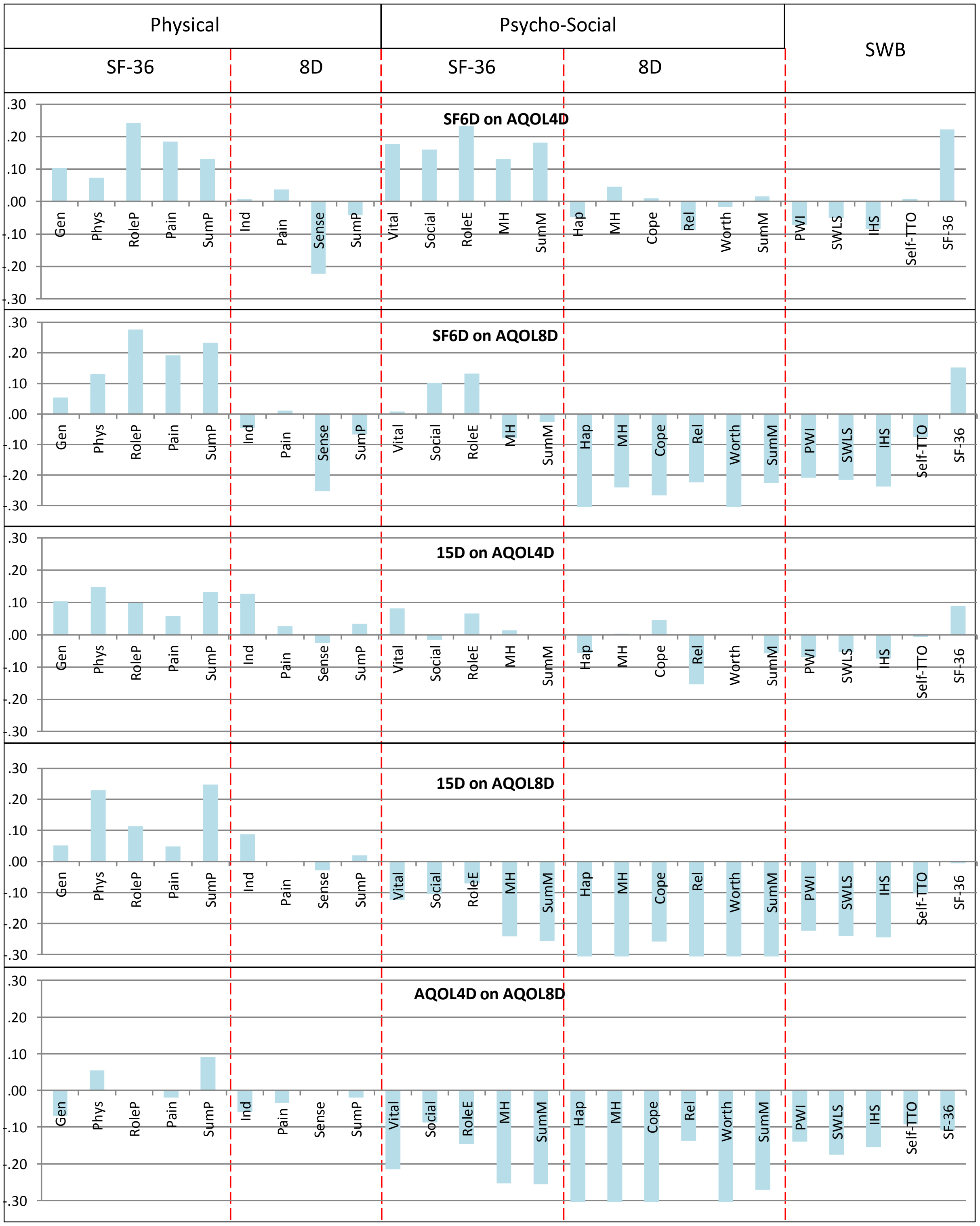
* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Figure 7.1 Pairwise instrument sensitivity: Correlations of dimension scores with the residual of one MAU instrument regressed upon a second




Physical
Psycho-Social
SWB
SF-36
8D
SF-36
8D
HUI3 on SF6D
HUI3 on 15D
HUI3 on AQOL4D
HUI3 on AQoL8D
SF6D on 15D
.30
.20
.10
.00
-.10
-.20
-.30
-.40
Gen
Phys
RoleP
Pain
SumP
Ind
Pain
Sense
SumP
Vital
Social
RoleE
MH
SumM
Hap
MH
Cope
Rel
Worth
SumM
PWI
SWLS
IHS
Self-TTO
SF-36



Key

Gen = general health; Phy = physical function; Role P = role limit physical; BP = bodily pain; Vit = vitality; Soc = social functioning; Role E = role limit emotional; MH = mental health; Cope = Coping; Rel = relationships; Worth = self- worth; Pain = pain; Sen = senses; MSD = mental super dimension; PSD = physical super dimension; SF-36: 8 dimensions – 4 physical; 4 psycho-social. AQoL-8D: 8 dimensions - 3 physical; 5 psycho-social; Self- TTO = Self TTO; PWI = Personal Wellbeing Index; SWLS = Satisfaction with Life Survey; IHS = Integrated Household Survey.

