EXPLORING THE CAPABILITY APPROACH IN MODEL-BASED ECONOMIC EVALUATIONS

by

PAUL MARK MITCHELL

A thesis submitted to the
University of Birmingham
For the degree of
DOCTOR OF PHILOSOPHY

Health Economics Unit
Public Health Building
School of Health and Population Sciences
University of Birmingham
July 2013

ABSTRACT

This thesis develops the implementation of the capability approach within health economic evaluations. Until now, the focus of applying the capability approach within health economics has centred on its theoretical merits, as well as the development of capability questionnaires.

The aim of this research is to establish methods for applying the capability approach in an evaluation framework. Specifically, this is done by (i.) investigating how a measure of capability well-being, the ICECAP-O, can be incorporated into a health economic model and (ii.) establishing the objective of capability evaluations to aid the decision-making process in allocating scarce resources for health.

The relationship between capability and condition-specific health status for osteoarthritis patients is studied through statistical mapping. Methods from the capability literature are drawn upon to construct a methodology for generating capability outcomes that can be used to aid decision-making. This methodology is then tested on an existing economic model, the Birmingham Rheumatoid Arthritis Model (BRAM).

Key findings from this thesis are that (i.) it is feasible to predict capability from a condition-specific health instrument and (ii.) establishing “sufficient capability” as the objective for capability evaluations. Further research is required to see what difference a capability based evaluation would make in practice.

ACKNOWLEDGEMENTS

First, I would like to acknowledge my supervisors Joanna Coast, Tracy Roberts and Pelham Barton from the Health Economics Unit at the University of Birmingham. Without their expertise in the health economics discipline and their frequent constructive critiques of draft chapters, this thesis would not be what it is today without them. In particular, I would like to thank Jo for the original concept that led to the Ph.D. stendentship funding for this work. Her encouragement for me to think outside the box and develop personally by presenting work at national and international conferences is something that I will ever be grateful for.

Thanks must also go to Terry Flynn and Beth Pollard for providing me with access to the Tayside Joint Replacement dataset. To those patients who filled in the questionnaires, I really do appreciate your time and effort. For useful comments on this thesis work throughout my studentship, I would like to pay particular gratitude to Hareth Al-Janabi, Philip Kinghorn and Sabina Sanghera from Birmingham. Additional praise must also go to Aki Tsuchiya (Sheffield) and Carole Siani (Lyon) and particiapants of the Health Economists' Study Group for their constructive comments on taking my research forward. Two anonymous referees on the mapping work in Chapter 6 also deserve recognition for their input. Other comments at journal clubs, seminars and progress reviews also added value to the final thesis.

To the legions of colleagues at the Health Economics Unit who proof-read a Chapter of an earlier draft of this thesis (Lazaros Andronis; Alastair Canaway; Thomas Keeley; Eoin Moloney; Deirdre O'Brien; Raymond Oppong; Rosanna Orlando; Maria Cristina Peñaloza Ramos; Melcior Rossello-Roig) – I very much appreciate your time and effort too. To all of my colleagues who have had the misfortune of sharing an office with me, thanks for your patience, friendship and distraction from work when necessary.

Finally, to my parents, Mary and Padraig, and brothers, Colm and Stephen – thanks for your support over the last three and a half years. I hope you now think it was worth the hassle.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1

CHAPTER 2. HEALTH ECONOMICS: THEORY, EVALUATION AND OUTCOME MEASUREMENT 9

2.1 INTRODUCTION 9
2.2 HEALTH ECONOMIC THEORY: WELFARISM AND EXTRA-WELFARISM 11
2.2.1 WELFARISM 11
2.2.2 EXTRA-WELFARISM 13
2.3 MEASURING BENEFITS FOR ALLOCATING RESOURCES IN HEALTH ECONOMICS 16
2.3.1 QUALITY ADJUSTED LIFE YEARS 16
2.3.1.1 What attributes to value 18
2.3.1.2 How are the attributes valued 19
2.3.1.3 Who to value the attributes 21
2.3.2 ALTERNATIVES TO THE QALY 22
2.3.2.1 Willingness to Pay 22
2.3.2.2 Extra-welfarist QALY alternatives 24
2.4 EVALUATION FRAMEWORKS FOR HEALTH 27
2.4.1 COST-BENEFIT ANALYSIS (CBA) 27
2.4.2 COST-UTILITY ANALYSIS (CUA) 29
2.4.3 ALTERNATIVE EVALUATION FRAMEWORKS 31
2.5 MODELLING APPROACHES IN HEALTH ECONOMIC EVALUATIONS 33
2.6 DECISION RULES IN HEALTH ECONOMIC EVALUATIONS 38
2.7 CRITIQUING CURRENT HEALTH ECONOMICS PRACTICE 39
2.7.1 CRITIQUING THE QALY: EVALUATION SPACE 40
2.7.2 CRITIQUING THE QALY: UNDERLYING ASSUMPTIONS 42
2.7.3 CRITIQUING THE QALY: DECISION RULE 44
2.7.4 THE DURABILITY OF QALYS 45
2.8 CONCLUDING COMMENTS 46

CHAPTER 3. CAPABILITY, ECONOMICS AND HEALTH 48

3.1 INTRODUCTION 48
3.2 WHAT IS THE CAPABILITY APPROACH? 50
3.2.1 THEORETICAL ARGUMENT AGAINST WELFARE ECONOMICS: AMARTYA SEN AND THE CAPABILITY APPROACH 50
3.2.1.1 A Primer in Capability Terminology 51
3.2.1.2 Evaluation Base of the Capability Approach 53
3.2.1.3 Early Applications of the Capability Approach 56
3.2.1.4 Critiques of Applying the Capability Approach Quantitatively 57
3.2.1.4.1 Functionings or capabilities? 57
3.2.1.4.2 Which capabilities/functionings to measure? 58
3.2.1.4.3 The individual as the focus of analysis 60
3.2.1.4.4 Critiquing the link between the HDI and the capability approach 61
3.2.2 HOW IS THE CAPABILITY APPROACH DIFFERENT FROM HEALTH ECONOMIC THEORY? 61
3.3 CONCEPTUALISING THE CAPABILITY APPROACH FOR ASSESSING HEALTH 64
3.3.1 RUGER: HEALTH AND SOCIAL JUSTICE 65

3.3.1.1 Objectives of the Health Capability Paradigm ... 65
3.3.1.2 Decision Making Rules in the Health Capability Paradigm ... 68
3.3.1.3 Application requirements in the Health Capability Paradigm ... 70
3.3.1.4 Applying the Health Capability Paradigm in practice ... 70
3.3.2 VENKATAPURAM: HEALTH JUSTICE ... 73
3.3.3 COMPARING, CONTRASTING AND CRITIQUING CONCEPTUALISATIONS OF THE CAPABILITY APPROACH FOR HEALTH ... 77
3.4 ALIGNING THE CAPABILITY APPROACH WITH TRADITIONAL OUTCOMES ... 82
3.4.1 CAPABILITY AND THE QALY OUTCOME ... 82
3.4.2 CRITIQUING THE FORMULATION OF QALYS WITH THE CAPABILITY APPROACH ... 84
3.5 MEASURING CAPABILITY ... 86
3.5.1 CAPABILITY INDICATORS: THE OXCAP FAMILY OF INSTRUMENTS ... 87
3.5.2 ASCOT MEASURE OF SOCIAL CARE ... 92
3.5.3 ICECAP CAPABILITY QUESTIONNAIRES ... 94
3.5.4 CHOOSING A QUESTIONNAIRE FOR RESOURCE ALLOCATION ... 98
3.6 SUMMARY OF CAPABILITY, ECONOMICS AND HEALTH ... 101

CHAPTER 4. A REVIEW OF EMPIRICAL APPLICATIONS OF THE CAPABILITY APPROACH ACROSS DISCIPLINES ... 103

4.1 INTRODUCTION ... 103
4.2 PREVIOUS REVIEWS OF EMPIRICAL CAPABILITY APPLICATIONS ... 104
4.2.1 LITERATURE REVIEW (1): KUKLYS & ROBEYNS (2005) ... 105
4.2.2 LITERATURE REVIEW (2): ROBEYNS (2006) ... 106
4.2.3 LITERATURE SURVEY (3): CHIAPPERO-MARTINETTI & ROCHE (2009) ... 109
4.2.4 SUMMARY OF LITERATURE SURVEYS ON EMPIRICAL CAPABILITY RESEARCH ... 112
4.3 METHODS ... 114
4.3.1 SEARCH STRATEGY ... 114
4.3.2 INCLUSION CRITERIA AND PAPER CATEGORISATION ... 116
4.3.2.1 Stage I – Initial Categorisation of studies ... 117
4.3.2.2 Stage II – Further Categorisation of studies ... 119
4.3.3 SELECTION OF KEY PEARLS ... 119
4.3.4 DATA EXTRACTION ... 121
4.4 RESULTS ... 121
4.4.1 SUMMARY OF PEARL SEARCH ... 121
4.4.2 FINAL CLASSIFICATION OF STUDIES ... 123
4.5 ANALYSIS OF REVIEW ... 123
4.5.1 SUMMARY DATA ON STUDIES INCLUDED IN REVIEW ... 124
4.5.2 FOCUS OF STUDY: CAPABILITY, FUNCTIONING OR AGENCY? ... 125
4.5.3 ATTRIBUTES INCLUDED IN CAPABILITY RELATED MEASURES ... 126
4.5.3.1 Thematic Group i. General assessment of human development ... 127
4.5.3.2 Thematic Group ii. Assessing small scale development projects ... 128
4.5.3.3 Thematic Group iii. Identifying the poor in developing countries ... 128
4.5.3.4 Thematic Group iv. Poverty and well-being assessment in advanced economies ... 129
4.5.3.5 Thematic Group v. Deprivation of disabled people ... 130
4.5.3.6 Thematic Group vi. Assessing gender inequalities ... 130
4.5.3.7 Thematic Group vii. Debating policies ... 131
4.5.3.8 Thematic Group viii. Education ... 132
4.5.3.9 Thematic Group ix. Technology ... 132
4.5.3.10 Thematic Group x. Health ... 133
4.5.4 AGGREGATING CAPABILITY RELATED ATTRIBUTES WITHIN AN INDEX ... 134
4.5.4.1 Alkire and Foster Multidimensional Poverty Methods ... 134

