

Broadening the Measurement and Valuation of Health and Quality of Life

Brendan Mulhern

Doctor of Philosophy

Centre for Health Economics Research and Evaluation

Faculty of Business

University of Technology Sydney

First submitted in November 2019

Resubmitted in July 2020

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, BRENDAN MULHERN declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Business at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 03/07/2020

ACKNOWLEDGEMENTS

I would like to thank many people who have helped me complete this thesis over the last three or so years. Thanks to my supervisors, Prof's Rosalie Viney and Deborah Street for all the support, and for the insightful comments that have immeasurably improved this work. In particular I thank Rosalie for giving me the chance to move to CHERE in 2015 and encouraging me to (eventually) start my PhD. Particular thanks to Debbie for making me think outside of the box when it came to DCE methods, challenging my analysis skills, and for the long discussions in both of our offices about my thesis and beyond. Madeleine King assessed each stage of my thesis, and her comments were incredibly useful.

I would also like to acknowledge my career mentors Prof. John Brazier and Prof. Aki Tsuchiya for the ongoing research collaborations and support. Also, thanks to my collaborators on the projects I have included in my thesis, and also many other health economists around the world that have directly or indirectly helped me over the line. In particular, thanks go to Richard Norman and Richard De Abreu Lourenco for the advice about academic and PhD life in Australia, and for contributing to the papers published in this thesis. Juliette Malley and Mark Oppe also helped to improve the studies included. The EuroQol Group have also been incredibly supportive of the sort of empirical work reported here, and thanks to the group for inspirational discussions at the various annual meetings. Work in this PhD was funded by the EuroQol Group, and I thank them for financial support as well Nick Bansback and Koonal Shah have also encouraged me through to the end, and I value their support.

I would also like to thank everyone at CHERE for their support, and giving me space with my office door closed (and taking work off my hands) when it came towards the end! In particular I thank Phil Haywood, Tom Longden and Kathleen Manipis for the support, and distractions when required.

Since moving to Sydney, and during my PhD, I have been lucky to meet my now fiancée Dr. Emily Trimmer. I thank her for her support in all aspects of our lives.

I also thank my parents in England and my future in laws for their support. Also, to friends outside academia with no interest in health economics that allowed for an escape from this thesis when needed.

THESIS FORMAT

This is a conventional thesis including seven chapters of content, references and further appendices. Two of the chapters (a literature review and empirical study) three empirical studies have resulted in published journal articles as described below.

PUBLISHED PAPERS RESULTING FROM THIS THESIS

1. **Mulhern B**, Norman R, DeAbreu Lourenco R, Street D, Malley J, Viney R. Investigating the relative value of health and social care related quality of life using a Discrete Choice Experiment. *Social Science and Medicine*. 2019; 233: 28-37.

Author contribution for this paper: Brendan Mulhern led the conceptualisation of the study, developed the study design, constructed the experiment, and led the data analysis and interpretation and paper writing. Richard Norman, Richard De Abreu Lourenco, Deborah Street, Juliette Malley and Rosalie Viney supported the conceptualisation of the study and the development of the study design. They also contributed to the interpretation of the results and supported the development of the manuscript.

2. **Mulhern B**, Norman R, Street D, Viney R. One method, many methodological choices: A structured review of Discrete Choice Experiments for health state valuation. *Pharmacoeconomics*. 2019; 37(1):29-43.

Author Contributions for this paper: Brendan Mulhern conducted the literature search, extracted the data, led the data synthesis and interpretation and developed the first draft of the manuscript. Richard Norman, Rosalie Viney and Deborah Street supported the data extraction process and interpretation of the results and were involved in the development of the manuscript.