8 Discussion and Conclusion

MAU instruments were scored for this paper using the algorithms summarised in Box 4. Prima facie the use of weights derived in one country in a second country may appear to invalidate the results. However this is not necessarily true and the issue of utility weights is complex. First there is very significant within country variation in preferences as found in the UK between social and demographic groups (Kind, Hardman et al. 1999). At best, national weights are themselves an average from heterogeneous groups.

The difference between national averages is presently of unknown importance. More significantly the evidence suggests the variance in scores is relatively insensitive to differences in weights. Using pilot data for this project Richardson and Khan (2012) found that 85 percent of the difference between instruments could be explained by unweighted instrument values, leaving little to be explained by differences in weights.

As a further test of this, US and UK weights published by the EuroQol group for the new five level EQ-5D-5L have been applied to the present data and the results reported in Figure 8.1. The R^2 of 0.99 (US and UK) and 0.95 (Germany and UK) indicates that, overall, conclusions with respect to correlation and sensitivity could not change with the choice of weights. The significant difference in absolute score at the lower end of the scale suggests, prima facie, an error in the UK values. It appears very implausible that when US citizens assign a score of 0.29, UK citizens would prefer to be dead.

The two figures also indicate that the EQ-5D-5L does not overcome the problem of insensitivity in the region of good health (ceiling effects). The second highest possible UK and US utility scores are 0.906 and 0.888 respectively. This implies that moving 11 and 9 people respectively from the second highest health state to the highest would be equivalent to saving a life and returning a person to full health for the same period of time. Nevertheless some results might vary and the data available from this project could be reweighted with new scoring formula for difference countries.

The major conclusion to be drawn from this report is that, despite a similarity in the mean scores, the instruments are dissimilar with respect to virtually all other criteria used to compare them. Taking account of the fact that MAUI purport to measure the same quantity the correlation between instruments is low, the marginal relationships inconsistent and their relationship with health dimensions is variable. This suggests that, contrary to the impression generated by use of the generic term ‘utility’, the instruments are measuring different constructs. In effect each MAU instrument employs a different definition of ‘health’. The correlation which exists between instruments does not disconfirm this conclusion. Over a wide range of objects the height and weight of people correlate (the coefficient is about 0.81). But this does not demonstrate the existence of a common property (Chan 2003). A further important conclusion is that the evaluation of instruments is complex. Multiple criteria exist for their assessment many of which have not been discussed in this report.

Table 8.1 Summary of MAU order by criteria

Criteria	Instrument						
	EQ-5D	**HUI 3**	**SF-6D**	**15D**	**AQoL-4D**	**AQoL-8D**	**Ratio highest/ lowest**
Distribution							
Mean value	0.73	0.70	0.71	0.85	0.63	0.73	1.35
Ceiling (% 1.00)	18.4	6.2	1.8	6.9	6.0	2.2	8.27
Floor (%<0.4)	9.5	15.3	1.3	0.3	21.9	10.2	73.0
Correlation							
ICC (ave with other 7)	0.71	0.70	0.63	0.52	0.64	0.74	1.37
SWB (PWI)	0.46	0.52	0.50	0.50	0.55	0.62	1.35
SF-36	0.81	0.76	0.93	0.84	0.78	0.84	1.22
Self TTO	0.36	0.36	0.37	0.36	0.37	0.42	1.67
Discrepancies from b=1 in							
Pairwise regression (ave %)	35.0	47.4	58.8	74.6	47.0	34.0	2.19
Sensitivity							
b coefficient in mult reg on SF-36 dim (Table 6.4b)							
Pain	0.38	0.22	0.25	0.11	0.15	0.14	3.45
Gen Health	0.03	0.03	0.01	0.11	0.08	0.08	11.0
Physical function	0.33	0.33	0.11	0.35	0.24	0.15	3.18
Vitality	0.07	0.03	0.15	0.13	0.05	0.17	5.67
Mental health	0.18	0.34	0.18	0.22	0.26	0.42	2.33
Rank order sensitivity using residuals							
Physical sum (SF-36)	1	4	3	2	4	5	
Physical sum (AQoL-8D)	1	3	5	3	2	4	
Mental sum (SF-36)	5	4	1	3	3	2	
Mental sum (AQoL-8D)	6	5	3	4	2	1	
Self TTO	3	4	2	1	3	2	
SWB (PWI)	6	3	5	4	2	1	