4.5.4.2 Human Development Index and related measures 138
4.5.5 Exploring the objectives and decision rules within the capability approach 139
4.5.5.1 Is the QALY objective (health maximisation over time) compatible with the capability approach? 140
4.5.5.2 Capability Objectives and Decision Rules beyond health... 142
4.5.6 Outstanding issues.. 145
4.6 DISCUSSION 147

CHAPTER 5. CASE STUDY SELECTION AND MAPPING TERMINOLOGY.................................... 151

5.1 INTRODUCTION 151
5.2 CASE STUDY SELECTION.. 153
5.3 MAPPING TERMINOLOGY.. 158
5.4 GUIDANCE FOR MAPPING RESEARCH 161
5.4.1 Academic Guidance: Brazier et al. (2010).. 161
5.4.2 NICE Guidance: Longworth and Rowen (2011) 164
5.4.3 Summary of Guidance available at present 166
5.5 CONCLUDING COMMENTS... 167

CHAPTER 6. EXPLORING THE RELATIONSHIP BETWEEN HEALTH STATUS AND CAPABILITY: MAPPING FROM THE WOMAC OSTEOARTHRITIS INDEX TO THE ICECAP-O CAPABILITY INDEX 168

6.1 INTRODUCTION 168
6.2 CASE STUDY SELECTION.. 170
6.2.1 Western Ontario and McMaster (WOMAC) Osteoarthritis Index 170
6.2.2 ICECAP-O capability questionnaire 171
6.3 MAPPING FROM WOMAC TO HEALTH UTILITY... 171
6.3.1 Mapping from the WOMAC to the Health Utility Index Mark III (HUI3)............................... 172
6.3.1.1 Grootendorst et al. (2007) 172
6.3.2 Mapping from the WOMAC to the EQ-5D.. 174
6.3.2.1 Barton et al. (2008) 175
6.3.2.2 Xie et al. (2010) 177
6.3.3 Summary of Mappings from WOMAC to health utility 178
6.4 MAPPING FROM WOMAC TO ICECAP-O: A CASE STUDY 179
6.4.1 Methods 179
6.4.1.1 Dataset 179
6.4.1.2 Measures 179
6.4.2 Statistical Analysis... 180
6.4.2.1 WOMAC Imputation 180
6.4.2.2 Regression Specification... 181
6.4.2.2.1 Ordinary Least Squares (OLS) and ICECAP-O score 181
6.4.2.2.2 Multinomial Logistic (ML) Regression and ICECAP-O dimensions 182
6.4.2.3 Prediction Accuracy.. 184
6.4.3 Results 185
6.4.3.1 Demographics of Dataset 185
6.4.3.2 WOMAC Imputation 186
6.4.3.3 Relationship between ICECAP-O and WOMAC score 187
6.4.3.4 Model Specification 189
6.4.3.5 Internal Validation Results.. 190
6.4.3.6 Recommended Mapping Model 192
6.4.3.7 What aspects of ICECAP-O are WOMAC categories predicting? 192
6.5 DISCUSSION 194

CHAPTER 7. SUFFICIENT CAPABILITY 197

7.1 INTRODUCTION 197
7.2 METHODS 199
7.2.1 MEASURE OF CAPABILITY: ICECAP-O 199
7.2.2 DATASET 200
7.2.3 ALKIRE-FOSTER METHOD OF MULTIDIMENSIONAL POVERTY MEASURES 200
7.2.4 THRESHOLD OF SUFFICIENT CAPABILITY 201
7.2.5 SUFFICIENT CAPABILITY SCORE 206
7.2.6 SUFFICIENT CAPABILITY OVER TIME 208
7.2.6.1 Poverty Free Years (capability) 209
7.2.6.2 Years of Sufficient Capability (equivalent) 210
7.2.6.3 Years of Insufficient Capability (equivalent) 210
7.2.6.4 Years of Full Capability (equivalent) 211
7.3 RESULTS 211
7.3.1 DEMOGRAPHICS 212
7.3.2 ALKIRE-FOSTER MEASURES 212
7.3.3 SUFFICIENT CAPABILITY FOR THRESHOLD OPTION 1 (33333) 214
7.3.4 SUFFICIENT CAPABILITY FOR THRESHOLD OPTION 2 (22222) 216
7.3.5. SUFFICIENT CAPABILITY OUTCOMES OVER TIME 218
7.3.5.1 Poverty Free Years (capability) – (PFY(c)) 218
7.3.5.2 Years of Sufficient Capability – (YSC) 219
7.3.5.3 Years of Insufficient Capability – (YIC) 220
7.3.5.4 Years of Full Capability – (YFC) 221
7.4. DISCUSSION 222

CHAPTER 8. METHODOLOGY FOR OBTAINING CAPABILITY OUTCOMES FROM AN ECONOMIC MODEL: A CASE STUDY 227

8.1 INTRODUCTION 227
8.2 BACKGROUND TO THE BIRMINGHAM RHEUMATOID ARTHRITIS MODEL (BRAM) 228
8.2.1 A REVIEW OF PREVIOUS BRAM VERSIONS 228
8.3 METHODS 237
8.3.1 BRAM VERSION 237
8.3.2 DATASET 238
8.3.3 INSTRUMENTS 238
8.3.3.1 Starting Measure: Amended version of the Health Assessment Questionnaire – Disability Index (HAQ-DI) 239
8.3.3.2 Target Measure: ICECAP-O capability questionnaire 241
8.3.4 CAPABILITY OUTCOMES 241
8.3.4.1 Capability Outcome 1 (CO1): Years of Full Capability (YFC) 242
8.3.4.1.1 ICECAP-O as a continuous dependent variable 243
8.3.4.1.2 ICECAP-O attributes (5) as categorical dependent variables 243
8.3.4.1.3 Prediction Measures 244
8.3.4.1.4 Prediction Specification 245
8.3.4.2 Capability Outcome 2 (CO2): Poverty Free Years (capability) (PFY(c)) 246
8.3.4.3 Capability Outcome 3 (CO3): Insufficient Capability Score (ICS) 247
8.3.4.4 Capability Outcome 4 (CO4): Sufficient Capability Score (SCS) 249
8.4 RESULTS 250
8.4.1 DESCRIPTIVE STATISTICS 250
8.4.2 CAPABILITY OUTCOME (1): YEARS OF FULL CAPABILITY 251
8.4.3 CAPABILITY OUTCOME (2): POVERTY FREE YEARS (CAPABILITY) (PFY(C)) 254

8.4.3 Capability Outcome (3): Years of Insufficient Capability (YIC) ... 256
8.4.4 Capability Outcome (4): Years of Sufficient Capability (YSC) ... 257
8.5 DISCUSSION ... 258
8.5.1 Capability Outcome (1): Years of Full Capability (YFC) ... 258
8.5.2 Capability Outcome (2): Poverty Free Years (capability) (PFY(c)) ... 260
8.5.3 Capability Outcome (3): Years of Insufficient Capability (YIC) ... 262
8.5.4 Capability Outcome (4): Years of Sufficient Capability (YSC) ... 263
8.6 CONCLUDING THOUGHTS ... 264

CHAPTER 9. GENERATING CAPABILITY OUTCOMES FROM THE BIRMINGHAM RHEUMATOID ARTHRITIS MODEL: RESULTS ... 265

9.1 INTRODUCTION ... 265
9.2 MODEL SPECIFICATION ... 266
9.2.1 BRAM11 – Modelling Benefits ... 266
9.2.2 BRAM11 – Modelling Costs ... 269
9.2.3 BRAM11 Results ... 269
9.3 BRAM12 ... 271
9.4 RESULTS ... 272
9.4.1 Capability Outcome (1): Years of Full Capability (YFC) ... 272
9.4.2 Capability Outcome (2): Poverty Free Years (capability) (PFY(c)) ... 273
9.4.3 Capability Outcome (3): Years of Insufficient Capability (YIC) ... 274
9.4.4 Capability Outcome (4): Years of Sufficient Capability (YSC) ... 275
9.4.5 Comparison of Model results ... 276
9.5 DISCUSSION OF EACH MODEL RUN IN BRAM12 ... 277
9.5.1 Capability Outcome (1): Years of Full Capability (YFC) ... 277
9.5.2 Capability Outcome (2) Poverty Free Years (capability) (PFY(c)) ... 280
9.5.3 Capability Outcome (3): Years of Insufficient Capability (YIC) ... 283
9.5.4 Capability Outcome (4): Years of Sufficient Capability (YSC) ... 285
9.6 DISCUSSION COMPARING BRAM12 MODEL RESULTS ... 288

CHAPTER 10. THESIS DISCUSSION AND CONCLUSION ... 293

10.1 INTRODUCTION ... 293
10.2 KEY FINDINGS ... 294
10.2.1 The illustration and development of the sufficient capability methodology ... 295
10.2.2 Development of capability outcomes ... 299
10.2.3 Understanding the relationship between health and capability wellbeing ... 300
10.3 STRENGTHS OF RESEARCH ... 301
10.3.1 Aligning capability conceptual theory with questionnaire development ... 301
10.3.2 Implementing an explicit literature search strategy for capability reviews ... 302
10.3.3 The feasibility of mapping from health status to capability ... 303
10.3.4 The development of a capability objective for economic evaluations ... 303
10.3.5 The incorporation of capability outcomes within an economic model ... 304
10.4 LIMITATIONS OF RESEARCH ... 304
10.4.1 Capability empirical review search strategy ... 305
10.4.2 The relationship between health status and capability ... 305
10.4.3 Defining Sufficient Capability ... 306
10.4.4 Methods for incorporating ICECAP-O into an economic model case study ... 307
10.4.5 The BRAM case study ... 308
10.5 IMPLICATIONS OF FINDINGS FOR HEALTH ECONOMICS AND DECISION-MAKERS IN HEALTHCARE ... 310
10.5.1. A new method for allocating resources across healthcare ... 310