TABLE OF CONTENTS

Chapter	Page
1. Measuring and valuing health – What are the issues?	1
1.1. Research question	4
1.2. Aims of thesis	5
1.3. Plan of thesis	5
2. Background to the measurement and valuation of health and quality of life	7
2.1. Chapter summary	7
2.2. The economic evaluation of health interventions	7
2.3. Key theoretical approaches and concepts important for economic evaluation	8
2.3.1. The utility principle	8
2.3.2. The Welfarist approach to economic evaluation	9
2.3.3. The Extra-Welfarist approach to economic evaluation	10
2.4. Cost utility analysis and the Incremental Cost Effectiveness Ratio	11
2.5. The Quality Adjusted Life Year	12
2.6. Health-related quality of life	14
2.7. Measuring and valuing health-related quality of life	15
2.8. Preference elicitation methods	15
2.8.1. General principles	15
2.8.2. Visual Analogue Scale	16
2.8.3. Standard Gamble	17
2.8.4. Time Trade-Off	18
2.8.5. Discrete Choice Experiments	20
2.9. Description of the DCE methodology	22
2.9.1. Developing descriptive systems for valuation	23
2.9.2. Constructing the choice sets – General format	24
2.9.3. Constructing the choice sets – Anchoring	25
2.9.4. Constructing the experimental design	27
2.9.5. Implementing the DCE – Task allocation	28
2.9.6. Implementing the DCE – Sample size, choice set and observation numbers	29
2.9.7. Implementing the survey – Mode of administration	29
2.9.8. Implementing the DCE - Survey format	30
2.9.9. Data analysis and modelling	31
2.9.10. Conditional logit model	31
2.9.11. Scale assessment modelling	32
2.9.12. Latent Class model to assess heterogeneity	32
2.9.13. Mixed logit model to assess heterogeneity	33

2.9.14.	Generalised Multinomial Logit Model to assess heterogeneity	33
2.10.	Instruments used to measure health and quality of life – Profile measures	34
2.11.	Preference-based measures of health	35
2.11.1.	General structure and principles.....	35
2.11.2.	Who should value quality of life?.....	36
2.11.3.	Example of a generic PBM - EQ-5D	37
2.11.4.	Example of a generic PBM - Short Form-6 Dimension (SF-6D)	39
2.11.5.	Example of a generic PBM - Health Utilities Index (HUI 2 and HUI-3)	40
2.11.6.	Example of a generic PBM - Assessment of Quality Of Life (AQoL)	41
2.11.7.	Multiplicative and additive value set modelling	42
2.11.8.	Limitations of HRQoL focused descriptive systems	42
2.11.9.	Limitations of HRQoL focused value sets	43
2.12.	Issues with the health-related QALY framework.....	44
2.13.	Moving beyond the health-focused QALY.....	44
2.13.1.	Why is moving beyond the health-focused QALY important?.....	44
2.13.2.	Adult Social Care Outcomes Toolkit	45
2.13.3.	ICECAP-A.....	47
2.13.4.	Limitations of broader PBMs	47
2.14.	What about combining or broadening existing outcome measures?.....	47
2.15.	The empirical work reported in the thesis	48
3.	Using Discrete Choice Experiments to value health states: A structured literature review	50
3.1.	Summary	50
3.1.1.	Aims and objectives:	50
3.2.	Structured review methods.....	51
3.2.1.	Literature search	51
3.2.2.	Inclusion and exclusion criteria.....	51
3.2.3.	Assessing paper content and quality	52
3.2.4.	Data extraction process.....	52
3.3.	Results.....	53
3.3.1.	Studies identified	53
3.3.2.	Findings – General study information.....	54
3.3.3.	Findings – Paper content and quality.....	60
3.3.4.	Findings – Choice set and study design.....	60
3.3.5.	Findings – Type of designed experiment.....	61
3.3.6.	Findings – Data analysis and modelling.....	63
3.4.	Discussion.....	64

3.4.1.	Summary.....	64
3.4.2.	What are the recurring limitations?.....	65
3.4.3.	Where is there consensus?.....	66
3.4.4.	What are the remaining questions and what further work is required?	66
3.4.5.	Review limitations.....	67
3.4.6.	Conclusions.....	68
4.	Assessing the relationship between QoL outcome measures using Item Response Theory methods	69
4.1.	Summary	69
4.2.	Introduction.....	69
4.3.	Defining a ‘layered’ approach to measurement.....	70
4.4.	Why is this research important?	71
4.5.	Aims and objectives	72
4.6.	Item Response Theory	72
4.6.1.	What is the IRT model, and what does it estimate?	72
4.6.2.	What IRT models are available?.....	76
4.6.3.	Assumptions of the IRT model.....	77
4.6.4.	How is IRT useful in developing and assessing QoL outcome measures?.....	78
4.6.5.	Overview of literature using IRT methods in PBM and HRQoL measure development	78
4.7.	Description of empirical study.....	80
4.7.1.	Data and study design	80
4.7.2.	Respondents and recruitment	82
4.8.	Descriptive assessment of the sample and outcome measures	83
4.8.1.	Sample demographics and survey completion process	83
4.8.2.	Scoring the measures	83
4.8.3.	Scoring the PBMs of HRQoL.....	84
4.8.4.	Scoring the PBMs of wider QoL	84
4.8.5.	Scoring the profile based HRQoL instruments – SF-36	84
4.8.6.	Scoring the profile based HRQoL instruments – PROMIS-29	84
4.8.7.	Scoring the WEMWBS	85
4.8.8.	Descriptive analysis of the measures and sample	85
4.9.	Results – Sample and measure descriptive statistics	85
4.9.1.	Sample characteristics and survey completion process.....	85
4.9.2.	Descriptive analysis of measures	87
4.10.	Extension 1 – Data analysis	89
4.10.1.	Estimation procedures	89