Figure 8.1a Comparison of EQ-5D using US and UK weights

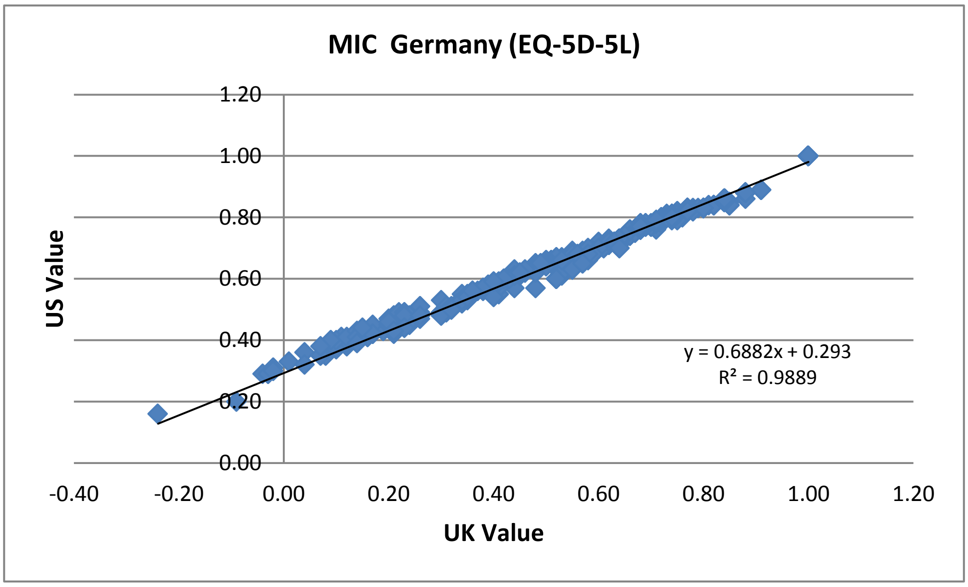
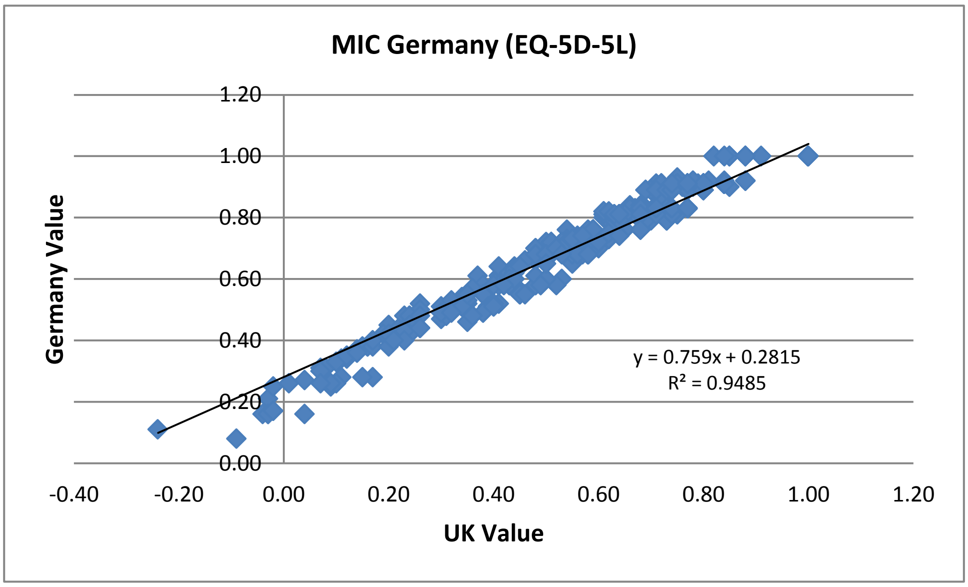


Figure 8.1b Comparison of EQ-5D using UK and German weights



Appendix 1 Frequency distribution of MAU instruments

Figure A.1.1 Frequency distribution of MAU instruments (Total n=1269)