10.5.2 SUFFICIENT CAPABILITY AND EXTRA-WELFARISM 311
10.6 FUTURE RESEARCH RECOMMENDATIONS 313
10.6.1 THE RELATIONSHIP BETWEEN CAPABILITY AND HEALTH MEASURES.. 313
10.6.2 DEFINING SUFFICIENT CAPABILITY FOR A SOCIETY 314
10.7 CONCLUSION.. 317
APPENDIX 1: THE ICECAP-O QUESTIONNAIRE(COAST ET AL., 2008A) 318
APPENDIX 2: DATA EXTRACTION SHEET FOR CAPABILITY APPLICATION REVIEW 319
APPENDIX 3: STUDIES INCLUDED IN REVIEW OF CAPABILITY APPLICATIONS 320
APPENDIX 4: STUDIES EXCLUDED AT THE SECOND STAGE OF CATEGORISATION 330
APPENDIX 5: WESTERN MCMASTERS AND ONTARIO (WOMAC) OSTEOARTHRITIS INDEX 333
APPENDIX 6: THE HAQ-DI (ADJUSTED FROM BARTON ET. AL., 2004A) 334
APPENDIX 7: AMENDED HAQ-DI (WORKED EXAMPLE).. 335
APPENDIX 8: CONSOLIDATED HEALTH ECONOMIC EVALUATION REPORTING STATEMENT FOR BRAM12 336

List of Figures

FIGURE 1 THE INTERRELATIONSHIPS BETWEEN CHAPTERS IN THIS THESIS 8
FIGURE 2 OPTIONS FOR ASSESSING SOCIETAL WELFARE IN THE CAPABILITY APPROACH 56
FIGURE 3 THE HEALTH CAPABILITY PARADIGM (RUGER 2010A) 66
FIGURE 4 AN ILLUSTRATION OF THE COMPREHENSIVE PEARL GROWING METHOD 118
FIGURE 5 SUMMARY STATISTICS OF COMPREHENSIVE PEARL GROWING REVIEW RESULTS 122
FIGURE 6 POPULATION WHERE THE CAPABILITY STUDIES IDENTIFIED WERE APPLIED 124
FIGURE 7 NUMBER OF STUDIES PER CAPABILITY THEMATIC GROUP 125
FIGURE 8 STUDY FOCUS OF CAPABILITY APPLICATION REVIEW (2006-2012) 126
FIGURE 9 DISTRIBUTION OF ICECAP-O IN PREDICTION DATASET 185
FIGURE 10 RELATIONSHIP BETWEEN ICECAP-O AND WOMAC SCORES (BASELINE) 188
FIGURE 11 EXAMPLE OF ALKIRE-FOSTER MEASURES OF MULTIDIMENSIONAL POVERTY 202
FIGURE 12 SHORTFALLS IN SUFFICIENT CAPABILITY FOR THRESHOLD OPTION 1 (“33333”) 208
FIGURE 13 EXAMPLE OF YEARS OF SUFFICIENT CAPABILITY (YSC) FOR OSTEOARTHRITIS PATIENTS (N=42) FOR THRESHOLD OPTION 1 “33333” 220
FIGURE 14 TREATMENT STRATEGY OPTIONS AND PATIENT PATHWAY REPRESENTED IN BPM02 230
FIGURE 15 INDIVIDUAL PATIENT PATHWAY IN BRAM04, BRAM06 AND BRAM11 MODELS(BARTON ET AL., 2004A) 232
FIGURE 16 REPRESENTATION OF HAQ-DI CHANGE OVER TIME FOR PATIENTS IN BRAM MODELS (BRAM04, BRAM06 AND BRAM11) (MALOTTKI ET AL., 2011) 234
FIGURE 17 TWO PART MODEL PROCESS TO CALCULATE SUFFICIENT CAPABILITY SCORE (SCS) 249
FIGURE 18 RELATIONSHIP BETWEEN ICECAP-O AND HAQ-DI 251

List of Tables

TABLE 1 EXAMPLES OF CAPABILITY LISTS 59
TABLE 2 THE OXCAP FAMILY OF QUESTIONNAIRES 89
TABLE 3 ROBEYNS' (2006) OVERVIEW OF WHERE THE CAPABILITY APPROACH HAS BEEN APPLIED 108
TABLE 4 KEY PEARL REFERENCES FOR IDENTIFYING STUDIES IN WAVE 1 120
TABLE 5 MODEL CASE STUDY SELECTION 157
TABLE 6 MAPPING MODELS FOR ICECAP-O PREDICTION FROM WOMAC 184
TABLE 7 DESCRIPTIVE STATISTICS FOR PREDICTION AND VALIDATION DATASET 186
TABLE 8 PREDICTION DATASET - ICECAP-O RESPONSES AND WOMAC SCORES 189
TABLE 9 INTERNAL VALIDATION: MAPPING MODEL RESULTS 191
TABLE 10 BEST PERFORMING MAPPING ALGORITHM: 192
TABLE 11 PREDICTION OF ICECAP-O DIMENSIONS FROM WOMAC CATEGORIES: MULTINOMIAL LOGISTIC (ML) REGRESSION P-VALUES 193
TABLE 12 THRESHOLDS OF SUFFICIENT CAPABILITY (TSC) ON THE ICECAP-O 204
TABLE 13 SHORTFALL IN SUFFICIENT CAPABILITY ON ICECAP-O FOR INDIVIDUAL A; (THRESHOLD OPTION 1 "33333") 207
TABLE 14 BASELINE DESCRIPTIVE STATISTICS FOR TAYSIDE REPLACEMENT DATASET 212
TABLE 15 AF POVERTY METHODS APPLIED TO ICECAP-O LEVELS USING "33333" AS THE POVERTY THRESHOLD (N=42) 213
TABLE 16 AF POVERTY METHODS APPLIED TO ICECAP-O LEVELS USING "22222" AS THE THRESHOLD (N=42) 213
TABLE 17 ICECAP-O VALUATION DATASET RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY (N=313) 214
TABLE 18 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "33333"(BASELINE) 214
TABLE 19 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "33333" (BASELINE AND 1 YEAR POST OPERATION; N=58) 215
TABLE 20 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "33333" (BASELINE AND THREE YEAR POST-OPERATION; N=55) 216
TABLE 21 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "22222"(BASELINE) 217
TABLE 22 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "22222" (BASELINE AND 1 YEAR POST OPERATION; N=58) 217
TABLE 23 RESPONSES BELOW THRESHOLD OF SUFFICIENT CAPABILITY "22222" (BASELINE AND THREE YEAR POST-OPERATION; N=55) 217
TABLE 24 CAPABILITY OUTCOMES FOR TAYSIDE REPLACEMENT DATASET (N=42) 222
TABLE 25 BRAM MODELS AND QUALITY OF LIFE CALCULATION FOR QALYS 237
TABLE 26 AGE AND SEX DISTRIBUTION WHEN STARTING TREATMENT IN BRAM11 238
TABLE 27 DISTRIBUTION OF HAQ-DI SCORES AMONG POPULATION WHEN STARTING TREATMENT IN BRAM11 238
TABLE 28 AGE AND SEX POPULATION FOR PREDICTION DATASET 241
TABLE 29 DISTRIBUTION OF AMENDED HAQ-DI SCORES IN PREDICTION DATASET 241
TABLE 30 DESCRIPTIVE STATISTICS FOR PREDICTION DATASET 250
TABLE 31 PREDICTION OF ICECAP-O SCORE AND DIMENSIONS TO GENERATE YFC 252
TABLE 32 PREDICTING THE ICECAP-O FROM THE AMENDED HAQ-DI: MAPPING MODEL 253
TABLE 33 PREDICTED YEARS OF FULL CAPABILITY FOR GIVEN HAQ-DI SCORE 253
TABLE 34 CAPABILITY POVERTY STATISTICS IN TAYSIDE JOINT REPLACEMENT COHORT 254
TABLE 35 PROBABILITY THAT INDIVIDUAL HAS REACHED THRESHOLD OF SUFFICIENT CAPABILITY (TSC) GIVEN THEIR HAQ-DI SCORE FOR ALL, MALES AND FEMALES 255
TABLE 36 PREDICTED INSUFFICIENT CAPABILITY SCORE FOR GIVEN HAQ-DI SCORE 257
TABLE 37 PREDICTED SUFFICIENT CAPABILITY SCORE FOR GIVEN HAQ-DI SCORE 257
TABLE 38 SUFFICIENT CAPABILITY AMONG ICECAP-O VALUATION SAMPLE OF UK ELDERLY POPULATION (COAST ET AL., 2008A) 262
TABLE 39 BRAM11 DETERMINISTIC RESULTS 270
TABLE 40 YEARS OF FULL CAPABILITY (YFC) & ICER (100,000 PATIENTS) 273
TABLE 41 POVERTY FREE YEARS (CAPABILITY) - PFY(C) & ICER (40,000 PATIENTS) 274
TABLE 42 YEARS IN INSUFFICIENT CAPABILITY (YIC) & ICER (100,000 PATIENTS) 275
TABLE 43 YEARS OF SUFFICIENT CAPABILITY (YSC) & ICER (100,000 PATIENTS) 276
TABLE 44 SUMMARY RESULTS FROM BRAM12 CASE STUDY 277