4.10.2.	Model specifications	90
4.10.3.	Rotation method and criteria	92
4.10.4.	Generating a dimension structure for further testing	92
4.11.	Results - Dimensionality assessment	92
4.12.	Broadening the standard framework.....	95
4.13.	Establishing a dimensionality for Extension 2 analyses	98
4.14.	Extension 2 – Investigating a layered approach to measurement	98
4.15.	Description of general IRT approach.....	98
4.15.1.	Which IRT model was used?	98
4.15.2.	How is IRT used in this study?.....	99
4.16.	Data analysis process	99
4.16.1.	Preliminary descriptive analysis – Data inspection	99
4.16.2.	Test common IRT assumptions – Local independence.....	100
4.16.3.	Test functional form and model-data fit	100
4.16.4.	Testing for DIF	101
4.16.5.	Item level thresholds.....	101
4.16.6.	Item level information and the total information provided	102
4.16.7.	Estimated calibrated dimension score curve	102
4.17.	Results overview	103
4.18.	Results – Physical Functioning dimension.....	103
4.18.1.	Justification of dimensionality	103
4.18.2.	Initial data inspection	103
4.18.3.	Assessing local independence.....	103
4.18.4.	Assessing model-data fit – Item level.....	104
4.18.5.	Assessing model-data fit – Model level.....	105
4.18.6.	Assessing DIF.....	105
4.18.7.	IRT item calibrations - Assessing item level thresholds and ordering	106
4.18.8.	IRT item calibrations - Assessing item information.....	108
4.18.9.	IRT Item calibrations - Assessing IRT score estimates	108
4.18.10.	Summary and implications of results	109
4.19.	Results – Mental health dimension	110
4.19.1.	Justification of dimensionality	110
4.19.2.	Initial data inspection	110
4.19.3.	Assessing local dependence.....	111
4.19.4.	Assessing model-data fit – Item level.....	112
4.19.5.	Assessing model-data fit – Model level.....	112
4.19.6.	Assessing DIF.....	113

4.19.7.	IRT item calibrations - Assessing item level thresholds and ordering	113
4.19.8.	IRT item calibrations - Assessing item information	113
4.19.9.	IRT item calibrations - Assessing IRT score estimates	116
4.19.10.	Summary and implications of results	117
4.20.	Results – Pain dimension	118
4.20.1.	Justification of dimensionality	118
4.20.2.	Initial data inspection	118
4.20.3.	Assessing local dependence	118
4.20.4.	Assessing model-data fit – Item level	119
4.20.5.	Assessing model-data fit – Model level	119
4.20.6.	Assessing DIF	121
4.20.7.	IRT item calibrations - Assessing item level thresholds and ordering	121
4.20.8.	IRT item calibrations - Assessing item information	122
4.20.9.	IRT item calibrations - Assessing IRT score estimates	123
4.20.10.	Summary and implications of results	124
4.21.	Results – Activities dimension	124
4.21.1.	Justification of dimensionality	124
4.21.2.	Initial data inspection	125
4.21.3.	Assessing local dependence	125
4.21.4.	Assessing model-data fit – Item level	125
4.21.5.	Assessing model-data fit – Model level	126
4.21.6.	Assessing DIF	126
4.21.7.	IRT item calibrations - Assessing item level thresholds and ordering	126
4.21.8.	IRT item calibrations - Assessing item information	127
4.21.9.	IRT item calibrations - Assessing IRT score estimates	128
4.21.10.	Summary and implications of results	129
4.22.	Overall discussion and implications for extending the QoL measurement framework	129
4.22.1.	Summary	129
4.22.2.	Implications for the measurement of QoL	130
4.22.3.	Implications for developing a flexible approach to measuring outcomes	131
4.22.4.	Study limitations and suggestions for future research	132
4.23.	How this study informs this thesis	134
5.	Testing the performance of existing preference elicitation methods to develop a value set for a measurement system combining health and social care related quality of life	135
5.1.	Summary	135
5.2.	Introduction	135