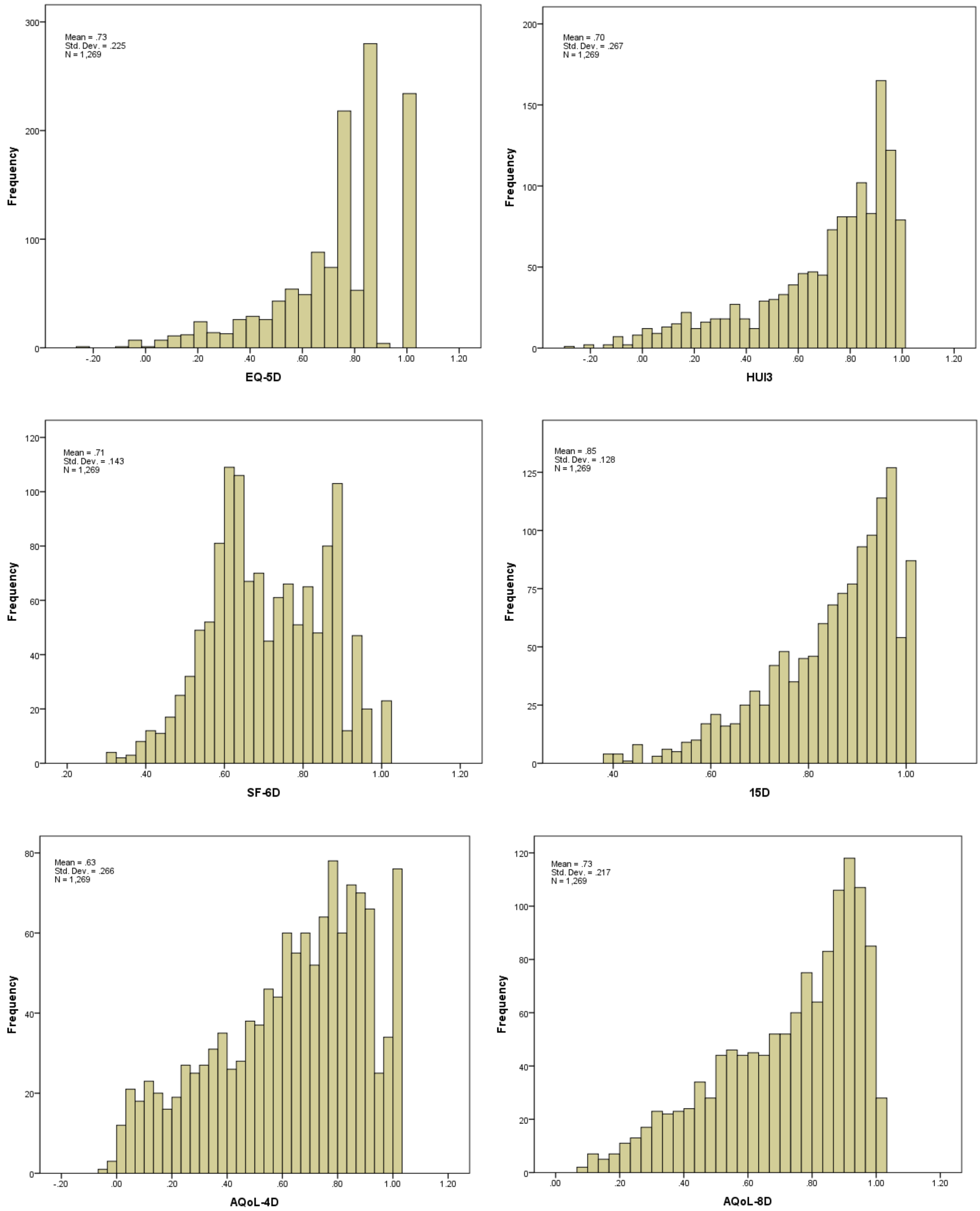
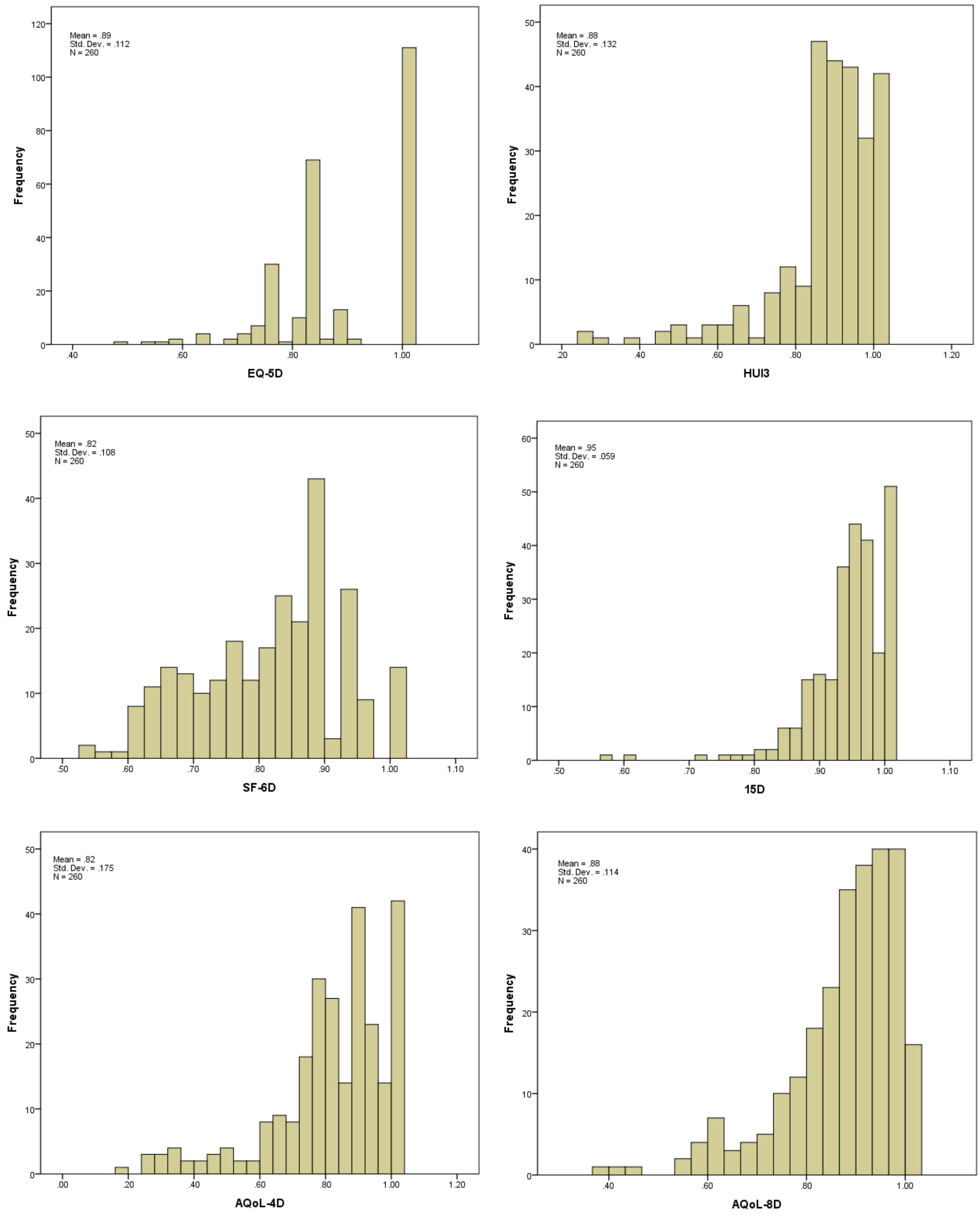