Abbreviations

anti-TNFs	anti-bodies against tumour necrosis factor	M_1	Adjusted Poverty Gap
ABT	abatacept	M_2	Adjusted Foster-Greer-Thorbecke measure
ADA	adalimumab	MAE	Mean Absolute Error
AF	Alkire-Foster methods of multidimensional poverty	MCDA	Multi Criteria Decision Analysis
ASCOT	Adult Social Care Outcomes Toolkit	ML	Multinomial Logistic Regression
BPM02	Birmingham Preliminary Model 2002 version	MOPSU	Measuring Outcomes for Public Service Users
BRAM	Birmingham Rheumatoid Arthritis Model	MPI	Multidimensional Poverty Index
BRAM04	BRAM 2004 version	MTX	methotrexate
BRAM08	BRAM 2008 version	NHS	National Health Service
BRAM11	BRAM 2011 version	NMB	Net Monetary Benefit
BRAM12	BRAM 2012 version	OA	Osteoarthritis
BWS	Best-worst scaling	OCAP	Oxford Capability Questionnaire Survey
CAPINDEX16	Mental Health Capability Index	OCAP 18	18 item capability questionnaire for public health
CBA	Cost Benefit Analysis	OL	Ordinal Logistic Regression
CCA	Cost Consequences Analysis	OPUS	Older People's Utility Scale
CEA	Cost Effectiveness Analysis	OxCAP-MH	Oxford Capability Measure for Mental Health
CMA	Cost Minimisation Analysis	PFY(c)	Poverty Free Years (capability)
CUA	Cost Utility Analysis	PSA	Probabilistic Sensitivity Analysis
DALY	Disability-Adjusted Life Year	PTO	Person Trade-Off
DII	Disaster Impact Index	QALY	Quality-adjusted life year
DMARDs	Disease modifying anti-rheumatic drugs	QoL	Quality of Life
EQ-5D	EuroQol 5 dimension	QSE	Quasi Standard Error
EQ-5D-3L	EQ-5D 3 level version	Quasi-CI	Quasi Confidence Intervals
EQ-5D-5L	EQ-5D 5 level version	R^2	Pearson product-moment correlation coefficient
GDP	Gross Domestic Product	RMSE	Root Mean Squared Error
GST	Gold/sodium aurothiomalate	RTX	Rituximab
H	Headcount ratio	SAVE	Saved Young Life Equivalent
HAQ	Health Assessment Questionnaire	SC-QALY	Social Care QALY
HAQ-DI	Health Assessment Questionnaire-Disability Index	SCRQOL	Social Care Related Quality of Life
HDI	Human Development Index	SF-6D	Short Form 6 dimension
HDR	Human Development Report	SF-12	Short Form 12 health survey
HRQoL	Health Related Quality of Life	SF-36	Short Form 36 health survey
HUI	Health Utilities Index	SCS	Sufficient Capability Score
HUI3	Health Utilities Index Mark III	SG	Standard Gamble
HYE	Healthy Year Equivalent	SSZ	sulphasalazine
ICECAP-A	ICEpop CAPability measure for Adult population	TSC	Threshold of Sufficient Capability
ICECAP-O	ICEpop CAPability measure for Older population	TTO	Time trade-off
ICECAP-SCM	ICEpop CAPability Supportive Care Measure	UN	United Nations
ICER	Incremental Cost Effectiveness Ratio	UNDP	United Nations Development Programme
ICS	Insufficient Capability Score	WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
IFX	infliximab	WTP	Willingness to pay
ITA	Incompletely theorized agreements	YIC	Years of Insufficient Capability
LEF	Leflunomide	YFC	Years of Full Capability (equivalent)
LIKP	Lifestyle Intervention for Knee Pain	YSC	Years of Sufficient Capability
M_0	Adjusted Headcount Ratio		

CHAPTER 1. INTRODUCTION

The study of the economics of health and healthcare has grown significantly in the past fifty years, ever since Nobel Laureate in Economics Kenneth Arrow wrote his seminal paper on the welfare economics of medical care in 1963 (Arrow, 1963). Economic evaluations are an analytical approach to assessing the benefits of competing resources in relation to their cost (Morris et al., 2007). Health economics, as a standalone social science discipline, has developed a number of unique methods for measuring the benefits of health interventions, which are for the most part, focused on the quantification of health benefits from interventions. This is in contrast with the majority of UK public policy economic assessments, which continue to focus on the monetary valuation of benefits (HM Treasury, 2003).

The role of health economic evaluations in aiding decision-making has grown significantly within the UK since the foundation of the advisory body for health guidance, the National Institute for Health and Care Excellence (NICE) in 1999. Since then, NICE has stipulated the requirement for economic evaluations for new interventions to be conducted before these interventions can be recommended for use within the National Health Service (NHS) (NICE, 2004; NICE, 2009a; NICE, 2013).

The Quality Adjusted Life Year (QALY) is the primary outcome measure of benefit in the majority of economic evaluations for NICE (NICE, 2013). The QALY combines health related quality of life (HRQoL) with length of time in that health state and changes over time to generate a single outcome of health status over time (Weinstein et al., 2009). The QALY has faced a number of criticisms since it has been developed, from both the theoretical assumptions underpinning the outcome measure (Carr-Hill, 1989; Loomes & McKenzie, 1989), as well as the considerations that are overlooked within the measure (Nord, 1999).

The objective of the QALY within NICE is to provide a generic measure of health over time, so that all interventions across the health service are treated in an equivalent manner in terms of the potential benefit accrued from a given intervention (NICE, 2013). To implement such an outcome in practice, a “reference case” HRQoL measure is required. The EQ-5D questionnaire (Brooks, 1996) is the currently recommended measure by NICE (NICE, 2013). However, notable difficulties of using a generic health questionnaire to capture all changes in health status for a number of conditions have been found, so alternative HRQoL measures (Brazier et al., 2002; Furlong et al., 2001), bolt-on dimensions to the EQ-5D (Lin et al., 2013) and condition-specific instruments (Brazier & Tsuchiya, 2010) to generate QALYs have been proposed instead. The development of alternative HRQoL instruments has led to the difficulty in comparing QALY outcomes produced via the original EQ-5D compared to alternative HRQoL measures (Mortimer & Segal, 2008). This is because alternative HRQoL measures, bolt-on dimensions or condition-specific instruments are likely to capture different dimensions than the reference case advocated by guidance bodies like NICE.

Additionally, there are a number of theoretical arguments against the use of the QALY outcome for measuring the benefits from health interventions. One such argument is the focus on changes in individual health status only, rather than a more holistic measure of individual welfare which would capture the broader benefits to individual wellness from healthcare

(Dolan et al., 2005b). This has led to some health economists for an outcome of benefit more in line with traditional economic theory to focus on utility from health interventions, such as willingness to pay (WTP) questionnaires (McIntosh et al., 2010; Donaldson et al., 2012).

An alternative proposal to the welfarist (through WTP) and extra-welfarist (through HRQoL and QALYs) approaches to measure benefits is the capability approach. The capability approach, developed originally by Amartya Sen (Sen, 1985; Sen, 1992; Sen, 1993; Sen, 2009), is a prominent critique of standard welfare economic theory. Sen argues that standard welfare economic theory is used to evaluate societal well-being through a narrow focus on a person's utility levels (Sen, 1980). The capability approach has been used to justify the use of the QALY (Culyer, 1989; Cookson, 2005b; Bleichrodt & Quiggin, 2013). Nonetheless, a number of health economists believe that a broader outcome measure based on individual capabilities rather than HRQoL would be a more appropriate implementation of the capability approach within a health economics framework (Verkerk et al., 2001; Grewal et al., 2006; Coast et al., 2008c; Lorgelly et al., 2010a; Kinghorn, 2010; Smith et al., 2012; Payne et al., 2013).

The use of the capability approach directly in the health economics field has so far focused on the development of capability questionnaires (Coast et al., 2008a; Lorgelly et al., 2008; Anand et al., 2009; Al-Janabi et al., 2012a). Less progress has been made on how such questionnaires, once fully developed and validated, can be used within an economic evaluation framework to aid priority setting in healthcare for advisory bodies like NICE. This thesis aims to further develop the implementation of the capability approach within health economics by addressing how capability questionnaires can be incorporated within an evaluation framework to aid decision making by allocating resources from a capability perspective.

The thesis is structured as follows;

In Chapter 2, the theory and use of economic evaluations in aiding decision-making in healthcare are examined. The chapter begins by examining the theory which underpins the current standard evaluation framework in health economics, known as “extra-welfarism” (Culyer, 1989). The evaluation frameworks that have emerged from extra-welfarism are compared with welfare economic methods to analysing healthcare. The QALY outcome is explained in detail as well as alternative outcomes that have been proposed. The difficulties of capturing benefits from healthcare are discussed in terms of the types of economic models that have been developed to model costs and outcomes from the dynamic nature of individual welfare over time. Such decision models are crucial in aiding decision-making for new treatments of complex health interventions. The chapter closes with a critique of the QALY in terms of the narrow evaluation space inferred from its focus on health alone.

Chapter 3 discusses an alternative theoretical base to the current approaches used within health economics. The theory behind the capability approach is detailed and compared with the alternative frameworks discussed in Chapter 2. Two possible means by which the capability approach could be applied to health are elaborated and critiqued from a health economics perspective. The focus then shifts from theoretical arguments to practical application of the capability approach. Two attempts have been made to combine capability theory with the QALY instrument. Alternative approaches through developing capability questionnaires for assessing interventions are then explored.

Chapter 3 raises a number of potential methods and different interpretations as to how questionnaires are aligned to capability theory. In Chapter 4 a literature review is conducted to explore how the capability approach has been applied across disciplines to assess capability and inform decision-making. A summary of previous reviews of empirical capability applications are first detailed. The original literature review in Chapter 4 attempts to highlight how recent measures of capability have been developed and the objective of such instruments once aggregated.

Chapter 5 begins the process of implementing capability questionnaires within a health economic evaluation, by identifying an appropriate case study for this thesis. A difficulty with the lack of routine collection of new capability questionnaires meant that the availability of longitudinal data on capability questionnaires were scarce at the beginning of this research project. While routine collection of capability questionnaires in clinical trials is improving (Henderson et al., 2013), it will take time to build a portfolio of studies across the many different types of conditions that currently exist. Therefore, alternative methods for incorporating capability questionnaires within an economic evaluation are explored through a process known as mapping, which allow for the generation of economic outcomes. Mapping between two instruments allows for the prediction of one instrument from another instrument.

This approach is already used in health economics when measures used to generate QALYs have not been collected. The guidance for researchers new to mapping is limited, although two recent notable publications are discussed (Brazier et al., 2010; Longworth & Rowen, 2011).

Chapter 6 explores the possibility of mapping to predict capability from condition-specific health instruments. The collection of capability data longitudinally at the time of this research was scarce, so mapping offered a way of incorporating capability questionnaires into decision models. Before implementing a capability instrument in an economic evaluation, a predictive relationship between a generic measure of capability and a health status instrument was first required, so that a prediction of capability from health was feasible. A dataset of osteoarthritis patients requiring knee or hip replacement was identified containing a measure of capability, the ICEPOP Capability questionnaire for older people – the ICECAP-O (Coast et al., 2008a), and a condition-specific instrument, the WOMAC Osteoarthritis index (Bellamy et al., 1988). The objective of this chapter is to assess the feasibility of mapping between these two instruments and to produce a method of mapping to a decision model, which is applied later in the thesis.