5.3.	Aims and objectives	138
5.4.	Methods	138
5.4.1.	Development of the DCE valuation task.....	138
5.4.2.	Pilot study to test survey functioning	140
5.4.3.	Study design – Constructing the design	140
5.4.4.	Study design – Use of zero priors.....	141
5.4.5.	Study design – Issues around implausibility	142
5.4.6.	Study design – Blocking and dimension ordering	142
5.4.7.	Survey design and administration.....	143
5.4.8.	Recruitment and respondents	143
5.4.9.	Data analysis and modelling – Sample	144
5.4.10.	Data analysis and modelling – Conditional logit	144
5.4.11.	Data analysis and modelling – Testing interactions	145
5.4.12.	Data analysis and modelling – Investigating scale differences between subsamples	146
5.4.13.	Data analysis and modelling - Time taken	146
5.4.14.	Data analysis and modelling – Preference Heterogeneity	147
5.4.15.	Data analysis and modelling – Latent class.....	148
5.4.16.	Data analysis and modelling – Mixed logit	148
5.4.17.	Data analysis and modelling – Generalised Multinomial Logit Model	152
5.4.18.	Assessing model performance	154
5.5.	Results	154
5.5.1.	Pilot launch	154
5.5.2.	Sample – Completion process and time taken	155
5.5.3.	Sample – Demographics	156
5.5.4.	Conditional Logit models.....	158
5.5.5.	Models including interactions	161
5.5.6.	Scale testing across subsamples	163
5.5.7.	Sensitivity analysis - Time taken by task and overall	165
5.5.8.	Assessing heterogeneity – Latent class	165
5.5.9.	Assessing heterogeneity – Mixed logit.....	167
5.5.10.	Assessing heterogeneity – Generalised Multinomial Logit Model.....	170
5.6.	Discussion.....	170
5.6.1.	Summary and explanation of findings.....	170
5.6.2.	Comparison with other EQ-5D-5L and ASCOT value sets.....	174
5.6.3.	Study limitations and further research	175
5.6.4.	Conclusions.....	176

6.	Comparing DCE designs that could be used to value measures of QoL	178
6.1.	Summary	178
6.2.	Introduction.....	178
6.3.	Relevant past work comparing design constructions	181
6.4.	Aims and objectives	181
6.5.	Summary of methodological process undertaken.....	181
6.6.	Methods – Summary of design construction features	182
6.7.	Methods - Design construction method	183
6.7.1.	Level of overlap.....	183
6.7.2.	Use of priors	183
6.7.3.	Implementation platform.....	183
6.7.4.	Design property assessed – Level balance	183
6.7.5.	Design property assessed - Dominated choice sets.....	184
6.8.	<i>Methods - Design construction methods</i>	<i>184</i>
6.8.1.	Generator developed	184
6.8.2.	Modified Fedorov algorithm.....	184
6.8.3.	Coordinate exchange algorithm.....	185
6.8.4.	Random selection	185
6.9.	Design combinations excluded.....	186
6.10.	Summary of the simulation process	186
6.10.1.	Standardised bias	187
6.10.2.	Root Mean Squared Error	187
6.10.3.	Coverage.....	188
6.11.	Methods - Study design.....	188
6.12.	Methods - Sample size and respondents	188
6.13.	Methods - Data analysis.....	189
6.13.1.	Comparing sample characteristics	189
6.13.2.	Assessing respondent behaviour	189
6.13.3.	Comparing designs - Assessing feedback questions	190
6.13.4.	Comparing designs – Conditional logit analysis.....	190
6.13.5.	Comparing designs - Assessing poolability	191
6.13.6.	Comparing designs – Assessing preference heterogeneity using latent class 191	
6.13.7.	Comparing designs – Assessing preference heterogeneity using mixed logit 191	
6.14.	Results	192
6.14.1.	Completion and sample characteristics	192