Figure A.1.2 Frequency distribution of MAU instruments (Public n=260)



Appendix 2 Frequency distribution of non-MAU instruments

Figure A.2.1 Frequency distribution of non-MAU instruments (Total n=1269)

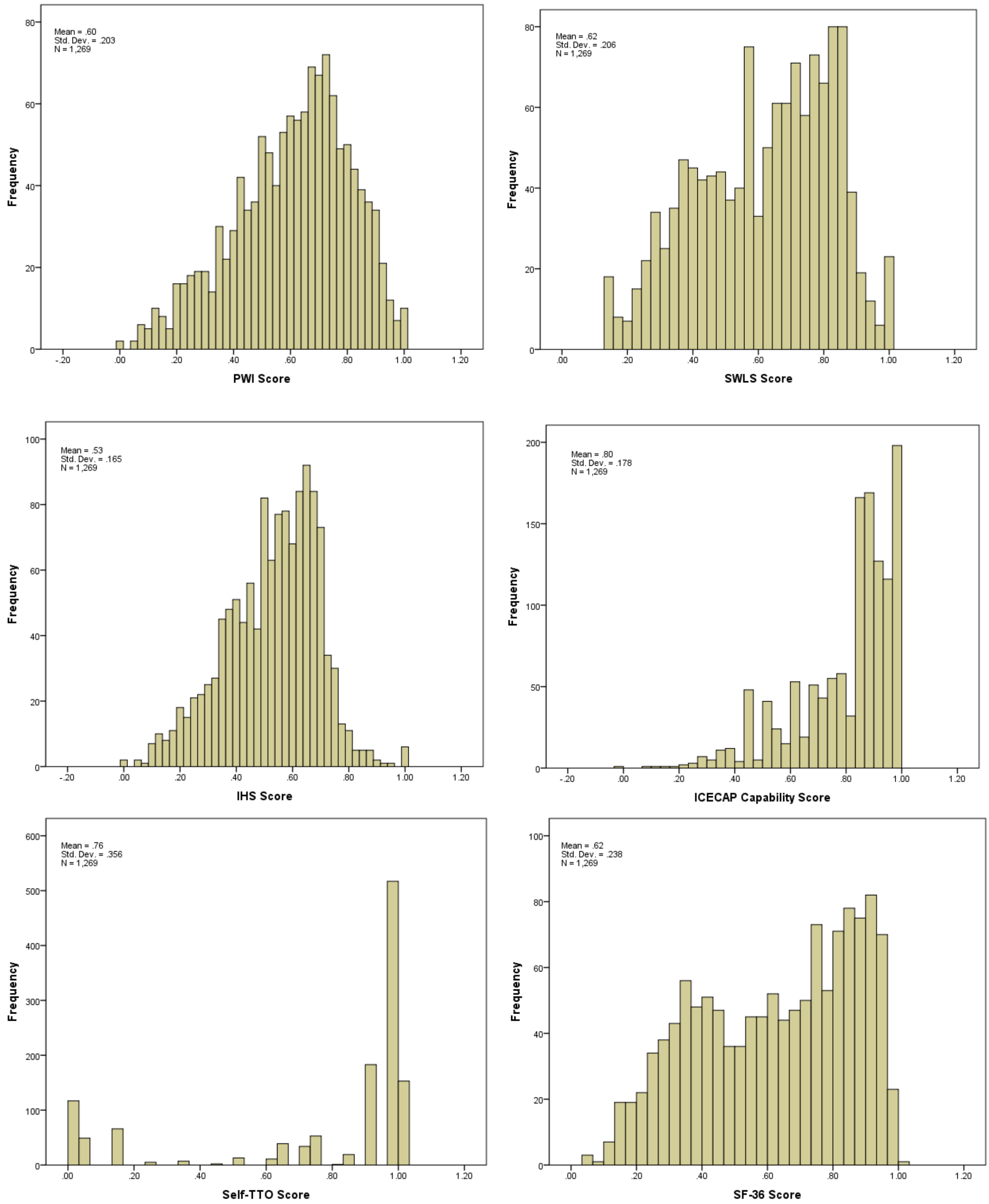
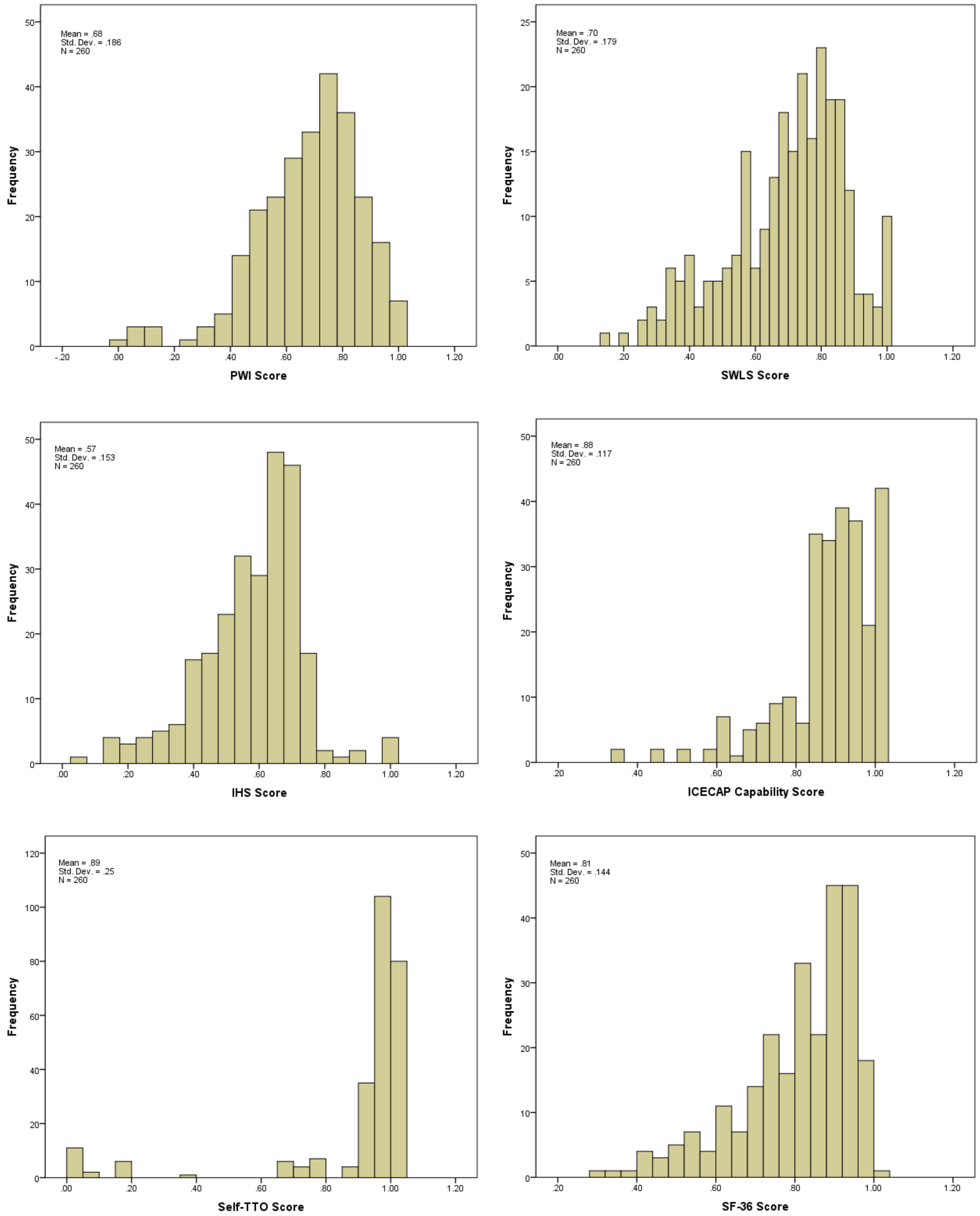
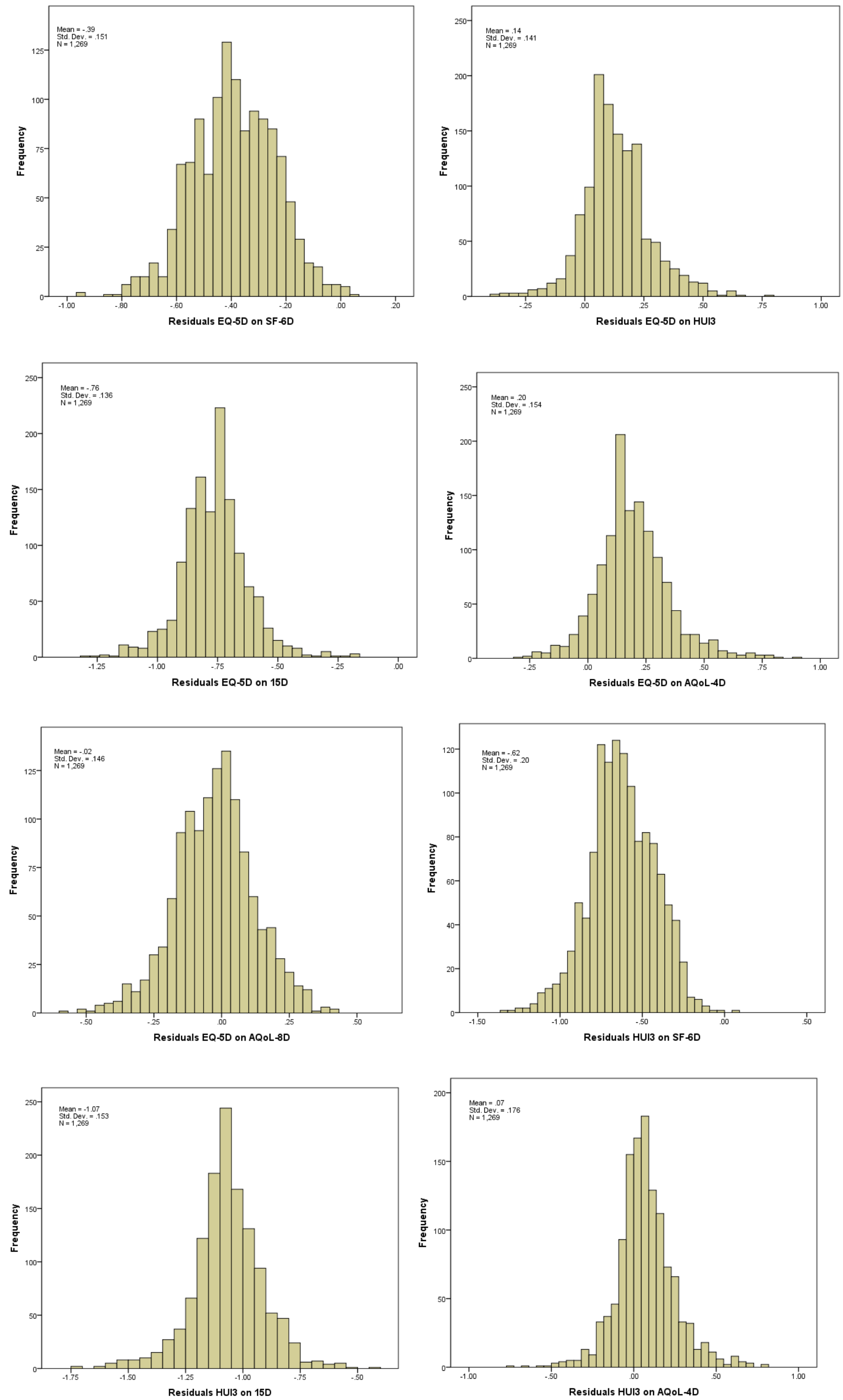