In Chapter 7, the findings from the review of capability questionnaire applications in Chapter 4 are drawn upon to develop an appropriate objective for capability outcomes in health economics. “Sufficient capability” is used to assess the improvements of capability below a threshold level deemed as sufficient. A number of potential capability outcomes are explored with a sample of patients completing the ICECAP-O at three time points to illustrate the calculations of these outcomes over time.

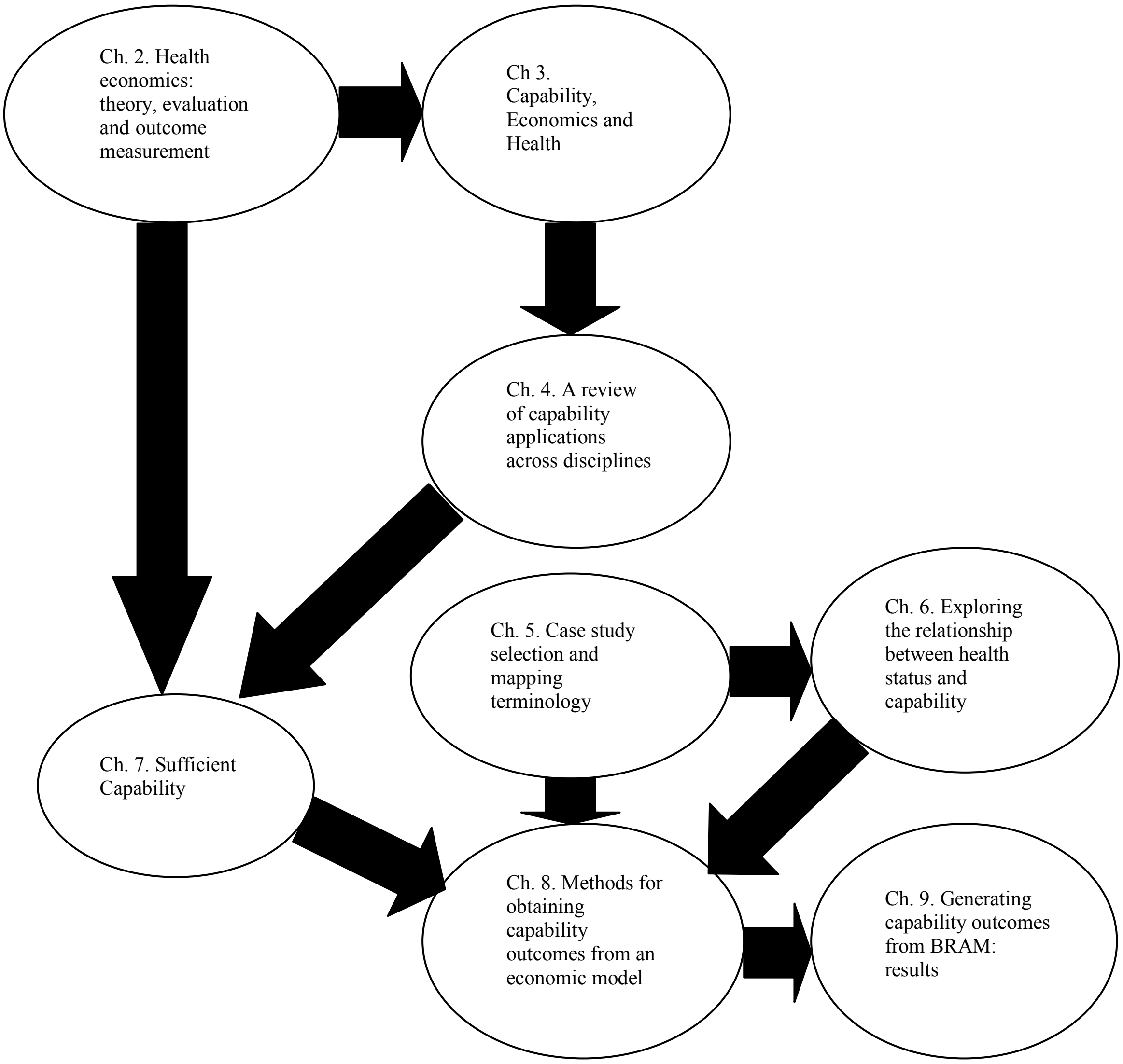
Building on the previous two chapters of implementing the use of a capability instrument within a mapping study (Chapter 6) and the sufficient capability objective (Chapter 7), in Chapter 8, a case study is used to illustrate the capability outcomes in a decision model. In Chapter 5, the Birmingham Rheumatoid Arthritis Model (BRAM) was identified as the most relevant case study given the data availability at the time (Malottki et al., 2011). The BRAM relied on a mapping between the Health Assessment Questionnaire Disability Index (HAQ-DI) to predict HRQoL for QALY calculations to assess different drug treatment strategies for Rheumatoid Arthritis patients within the UK. This case study allows the direct application of a capability instrument through a similar mapping process used in Chapter 6 and to test the alternative capability outcomes developed in Chapter 7 within an economic model used previously to aid decision-making.

Chapter 9 presents the results of the case study. Previous BRAM results are compared with the capability outcomes generated in the mapping study detailed in the preceding Chapter (Chapter 8). The cost per unit change of capability gained varied considerably depending on the outcome implemented within the model.

In Chapter 10 an overall discussion and conclusion to the thesis is presented. The principal findings of the thesis are highlighted. Strengths and weaknesses of the thesis are also discussed. For decision-makers interested in using a capability instrument, a number of policy implications from this research are outlined. Finally, future research directions on the further implementation of the capability approach within health economics are suggested.

Figure 1 presents the inter-relationships between the thesis and how they relate to one another in developing the capability approach within model-based economic evaluations.

Figure 1 The interrelationships between Chapters in this thesis



CHAPTER 2. HEALTH ECONOMICS: THEORY, EVALUATION AND OUTCOME MEASUREMENT

2.1 INTRODUCTION

While there are many interpretations about what economic analysis involves, Backhouse and Medema (2009) explored some of the common definitions of economic analysis in textbooks. The following quotation gives a flavour of some of the areas which are covered within the economic discipline (Backhouse & Medema, 2009):

> *"Thus, economics is apparently the study of the economy, the study of the coordination process, the study of the effects of scarcity, the science of choice, and the study of human behaviour"* (Backhouse & Medema, 2009, p. 222)

Economics has evolved dramatically since the work of classical economists in the eighteenth century, like Adam Smith, who was primarily concerned with factors which influenced the wealth of nations and societal welfare from economic progress more generally (Smith, 1776). Increasingly, economics has been applied in areas which have less of a direct link to the traditional role of economic attempts to understand society. The primary concern of economic analysis in healthcare is to do with understanding the role of allocating scarce healthcare resources across a national service (Morris et al., 2007). This chapter aims to explore the leading theories within health economics today, with a particular focus on the justifications provided for economic evaluations used to aid the decision-making process in healthcare resource allocation presently.

In Section 2.2, the role of economic theory in developing a framework by which healthcare ought to be examined is summarised. Particular attention is given to the shift in health economics from standard welfare economic theory to something which has become known as extra-welfarism. The reasoning behind the need for an alternative theoretical framework within health is detailed within this section. In Section 2.3, focus turns to the types of outcomes applied within health economic evaluations to measure individual improvements in well-being, how these outcomes are formulated as well as the differences between the outcomes. In Section 2.4, the types of economic evaluations which have been applied within the health economics literature are described. Primary attention is given to the dominant evaluation framework within the UK, cost-utility analysis (CUA). The reasons CUA has developed a dominant role within health care evaluation is further elaborated in this section. In Section 2.5, the role of modelling within health economics is detailed. Due to the complex nature of how individual health can change over time, a number of different ways of accounting for such changes within an evaluation based on economic models have been developed. Such modelling approaches, as well as how outcomes are incorporated within these evaluations are dealt with in section 2.5. In Section 2.6, the types of decision rules used to compare competing interventions are discussed. A review of critiques of the primary health economic outcome, the Quality-Adjusted Life Year (QALY), is presented in Section 2.7. Attention is focused on the evaluative space, the underlying assumptions and decision rules used in applying the outcome in economic assessments. The conclusion to the chapter summarises the state of affairs of health economic evaluation practice at present.

2.2 HEALTH ECONOMIC THEORY: WELFARISM AND EXTRA-WELFARISM

Any normative formulation of what society ought to be needs a form of ethical theory (Broome, 2009). This section focuses on the two prevailing theoretical bases for conducting economic evaluation in healthcare analysis. The first theory considered is the use of standard welfare economic theory as the basis for evaluation, generally referred to as welfarism (Brouwer et al., 2008). The second approach looks at tackling some of the problems associated with the application of standard welfare economic theory to healthcare, but still maintaining principles related to the ideals from welfare economics. This theoretical basis is generally referred to as extra-welfarism (Brouwer et al., 2008).

2.2.1 Welfarism

Alongside the numerous definitions used to define economics, welfarism is a term that has many interpretations, as it has been applied in a variety of ways. When referring to welfarism, welfarist or welfare economics in this thesis, it is within the interpretation as noted by Sen as a focus on individual utilities only, in terms of desire and satisfaction based on people's preferences, as the function of welfare (Sen, 1992). Welfare economics is the standard theoretical framework for assessing market based solutions in areas such as environmental economics (Hanley & Barbier, 2009) and transport economics (Button, 2003), and is the theoretical basis for the majority of economic evaluations applied in public policy by the UK government (HM Treasury, 2003). Welfare economics is primarily concerned with economic efficiency, meaning:

> *"to order social states on the basis of some minimal value judgements"* (Boadway & Bruce, 1984, p. 2)

There are four key principles upon which welfarism attempts to achieve economic efficiency (Hurley, 1998; Hurley, 2000):

The first principle is known as utilitarianism. Utilitarianism means that each individual in society is a rational agent. Under utilitarianism, each agent acts to maximise their utility or preferences to their optimum or highest possible level (Hurley, 1998).

The second principle of welfarism is individualism. This is where individuals themselves are thought to be the best judges of how to maximise their utility, with a laissez-faire approach from the state which permits utility maximisation by individuals (Hurley, 2000).