6.14.2.	Comparing designs – Respondent behaviour	193
6.14.3.	Comparing designs – Feedback questions.....	197
6.14.4.	Comparing designs – Conditional logit models.....	198
6.14.5.	Comparing designs - Poolability.....	206
6.14.6.	Comparing designs - Assessing preference heterogeneity using latent class.....	207
6.14.7.	Comparing designs - Assessing preference heterogeneity using mixed logit.....	211
6.15.	Summary and discussion.....	214
6.15.1.	Overlap of severity levels	215
6.15.2.	Use of prior information.....	216
6.15.3.	Theoretical or algorithmic approaches	216
6.15.4.	Impact of different software packages.....	217
6.15.5.	Are certain designs better for certain models?	217
6.15.6.	Study limitations and future research.....	217
6.15.7.	Conclusions and recommendations for study designs.....	218
7.	Discussion	220
7.1.	Summary	220
7.2.	Broadening the measurement of health and QoL.....	221
7.3.	Developing a layered approach to measurement	225
7.4.	Using IRT methods in the assessment of PBMs.....	226
7.5.	Findings related to the valuation of QoL.....	228
7.6.	Using DCE to value broader QoL measures.....	228
7.7.	Testing design construction approaches.....	229
7.8.	Broadening the measurement and valuation of QoL using existing measures.....	230
7.9.	Implications of the findings.....	231
7.9.1.	What do the results mean for the concept of QoL, and decision making based on QoL?	231
7.9.2.	Can decision makers use existing PBMs with confidence?	232
7.9.3.	Should a broader approach to measuring QoL be advocated?	233
7.9.4.	If a broader measure of QoL was developed, what format should it take?.....	233
7.9.5.	Are DCE's an acceptable method for the valuation of health and QoL?.....	233
7.9.6.	What questions should decision makers and researchers ask when assessing the results from a PBM?	234
7.10.	Limitations of thesis and suggestions for further research.....	234
7.11.	Conclusions.....	236
8.	Appendices	238
8.1.	Appendix 1: HUI classification systems.....	238
8.2.	Appendix 2: AQoL-8D classification system	241

8.3.	Appendix 3: Structured review search terms.....	242
8.4.	Appendix 4: CREATE checklist for reporting valuation studies	243
8.5.	Appendix 5: Paper identification process for structured review.....	244
8.6.	Appendix 6: PRISMA Checklist.....	245
8.7.	Appendix 7: Health Measurement Study - Survey Outline	247
8.8.	Appendix 8: Orthogonal EFA models	267
8.9.	Appendix 9: Item coding and further item description for the measurement chapter 274	
8.10.	Appendix 10: Local independence values (all item pairs)	276
8.11.	Appendix 11: One Block of choice sets from the DCE design.....	279
8.12.	Appendix 12: EQ-5D-5L and ASCOT valuation - Survey content.....	280
8.13.	Appendix 13: HRQoL and SCRQoL interaction models	293
8.14.	Appendix 14: Scale testing for demographic variables	295
8.15.	Appendix 15: Time taken sensitivity analysis	298
8.16.	Appendix 16: Latent class models with between three and six classes.....	303
8.17.	Appendix 17: Further exploratory mixed logit models	307
8.18.	Appendix 18: Latent class demographic parameters (design comparison study).....	311
9.	References	313

LIST OF FIGURES

Figure 1: Cost effectiveness plane	12
Figure 2: Stylised QALY profile	14
Figure 3: Example VAS scale	17
Figure 4: The Standard Gamble valuation process	18
Figure 5: The Time Trade-Off process.....	18
Figure 6: Representation of the Lead Time TTO process	20
Figure 7: DCE choice set example 1	21
Figure 8: DCE choice set example 2	21
Figure 9: Conjoint analysis development process	23
Figure 10: DCE study design process	23
Figure 11: Example of Item Characteristic Curves	75
Figure 12: Example of Item Information Curve	75
Figure 13: Threshold and information curves - Physical functioning dimension	107
Figure 14: Total information curve - Physical functioning dimension	109
Figure 15: Expected score curve - Physical functioning dimension.....	109
Figure 16: Threshold and information curves - Mental health dimension	115
Figure 17: Total information curve - Mental health dimension.....	116
Figure 18: Estimated score curve - Mental health dimension	117
Figure 19: Threshold and information curves - Pain dimension	122
Figure 20: Total information curve - Pain dimension	123
Figure 21: Estimated score curve - Pain dimension	124
Figure 22: Threshold and information curves - Activities dimension	127
Figure 23: Total information curve - Activities dimension.....	128
Figure 24: Estimated score curve - Activities dimension	129
Figure 25: Example DCE choice set	140
Figure 26: Time taken per task	155
Figure 27: Frequency charts of respondent reported usability questions	156
Figure 28: Percentage of respondents answering at each level of each dimension .	158
Figure 29: Parameter estimates scaled on the value of health state 55555 (overlap designs)	202
Figure 30: Parameter estimates scaled on the value of health state 55555 (non-overlap designs)	203
Figure 31: Overall magnitude of the scaled level 5 parameter for each dimension (overlap).....	204
Figure 32: Overall magnitude of the scaled level 5 parameter for each dimension (non-overlap)	205
Figure 33: Latent class models with the lowest BIC (overlap designs)	209
Figure 34: Latent class models with the lowest BIC (non-overlap designs)	210
Figure 35: Mixed logit models (overlap designs)	212
Figure 36: Mixed logit models (non-overlap designs H to M).....	213
Figure 37: Mixed logit models (non-overlap designs N to S)	214
Figure 38: Structured review paper identification process.....	244
Figure 39: Sensitivity analysis of time taken per task.....	298
Figure 40: Sensitivity analysis of time taken to complete survey	299