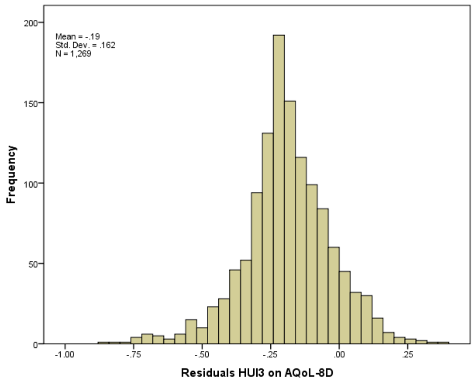
Figure A.2.2 Frequency distribution of non-MAU instruments (Public n=260)

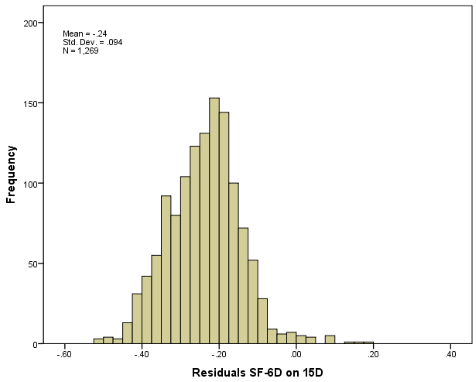


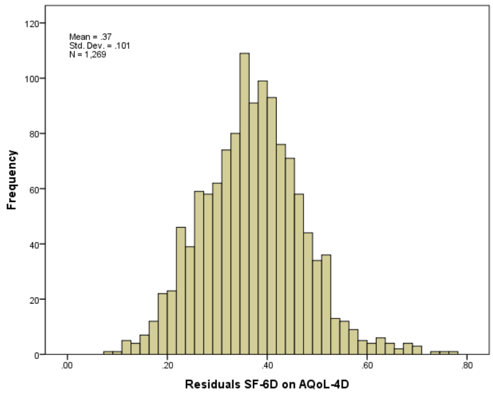
Appendix 3 Frequency distribution of residuals from pairwise regression of MAUI