Principle number three is consequentialism. Consequentialism is where the outcome of choices made by individuals is the only consideration for assessing their goodness. The means to how the ends or outcomes are reached are deemed irrelevant (Hurley, 1998).

The final principle is welfarism itself. Welfarism can be defined in many different ways, but the principle tenet is concerned with the judgement that is made for allocating resources to improve human welfare. Within welfarism this judgement is based only on individual utility (Hurley, 1998).

The evaluation of individual utility within the welfare economic framework is mainly concerned with the notion of Pareto efficiency or Pareto optimality, named after the Italian economist Vilfredo Pareto (1848-1923). The theory behind Pareto optimality in welfare economics is that there is a socially optimum point at which efficiency is reached, which represents societal welfare whereby no change in individual utility level can be improved without making someone else's utility worse (Boadway & Bruce, 1984). Any change in utility which moves society closer to the point of Pareto optimality is referred to as a Pareto improvement (Morris et al., 2007). This objective is closely associated to the writings of utilitarian economist Léon Walras (1834-1910) and his theory on general economic equilibrium (Hunt, 2002).

While the Pareto principle allows for a judgement on welfare levels to be made where no one loses utility and there are only utility gainers, it does not help to make a judgement under a fixed healthcare budget as there will be both gains and losses, in terms of individual utility, from the impact of a particular policy. In reality, most policy interventions will have winners and losers, which the Pareto improvement rule does not take into consideration (Coast et al., 2008d). To counteract this problem welfare economists have proposed a compensation principle (Hicks, 1939; Kaldor, 1939), whereby a potential Pareto improvement is achieved if those who gain utility are able to hypothetically compensate the individuals who lose utility and still be better off themselves (Morris et al., 2007). In this case, the policy intervention should be implemented.

2.2.2 Extra-Welfarism

The application of the normative theoretical framework of welfarism to a healthcare setting is controversial because there are a number of principles in welfarism that conflict with the nature of healthcare. The principle underlying welfarism that has been most strongly challenged within health economics is the principle of utilitarianism, i.e. relying solely on utility information to judge individual wellbeing. While early applications of the extra-welfarist approach evolved out of the importance of valuing health from healthcare (Coast, 2004), more recently the theoretical critique of welfarism for use in healthcare has been drawn primarily from the critique of utility as a basis for assessing societal welfare by Amartya Sen (Sen, 1977). In his critique of welfare economics, Sen referred to capturing additional information beyond individual utility as extra-welfarist. From this critique, the term extra-welfarist has become associated with the health economics alternative to welfarism.

Brouwer and colleagues (2008) identified four ways in which extra-welfarism can be distinguished from welfare economic theory, as presented in Section 2.2.1 (Brouwer et al., 2008).

First, extra-welfarism permits the use of non-utility outcomes. Given that the focus on the healthcare sector is on improving health, Brouwer and colleagues (2008) argue that a sole focus on utility is too narrow for health analysis. The primary normative framework for extra-welfarism in health economics is mainly based on incorporating additional information beyond utility information into outcome measurement for healthcare provision (Culyer, 1989).

Second, extra-welfarism allows for the valuation of outcomes from those not directly affected by the outcome of interest. Within extra-welfarism, a number of different population groups could be considered relevant for valuing outcomes using this theory, not only the values (in terms of utility) of those directly affected within the welfarist tradition (Brouwer et al., 2008). Alternative values from the individual(s) under consideration can be appropriate within state provision of healthcare, for example, where the general population is funding the treatment of those who receive treatment, so it could be argued that they are stakeholders in the benefit obtained from such interventions and should be involved in the valuation of outcomes (Gold et al., 1996).

Thirdly, Brouwer et al. (2008) consider extra-welfarism to be different from welfarism because it allows the weight of outcomes to vary from individual preferences. For example different weights could be applied based upon a socio-demographic characteristic of the individuals receiving the intervention (Brouwer et al., 2008), or additional weight could be added if priority was advocated for a particular patient group (e.g. children).

Finally, extra-welfarism is different from welfarism because it permits interpersonal comparison in a number of dimensions of well-being (Brouwer et al., 2008). The primary difference in outcomes used within the extra-welfarism umbrella is that comparisons of wellness in terms of health are possible within this framework and comparisons between the health of different people can be made, unlike the comparability of utility scores (Brouwer et al., 2008).

While it has been argued that there are a number of differences between the extra-welfarist and welfarist frameworks, a number of similarities between the applications of the two theories remain. The objective within the extra-welfarist framework remains consequential in evaluation (i.e. maximisation), mirroring the same form of consequentialism as applied in welfarism. The only difference is what is maximised, with the maximisation of utility in welfarism replaced with the maximisation of health in extra-welfarism (Hurley, 1998). While the extra-welfarism framework argues for the multidimensionality of outcomes to be accounted within evaluation, the practical application of extra-welfarism focuses on a single dimension (Hurley, 1998). This is particularly true within the extra-welfarism theoretical framework applied within health economics presently, with the objective of the maximisation of health using health related outcomes as the primary objective of interest (Culyer, 1989).

2.3 MEASURING BENEFITS FOR ALLOCATING RESOURCES IN HEALTH ECONOMICS

The first section on outcome measurement in economic evaluation focuses on the Quality Adjusted Life Year (QALY). In particular, this section addresses what the QALY is, how it is calculated and the history of its use in health economic evaluations. The second part of this section focuses on outcomes that have been proposed as an alternative to the QALY in economic evaluations.

2.3.1 Quality Adjusted Life Years

The QALY is the recommended measure of health benefit by the UK health guidance body NICE to be calculated for economic evaluations (NICE, 2013). The use of a cost-effectiveness outcome for the UK healthcare system was originally recommended in the first guidance to manufacturers of new technologies by NICE (NICE, 2001). This recommendation by NICE remains for new healthcare technologies (NICE, 2004) and also for public health interventions (NICE, 2011). These recommendations have led to a significant increase in the use of the QALY within the UK, with their use increasing globally too (Neumann et al., 2009).

The QALY is a single outcome comprising a combination of two key components: health related quality of life (HRQoL) and length of life. Originally the idea of incorporating quality of life into economic outcomes was mooted 45 years ago (Klarman et al., 1968). In more recent times where QALYs have been identified as 'reference cases' by advisory bodies in the UK (NICE, 2004), the concept of incorporating quality as well as quantity of life into resource allocation decisions has been the primary outcome of economic evaluations.

The QALY as it was defined first in 1977 (Weinstein & Stason, 1977) has changed relatively little over time (Johnson, 2009). The QALY takes account of both quality of life in terms of health (quality or Q) and length of life (i.e. life years LY). The quality part of the QALY is measured on a scale with the common anchoring of full health anchored to one and health states equivalent to being dead anchored to zero (Drummond et al., 2005). The quality part of the QALY is collected over time and combined with time spent in health states to measure QALYs, where 1 QALY is equivalent to one year in full health. When applied to patient populations, the QALY seeks to find the additional health benefit of receiving a new treatment in comparison to an alternative by measuring the change in quality and quantity of life if a new treatment were introduced (Weinstein et al., 2009).

Three steps are required to value the quality part of the QALY: 1. what attributes of quality need to be valued; 2. how are these attributes to be valued; 3. who is to value them (Dolan et al., 2009). Each of these three issues are addressed below:

2.3.1.1 What attributes to value

To calculate what is to be valued in the QALY, a generic measure of health status is usually collected from patients. The recommended method by NICE for measuring quality for QALYs is the EuroQol (EQ-5D) (Brooks, 1996). The EQ-5D is a five item questionnaire of health status which assesses mobility, self-care, usual activities, pain/discomfort and anxiety/depression (Brooks, 1996). The dimensions on the EQ-5D were originally developed on three levels (i.e. no problems, some problems and a lot of problems on a given dimension). The EQ-5D has recently been expanded to a five level version, the EQ-5D-5L (Herdman et al., 2011) and is now recommended instead of the three level (EQ-5D-3L) version (NICE, 2013).

Other generic health status instruments to calculate the quality part of the QALY formula include the Short Form 6 dimension (SF-6D), the health utilities index Mark II (HUI2) or Mark III (HUI3) and the Assessment of Quality of Life (AQoL). The SF-6D is derived from the SF-36 or SF-12 generic health questionnaire. Dimensions on the SF-6D are physical functioning, role limitations, social functioning, pain, mental health and vitality, with four to six levels for each dimension (Brazier et al., 2002). The HUI3 consist of eight attributes which are vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain/discomfort. Each attribute has five to six levels ranging from normal to highly impaired/disabled (Feeny et al., 1995). The HUI2 and HUI3 are mainly applied in North America (Furlong et al., 2001). Four versions of the AQoL exist, ranging from the AQoL-4D with four dimensions (independent living, mental health, relationships and stress) assessed across twelve questions, to the AQoL-8D with eight dimensions (four dimensions of AQoL-4D plus happiness, coping, self worth and pain) assessed across thirty five questions (Richardson et al., 2011). The AQoL questionnaires are predominantly used in Australia (Hawthorne et al., 1999).

2.3.1.2 How are the attributes valued

Second, generic health status instruments need to be valued. NICE stipulates that the method for valuing between different health states must be choice based (NICE, 2013). This is the reason why rating scales of health such as the EuroQol Visual Analogue Scale (EQ-VAS), a scale of 0 (worst health state imaginable) to 100 (best health state imaginable) cannot be used to value health states, as respondents are not presented with a choice in the task. Preferences for health states are used to compare different interventions to represent a societal value of changes in health status (Gold et al., 1996).

For the EQ-5D-3L, the values associated with each of the 245 possible health states (3^5 or 243 health states and two additional health states for ‘unconscious’ and ‘dead’) were generated in the UK by Dolan from a representative sample of the general UK adult population (Dolan, 1997). These preferences were elicited using the time-trade off (TTO) technique developed by Torrance and colleagues to generate health preferences between quality and quantity of life (Torrance et al., 1972). The TTO method asks participants how much quantity of life they are willing to trade off in a worse state of full health (i.e. less than 1) to improve their quality of life to its optimum level of full health (Torrance et al., 1972).