LIST OF TABLES

Table 1: The EQ-5D-3L descriptive system	38
Table 2: The EQ-5D-5L descriptive system	38
Table 3: The SF-6D classification system	40
Table 4: ASCOT Descriptive System	46
Table 5: ICECAP descriptive system	47
Table 6: Study Categorisation	55
Table 7: Study design characteristics	62
Table 8: Choice set selection methods	63
Table 9: Modelling and analysis characteristics	64
Table 10: Summary of IRT models	77
Table 11: Measures included in the Health Measurement Study	82
Table 12: Survey completion process	86
Table 13: Sample demographics	87
Table 14: Descriptive statistics for each of the value sets estimated	88
Table 15: SF-36 and PROMIS dimension, WEMWBS and ONS-4 scores	89
Table 16: EFA models tested	91
Table 17: Dimension structure - EFA Oblique quartimax rotation (Model 1)	96
Table 18: Dimension structure - EFA Oblique Varimax rotation (Model 2)	97
Table 19: Initial data inspection - Physical functioning dimension	104
Table 20: Item pair descriptors with a dependency > 10 – Physical functioning dimension	104
Table 21: Item calibrations - Physical functioning dimension	105
Table 22: DIF assessment by gender and condition – Physical functioning dimension	106
Table 23: Total test information at key points of the latent scale – Physical Functioning dimension	108
Table 24: Initial data inspection - Mental health dimension	111
Table 25: Item pairs with local dependence estimates > 10 - Mental health dimension	111
Table 26: Item calibrations – Mental health dimension	112
Table 27: DIF by gender and condition - Mental health dimension	113
Table 28: Total test information at key points of the latent scale - Mental Health dimension	116
Table 29: Initial data description - Pain dimension	118
Table 30: Item pairs with local dependence estimates > 10 - Pain dimension	119
Table 31: Item calibrations - Pain dimension	120
Table 32: DIF by gender and condition - Pain dimension	121
Table 33: Total test information at key points of the latent scale - Pain dimension	123
Table 34: Initial data inspection – Activities dimension	125
Table 35: Item pairs with local dependence estimates > 10 - Activities dimension	125
Table 36: Item calibrations - Activities dimension	126
Table 37: DIF by gender and condition - Activities dimension	126
Table 38: Total test information at key points of the latent scale - Activities dimension	128