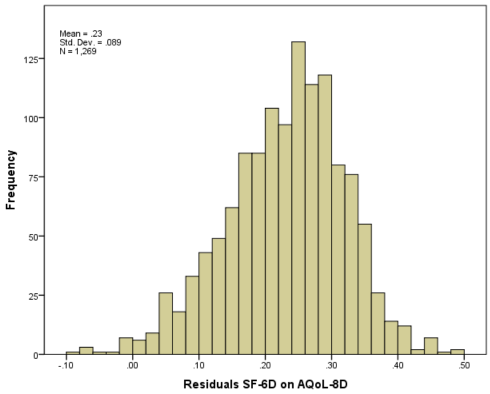
Figure A.3.1 Frequency distribution of residuals from pairwise regression of MAU instruments (Total n=1269)

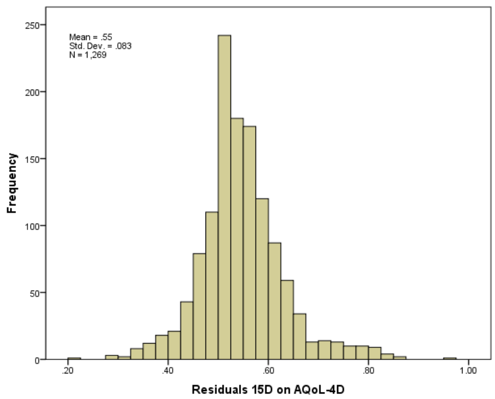


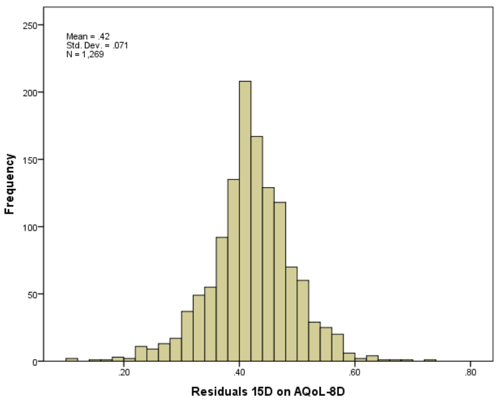

Mean = -.19
Std. Dev. = .162
N = 1,269
Frequency
Residuals HUI3 on AQoL-8D

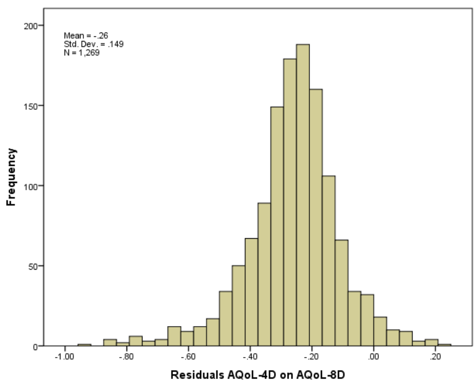

Mean = -.24
Std. Dev. = .094
N = 1,269
Frequency
Residuals SF-6D on 15D


Mean = .37
Std. Dev. = .101
N = 1,269
Frequency
Residuals SF-6D on AQoL-4D


Mean = .23
Std. Dev. = .089
N = 1,269
Frequency
Residuals SF-6D on AQoL-8D


Mean = .55
Std. Dev. = .083
N = 1,269
Frequency
Residuals 15D on AQoL-4D


Mean = .42
Std. Dev. = .071
N = 1,269
Frequency
Residuals 15D on AQoL-8D


Mean = -.26
Std. Dev. = .149
N = 1,269
Frequency
Residuals AQoL-4D on AQoL-8D

References

Chan, Y. H. (2003). "Biostatistics 104: Correlational Analysis." Singapore Medical Journal **44**(12): 614-619.

Cronbach, L. (1951). "Coefficient alpha and the internal structure of tests." Psychometrika **16**(3): 297-334.

Hawthorne, G., S. Korn, et al. (2012). "Population norms for the AQoL derived from the 2007 Australian National Survey of Mental Health and Wellbeing." Australian and New Zealand Journal of Public Health (submitted).

Kind, P., G. Hardman, et al. (1999). US Population Norms for EQ-5D, Discussion Paper 172. York, Centre for Health Economics, University of York.

Pallant, J. (2010). SPSS Survival Manual: A step by step guide to data analysis using SPSS. Crows Nest, Allen and Unwin.

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

Richardson, J., A. Iezzi, et al. (forthcoming). The Self TTO: Its properties, validity, reliability and relationship with MAU scores, Research Paper. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., A. Iezzi, et al. (2012). Cross-national comparison of twelve quality of life instruments: MIC Paper 1 Background, questions, instruments, Research Paper 76. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., M. A. Khan, et al. (2012). Population norms and Australian profile using the Assessment of Quality of Life (AQoL) 8D utility instrument, Research Paper 72. Melbourne, Centre for Health Economics, Monash University.

Richardson, J., A. Maxwell, et al. (2012). The Myth of Maximisation: Quantifying trade-offs between social objectives using the Conflict Scale, Research Paper 74. Melbourne, Centre for Health Economics, Monash University.