Values for the SF-6D (Brazier et al., 2004a) and child-orienteted HUI2 (McCabe et al., 2005b) questionnaires are calculated by adopting the standard gamble (SG) approach, which is based on von Neumann and Morgenstern utility theory (Drummond et al., 2005). Cardinal preferences are measured by choosing between two options for a specific health state: one option is the current option or the ‘do nothing’ approach; the second option is a new treatment with probabilities attached to the likelihood of outcomes if this new treatment replaces current practice (Gafni, 1994). Other alternative valuation methods include discrete choice experiments, whereby respondents are asked to choose between alternative states, and thus deriving a latent (unobservable) utility function of their preferences (de Bekker-Grob et al., 2012).

Once a health status questionnaire has been completed to give a profile of an individual for a given condition, values are then assigned to the patient profile to generate an index score for that state of being (Morris et al., 2012). Index scores of individual health states can then be combined with the length of period a given individual spends within this health state to calculate the QALY. For example, an individual who scores an EQ-5D score of 0.5 and is in this health state in one year generates 0.5 QALY.

2.3.1.3 Who to value the attributes

Finally, there is the question of who should value health. As was already mentioned, the EQ-5D values were derived from the general population, which is the recommended approach by NICE (NICE, 2013). There is a debate between who should value health, with some arguing for the preferences of the general population in health outcomes ("decision utility"), while others believe that patients experiences ("experienced utility") should be used instead (Dolan & Kahneman, 2008).

Theoretical justification of the valuation by the general population is contributed mostly by a US panel of experts who suggested population valuations as the preferred approach due to the "veil of ignorance" of the general population. The general population would therefore maximise the aggregation of "utility" across the population lives within a given society (Garber et al., 1996). Additionally, there is the argument that the public should have a role within deciding what values are implemented within a publicly funded healthcare system, such as the NHS (Hadorn, 1991b).

There are a number of arguments against using general population values. The main alternative proposed is that patients with experience of the condition should be valuing health states rather than the hypothetical values the population perceive for the same conditions (Dolan & Kahneman, 2008). There are also questions over the theoretical grounding of the population value approach within welfare economics (Gandjour, 2010). However, there are a number of issues with using patients' values too. There are arguments that patients' values are affected by adapting to the condition over time (Menzel et al., 2002), as well as a "response shift" down from what the best health state imaginable was before the condition (Sprangers & Schwartz, 1999). These concerns are reflected in research which shows that patients record higher values (i.e. in a better health state) than the general population do for the same condition (Ubel et al., 2003; Mann et al., 2009).

2.3.2 Alternatives to the QALY

A number of alternatives to the QALY outcome measure have been proposed within the health economics literature. Welfarists argue that the QALY is not consistent with standard welfare economic theory and development of outcomes which have theoretical grounding in welfarism are more appropriate for assessing allocative efficiency (see Section 2.3.2.1). Alternatively, other critiques of the QALY outcome have come from within extra-welfarism, arguing that aspects of the QALY calculation can be improved with the focus remaining on health status. Each additional proposal will be briefly discussed in Section 2.3.2.2. However, since the primary interest of this chapter is to demonstrate and critique the standard format of the health economic evaluation methods used most often in practice, critiques of the alternative measures are discussed only briefly.

2.3.2.1 Willingness to Pay

A major reason for the original deviation of health economics evaluation frameworks and outcomes from welfare economics was a difficulty of measuring individual benefits in monetary terms. Within a health setting, putting a direct monetary valuation on life is difficult on a number of levels, namely the ethical issues involved with valuing life monetarily as well as the acceptance of monetary valuations of life within the health service (Garber et al., 1996).

However, this has not deterred attempts within health economics to develop methodologies more in line with standard welfare economic theory.

Given that there is no market price for healthcare in the UK, due to the public funding of the NHS, alternative methods are required to ascertain monetary values of healthcare within the welfarist approach. The main method of valuing individuals' utility (benefit) in this format is eliciting willingness to pay (WTP) values from individuals who would benefit from a given intervention. WTP methodology has normative grounding in the application of welfare economics of John Hicks (Hicks, 1939). Unlike the revealed preference approach to valuing benefits monetarily, where values are inferred from choices that individuals make in real world scenarios, WTP methodology is based on the stated preference approach, derived from survey or experimental responses (Morris et al., 2012). WTP for healthcare is usually captured in contingent valuation surveys, which vary from open-ended questions on an individual's willingness to pay for an intervention, to iterative bidding games, payment scales and closed ended questions (Frew, 2010). The aim of all of these approaches, however, is to measure the maximum amount that an individual is willing to pay for the introduction of a new programme. Research within WTP studies has tended to focus on developing a greater understanding of which methodology is most appropriate for capturing individual's WTP, with no overall consensus amongst welfare economists concerned, at present, as to the optimal method of measuring WTP (McIntosh et al., 2010).

2.3.2.2 Extra-welfarist QALY alternatives

A number of alternatives to the QALY have been suggested within the health economics literature. The most well-known of these is the Disability Adjusted Life Year (DALY), which has been the measure of choice for assessing the global burden of disease by the World Health Organisation (WHO) since the early 1990s (Murray & Lopez, 1996). The calculation of QALYs and DALYs are somewhat similar. However, the objective of maximising health within the QALY approach is substituted in the DALY approach with minimising disease burden by reducing DALYs lost (Murray & Lopez, 1996).

DALYs consist of four components (Fox-Rushby, 2002):

1. Life expectancy measured through years of life lost from what would be an expected average of life expectancy
2. Age values – greater weight is given to individuals with diseases between 20 and 40 years old as they are the most likely to have caring responsibility for others. Less weight is given to those under five and over ninety
3. Value of future time – DALYs are discounted at a rate of 3% per annum (this type of discounting is also included in QALYs)
4. Value of avoiding disability, which is the inverse of the QALY calculation measuring years lived with disability, with the aim of minimising such occurrences.

The DALY has been developed to assess population health within developing countries, which is easier to measure where information on HRQoL may not be easily accessible (Murray & Lopez, 1996). The DALY provides more information than focusing on mortality data alone (Morris et al., 2012).

Originally, the values associated with given health conditions measured by DALYs were derived from a person trade off (PTO) method from healthcare practitioners, as opposed to valuations from the general population through TTO or standard gamble for health status instruments used for QALYs. The PTO asks individuals how much of a particular outcome for disease X is worth compared to particular outcomes for disease Y (Nord, 1995). However, new values for DALYs have been generated from a general population sample across five countries (Salomon et al., 2012).

Originally, the PTO was developed for another alternative outcome measure to QALYs, the Saved Young Life Equivalent (SAVEs) (Nord, 1992). SAVEs equal to 1 is the value associated with saving the life of a young person which, in Nord's opinion, all other interventions should be compared against using PTO (Nord, 1995). While SAVEs allow for cost utility comparisons similar to QALYs, Nord (1992) argues that SAVEs are markedly different than QALYs by comparing all interventions against an intervention which is seen as the maximum benefit from healthcare (i.e. saving the life of a young person). However, SAVEs are not necessarily a replacement for the QALY, as decision-makers may still be interested in health gain from different interventions (Nord, 1992)

Finally, the healthy-years equivalent (HYE) is an outcome instrument considered by some as a theoretically superior measure to the QALY (Mehrez & Gafni, 1989). While HYEs also combines health and length of life within one measure, HYEs differ in two respects to the QALY (Drummond et al., 2005) :

1. HYEs measure preferences over all of the varying states a person would find themselves in during treatments rather than measuring preferences of each state on its own

2. HYEs use a two-stage standard gamble process taking account of utility for each health path and the healthy year gained, whilst QALYs are compiled of a one stage process of standard gamble or time trade-off

However, due to the extensive and more complex calculation of the HYEs (Hauber, 2009), like SAVEs, DALYs, and WTP measures, the HYE has failed to catch the decision-makers attention in the same way as the simpler QALY outcome, within the UK at least.

The majority of these non-QALY extra-welfarist outcomes have been proposed because of some of the underlying assumptions with the QALY. The shortcomings of the QALY will be re-visited in Section 2.7.

2.4 EVALUATION FRAMEWORKS FOR HEALTH

In this section, an account of the main economic evaluation types that have been employed to aid decision-making for healthcare provision are examined. While most attention is given to the recommended evaluation framework by NICE, the section starts with an introduction to the standard economic evaluation framework, cost-benefit analysis (Section 2.4.1). Other types of evaluation frameworks which have been implemented are also discussed in this section.

2.4.1 Cost-Benefit Analysis (CBA)

The main type of economic evaluation arising from the theoretical basis of welfare economics is cost-benefit analysis (CBA). The main aim of CBAs are to value the costs and benefits of different interventions/treatments against each other, usually in monetary terms (Drummond et al., 2005). This is the most clearly Paretian form of economic analysis in practice, where benefits are sought which deliver a societal increase in utility. It is the primary evaluation framework used by the UK in appraising and evaluating public policy projects outside of healthcare (HM Treasury, 2003).

The origins of CBA can be traced back to the first half of the nineteenth century when French economist and engineer, Jules Dupuit, enquired about how the use of toll roads would benefit the public at large (Boardman, 2006). Cost-benefit analysis plays a major role in aiding decision-making in areas concerning transport and other areas across the public sector, such as environment and education projects (Gafni, 2006). When it comes to the use of CBA in healthcare, however, it remains somewhat on the periphery in comparison to its use in other areas of the public sector.

Cost-benefit analysis can be defined as:

> *"a policy assessment method that quantifies in monetary terms the value of all consequences of a policy to all members of society"* (Boardman, 2006, p.2)

The aim of this type of analysis is to estimate monetary values for benefits, compare them with the monetary costs of providing a given project, for example the building of a new road. If the benefits outweigh the costs, then the project should go ahead. This is usually represented in terms of Net Monetary Benefit (NMB). If the monetary gains outweigh the cost of a proposal, then the project under review should be implemented on economic grounds (Morris et al., 2012). The CBA approach to evaluation relies on the assumptions of changes in individual utility as the key to aggregating social welfare and that individuals are the best judges of their own welfare (Drummond et al., 2005).

CBA focuses on allocative efficiency, that is, the overall impact of a project across the sector where resources are being allocated. This means that when CBA is applied within the health service, all health and non-health related cost and benefits are, in theory, accounted for within monetary outcomes. Allocative efficiency allows for comparison of welfare across multiple interventions for different population groups (Palmer & Torgerson, 1999). Practical examples of allocative efficiency studies linked with the CBA framework within healthcare include comparing helicopter ambulance services, heart operations and hip replacements (Olsen & Donaldson, 1998), and mental healthcare compared to cancer and elderly care (O'Shea et al., 2008). However, only one CBA in healthcare is known to have been completed in practice (Haefeli et al., 2008).