Table 39: Mixed logit parameter specifications	150
Table 40: GMNL model specifications	153
Table 41: Sample demographics	157
Table 42: Conditional logit models for the overall sample	160
Table 43: Exploratory analysis of interactions	162
Table 44: Conditional logit and heteroskedastic pooled models by measure order	164
Table 45: Latent class model performance statistics	165
Table 46: Two-class latent class model	166
Table 47: Mixed Logit models – EQ-5D-5L and ASCOT combinations as random	169
Table 48: Generalised multinomial logit models for the EQ-5D-5L and ASCOT data	172
Table 49: The 19 designs included in data collection	182
Table 50: The Krabbe priors (mean and standard error)	183
Table 51: Simulation indicators for the 19 designs	187
Table 52: Demographic characteristics of the overall sample, and seven overlap designs	194
Table 53: Demographic characteristics of the overall sample, and the 12 non-overlap designs	195
Table 54: Time taken overall and by design (in seconds)	196
Table 55: Drop out overall and by design (in seconds)	196
Table 56: Summary of feedback questions	198
Table 57: Conditional logit comparison of seven designs with overlap	199
Table 58: Conditional logit comparison of 12 non-overlapping designs (part 1)	200
Table 59: Conditional logit comparison of 12 non-overlapping designs (part 2)	201
Table 60: Assessment of scale across the overlap and no overlap designs	206
Table 61: Summary of latent class model performance	207
Table 62: The HUI-2 descriptive system	238
Table 63: The HUI-3 descriptive system	239
Table 64: How the AQOL-8D items contribute to the descriptive classification	241
Table 65: CREATE checklist	243
Table 66: PRISMA checklist	245
Table 67: EFA Orthogonal CF-Varimax model (Model 3)	267
Table 68: EFA Orthogonal quartimax model (Model 4)	270
Table 69: Item descriptions for the 91 items included in the IRT analyses	274
Table 70: Local dependencies Chi square values across item pairs – Physical functioning dimension^a	276
Table 71: Local dependence estimates - Mental health dimension^a	277
Table 72: Local dependence estimates - Pain dimension^a	277
Table 73: Local independence - Activities dimension^a	278
Table 74: Block of choice sets from the Chapter 5 DCE design	279
Table 75: Further exploratory interactions	293
Table 76: Conditional logit and pooled models by gender	295
Table 77: Conditional logit and pooled models by age	296
Table 78: Conditional logit and pooled models by condition status	297
Table 79: Conditional logit and pooled models by median time per task	300
Table 80: Conditional logit and pooled models by median overall completion time	301
Table 81: Three class latent class model (Model 67)	303

Table 82: Four class latent class model (Model 68)	304
Table 83: Five class latent class model (Model 69)	305
Table 84: Six class latent class model (Model 70)	306
Table 85: Mixed Logit models – Combining EQ-5D-5L and ASCOT as random parameters (overall level	307
Table 86: Further exploratory models with all parameters, EQ-5D-5L and ASCOT as random	308
Table 87: Mixed logit models – Further exploratory combinations of parameters in random	309
Table 88: Latent class demographic parameters (overlap designs A to D)	311
Table 89: Latent class demographic parameters (overlap designs E to G)	311
Table 90: Latent class demographic parameters (non-overlap designs H to I)	311
Table 91: Latent class demographic parameters (non-overlap designs J to K)	311
Table 92: Latent class demographic parameters (non-overlap designs L to M)	312
Table 93: Latent class demographic parameters (non-overlap designs N to O)	312
Table 94: Latent class demographic parameters (non-overlap designs P to S)	312

ABSTRACT

Economic evaluation is an important tool in health care resource allocation. Interventions are typically evaluated through a cost utility analysis (CUA) using the Quality Adjusted Life Year (QALY), a metric combining length of life and quality of life (QoL) into a single outcome. The quality aspect of the QALY is often provided by a preference-based measure (PBM) that includes a way of measuring health, and a preference-based value set. The most commonly used PBMs focus on health-related quality of life (HRQoL). However, there is a case for broadening what is measured and valued by including other aspects of QoL (such as social care related QoL) alongside HRQoL.

This thesis explores how methods for the measurement and valuation of health and QoL can be extended to inform the development of broader and more widely applicable instruments. This was investigated by first exploring how to incorporate QoL concepts into PBMs, and second by testing the further application of Discrete Choice Experiment (DCE) methods to value QoL. Three empirical studies were conducted

The first study assessed existing measures of health and QoL using Item Response Theory (IRT), and tested two ways in which PBMs could be broadened to incorporate wider QoL concepts. The results demonstrated overlap and divergence in what is measured. This informed where extra dimensions of QoL could broaden the information collected, and how the information collected within existing HRQoL frameworks could be extended.

The second study used DCE to understand respondent preferences for diverse dimensions of QoL. The results provided evidence respondents do trade across different concepts of QoL. This supports the need for broader measures, and also the use of DCE to value broader outcomes.

The third study focuses on DCE methods, and particularly on the construction of designs for DCEs. The results provided detailed information about different design strategies for the valuation of QoL outcomes.

The overall findings raise key issues about what should be captured in PBMs, and also provide novel information about methods that can be used to inform the assessment, development and valuation of future instruments. For example, the results inform how IRT can be used in PBM

development. They also suggest how DCE can be used to value diverse QoL concepts. This can inform the development and valuation of broader measurement systems of QoL outcomes that can increase the scope and enhance the applicability of QALY values used in resource allocation decision making.

LIST OF ABBREVIATIONS

Abbreviation	Description
15D	15 Dimension
A	Anxiety (PROMIS dimension)
ABC	Assessment of Burden of COPD
AC	Accommodation (ASCOT dimension)
AD	Anxiety/Depression (EQ-5D dimension)
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
AQoL	Assessment of Quality of Life
AQoL-8D	Assessment of Quality of Life – 8 Dimension
ASCOT	Adult Social Care Outcomes Toolkit
BIC	Bayesian Information Criterion
BPI	Behaviour Problems Index
BWS	Best Worst Scaling
CAT	Computer Adaptive Testing
CBA	Cost Benefit Analysis
CF	Crawford Ferguson
CFA	Confirmatory Factor Analysis
CL	Cleanliness (ASCOT dimension)
CO	Control (ASCOT dimension)
COPD	Chronic Obstructive Pulmonary Disease
CREATE	Checklist for Reporting Valuation Studies
CTT	Classical Test Theory
CUA	Cost Utility Analysis
D	Depression (PROMIS dimension)
DCE	Discrete Choice Experiment
DCE _{TTO}	Discrete Choice Experiment including duration
DEMQOL	Dementia Quality of Life
DI	Dignity (ASCOT dimension)
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
EORTC	European Organisation for Research and Treatment of Cancer
EQALY	Extended Quality Adjusted Life Year
FA	Fatigue (PROMIS dimension)
FD	Food and Drink (ASCOT dimension)
GDP	Gross Domestic Product
GH	General health (SF-36 dimension)
GMNL	Generalised Multinomial Logit model
HRQoL	Health-Related Quality of Life
HTA	Health Technology Assessment
HUI-2	Health Utility Index – Mark 2
HUI-3	Health Utility Index – Mark 3
ICECAP	ICEpop CAPability measure for Adults
ICER	Incremental Cost Effectiveness Ratio
IIA	Independence of Irrelevant Alternatives
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IRT	Item Response Theory

LL	Log-Likelihood
LLR	Log-Likelihood Restricted
LLU	Log-Likelihood Unrestricted
LR	Likelihood Ratio
LT-TTO	Lead Time – Time Trade-Off
MAUI	Multi-Attribute Utility Instrument
MH	Mental Health (SF-36 and SF-6D dimension)
MH-RM	Metropolis–Hastings Robbins-Munro algorithm
MIC	Multi Instrument Comparison
MNL	Multinomial Logit
MO	Mobility (EQ-5D dimension)
NICE	National Institute for Health and Care Excellence
OC	Occupation (ASCOT dimension)
ONS-4	Office of National Statistics – 4
OPUS	Older Person’s Utility Scale
PA	Pain (SF-36, SF-6D and PROMIS dimension)
PBAC	Pharmaceutical Benefits Advisory Committee
PBM	Preference-Based Measure
PD	Pain/Discomfort (EQ-5D dimension)
PF	Physical Functioning (SF-36, SF-6D and PROMIS dimension)
PICOS	Participants, interventions, comparisons, outcomes, and study
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMIS	Patient-Reported Outcome Measure Information System
PWI	Personal Wellbeing Index
QALY	Quality Adjusted Life Year
QoL	Quality of Life
RE	Role Emotional (SF-36 dimension)
RF	Role Functioning (SF-6D dimension)
RMSE	Root Mean Squared Error
RMSEA	Root Mean Squared Error of Approximation
RP	Role Physical (SF-36 dimension)
RUT	Random Utility Theory
SA	Safety (ASCOT dimension)
SAS	Statistical Analysis Software
SC	Self-Care (EQ-5D dimension)
SCRQoL	Social Care Related Quality of Life
SD	Standard Deviation
SF-6D	Short Form – 6 Dimension
SF-36	Short Form – 36
SF	Social Functioning (SF-36, SF-6D and PROMIS dimension)
SG	Standard Gamble
SL	Sleep (PROMIS dimension)
SP	Social Participation (ASCOT dimension)
SWLS	Satisfaction With Life Scale
TIF	Total Information Function
TTO	Time Trade-Off
UA	Usual Activities (EQ-5D dimension)
VAS	Visual Analogue Scale

VNM	Von Neumann and Morgenstern
VT	Vitality (SF-36 and SF-6D dimension)
WEMWBS	Warwick and Edinburgh Mental Wellbeing Scale
WHO	World Health Organisation