A major issue with the application of CBA within a healthcare setting is the monetary valuation on the benefits of health improvements to human life, thereby indirectly leading to a monetary value on a human life (Robinson, 1986). However, many economists believe it is the best way of evaluating outcomes because it is grounded within welfare economic theory. New methods of valuing improvements in health in monetary terms continue to be made to further develop this type of evaluation for healthcare (McIntosh et al., 2010).

2.4.2 Cost-Utility Analysis (CUA)

Since the establishment of the National Institute for Clinical Excellence (as it was originally known) within the UK in 1999, the use of economic evaluations, particularly alongside assessment of new technologies, has grown significantly. This is due partly to the recommendations for the conduct of economic evaluation for new pharmaceuticals within the earliest guidance by NICE (NICE, 2001). The role of NICE has expanded since it was founded, to advise on public health interventions by local government (NICE, 2011), social care and other areas of evidence in the NHS such as a Quality Outcomes Framework, an incentive scheme for GPs and an international branch of NICE, NICE International, too (NICE, 2013). Cost-utility analysis (CUA) is the preferred evaluation framework of NICE for baseline comparisons across the NHS (NICE, 2011; NICE, 2013).

CUA is a type of economic evaluation which focuses attention particularly on health related outcomes for healthcare treatments (Drummond et al., 2005). While CUA is also referred to as cost-effectiveness analysis (CEA) in the United States (Gold et al., 1996), CUA is differentiated from CEA in a number of ways (see Section 2.4.3 for more on CEA). CUA attempts to develop a framework which allows for a generic measure of health related utility (often referred to as preference based measures) from health related quality of life (HRQoL) instruments, which can be implemented and compared for priority setting decisions between interventions across a health service (Drummond et al., 2005).

CUA is the main evaluation framework of the extra-welfarism theory for healthcare as developed by Culyer (Culyer, 1989). Culyer believed that the objective of the extra-welfarism theory should be the output of health as the maximand (Culyer, 1989). While utility is referred to within the title of CUA, it is not utility as is commonly interpreted within welfare economics. Utilities within the CUA framework refer to the von Neumann-Morgenstern 'utility theory' as measures of cardinal utility which are interpersonally comparable (von Neumann & Morgenstern, 1953). Measures of HRQoL rely on preferences of individuals to value a generic health state in comparison to the anchors of full health and a state equivalent to being dead (Dolan et al., 1995). The index scores generated from HRQoL questionnaires are then combined with length of time, to form a QALY (or other outcome, see Section 2.3), which is used as the outcome of benefit from economic evaluation and provides the reference case outcome measure for NICE evaluations (NICE, 2013).

The CUA evaluation framework requires a consistent outcome measure to be applied across all interventions evaluated, so that decisions can be made that not only address technical efficiency between treatment options for the same health condition, but also allocative efficiency, so that funding can be justified in comparison with any other treatment across the health service (Palmer & Torgerson, 1999). This is of particular importance in the UK healthcare system which is mainly funded by the UK taxpayer and as decisions on the allocation of scarce resources need to ensure resources are appropriately allocated to different areas of the health service, so that taxpayers are getting value for money (Gerard, 1993).

2.4.3 Alternative Evaluation Frameworks

A number of alternative frameworks have also been used to evaluate health from an economics perspective when it has been felt that the appropriateness of CUA or CBA for a certain intervention has been challenged. The majority of these frameworks would be classified within the extra-welfarism framework primarily because their outcomes usually move away from utility as interpreted in the Paretian welfare economic sense of the word.

One framework which has already been mentioned in passing is cost-effectiveness analysis (CEA). What distinguishes CEA from CUA, is that outcomes within CEA are focused on a particular condition, rather than a generic measure applicable across many conditions (Drummond et al., 2005). The measurement of outcomes usually takes place in what are referred to "natural units" such as life years gained (Weinstein, 1990), or major outcomes averted (Roberts et al., 2007). While CEA can be a useful tool for a decision-maker with a budget for only one condition, at the macro level such outcomes are difficult to compare in natural units (e.g. comparing a rehabilitation service for drug addicts with hip replacements would be difficult as there is no common outcome to both interventions). CEA is therefore only useful in situations of technical efficiency within a health service, which is concerned with measuring the efficiency of interventions for a specific population group (Palmer & Torgerson, 1999). This is a more constrained measure of efficiency than allocative efficiency across a health service, which is in theory applied within CBA or CUA evaluations.[1]

Cost Minimisation Analysis (CMA) is sometimes used when CEA results produce similar natural unit outcomes and the focus switches to the cost side of the analysis (Drummond et al., 2005). However, such an analysis, which targets the reduction of costs only, has been widely discredited within the health economics literature due to the uncertainty of costs and effects for outcomes of different interventions (Briggs & O'Brien, 2001). Nonetheless, CMAs continue to be conducted and published, albeit on a small scale, even though the evaluation format continues to be questioned within the health economics literature (Dakin & Wordsworth, 2013).

A final evaluation framework used within health economics is known as cost-consequence analysis (CCA). CCA, like CEA, is concerned with technical efficiency. However, CCA is a disaggregated evaluation of all costs, resource use and outcomes in natural or generic units of the interventions under consideration. It is then left up to the decision-maker to weigh up for themselves how to value which intervention is best (Mauskopf et al., 1998). CCA has the

[1] While there is a debate as to whether CUAs are an extension from CEAs or a limited form of CBA (see Morris et al. 2007, p. 250), in terms of resource allocation within a health service, both evaluation practices can be distinguished from CEA, which can only be used as a technical efficient assessment for a given patient group.

advantage of all the necessary information given to the decision-maker, but again this framework would be difficult to use for allocating resources across a health service without a common metric (Coast, 2004).

An evaluation approach which is an extension of CCA called Multi Criteria Decision Analysis (MCDA), has also been suggested as a method for aiding decision-making within a health service (Devlin & Sussex, 2011; Thokala & Duenas, 2012). What makes MCDA different from CCA is that numerical values are attached to the multiple outcomes, with different outcomes (e.g. cost-effectiveness, severity, disease of the poor, age etc.) given weights of importance and aggregated to find the optimal intervention (Baltussen & Niessen, 2006). While the MCDA framework is in its infancy within health economics, it may offer a mechanism for allocating resources across a health service, which is beyond the reach of cost-consequence analysis.

2.5 MODELLING APPROACHES IN HEALTH ECONOMIC EVALUATIONS

To combine data on the costs and benefits of different interventions from multiple sources, decision models are frequently used to generate the required outcomes to aid decision-making. A model-based evaluation has been defined as:

> *"a formal quantified comparison of health technologies synthesising sources of evidence on costs and benefits, in order to identify the best option for decision makers to adopt"* (Brennan et al., 2006, p. 1296)

Economic modelling plays an important role in capturing changes in patient quality of life over time. Decision models are used as a method to simplify the complexities of reality across a multitude of academic disciplines, not least health economics. Two important aspects within evaluating health require the incorporation of economic models to capture changes in individual wellbeing (Buxton et al., 1997):

1. Where data which can be used to generate economic outcomes (i.e. QALYs) have not been collected prospectively within a clinical trial, models are used to combine the best available data to estimate such outcomes.

2. Where economic outcomes have been collected, due to resource constraints they may only be collected for patients over a short time period. Interventions which will have impacts over the individual's life cycle also need to be evaluated, so models allow for the extrapolation of data over time.

Moreover, decision models are also useful even when a clinical trial has been conducted as the trial in question may not compare all the relevant interventions (Sculpher et al., 2006). Additionally, relying on one clinical trial does not allow the incorporation of data from previously conducted related research from trials, meta-analysis and observational studies (Petrou & Gray, 2011).

Another key component as to why economic models are used is that they allow for a comprehensive assessment of uncertainty around the model parameters (Briggs et al., 2006). There are a number of types of uncertainty encountered within economic evaluation estimates.

The main focus of decision analysis is to address parameter uncertainty, i.e. the accuracy of the individual data inputs (e.g. HRQoL of individual at a certain time point), and decision uncertainty, the likelihood that the decision made from the data within the model is the correct one (Briggs et al., 2006). Each of these types of uncertainty can be addressed through what is known as sensitivity analysis, which are statistical techniques which assess the sensitivity of results due to uncertainty around the parameter inputs (Briggs et al., 1994). The uncertainty around the final decision from the model outputs can be ascertained by changing model parameters simultaneously through the use of probabilistic sensitivity analysis (PSA) (Claxton et al., 2005).

The underlying need for model-based economic evaluations is driven by the requirements of considerations necessary within economic appraisals for decision-making bodies like NICE in the UK, the Pharmaceutical Benefits Advisory Committee in Australia and the Canadian Drug Expert Advisory Committee (Fischer, 2012). For example, within the UK, the most recent edition of the methods for technology appraisal by NICE advises that models should be used when all relevant data (i.e. comparators of interventions, healthcare costs and benefits over technology lifetime etc.) are not contained within a single trial, where patients in trials are not representative of real-world patients, where HRQoL and survival data are not collected within trials and also where treatment switching of patients occurs in trials (NICE, 2013, p. 44-45).

There are a number of different types of models which have been employed within health economics that carry different assumptions when calculating outcomes. The simplest of these model types are known as cohort models, where the average expected cost and effects across the population are calculated at the cohort, rather than individual level (Briggs et al., 2006). The simplest type of decision analysis cohort model is a decision tree, consisting of decision strategies from which outcome nodes (e.g. costs, QALYs) are calculated for each decision strategy. A decision node is then used to compare the outcome nodes from each decision strategy from a pre-defined decision rule by the decision-maker (Stahl, 2008). However, the simplicity of the tree does not allow recursion, which can make decision trees unwieldy (Barton et al., 2004b; Brennan et al., 2006). Therefore, Markov models are a simpler method to capture cyclical events within a decision strategy as an alternative cohort model to decision trees. In Markov models, each individual within a decision strategy is assigned to a finite number of states (e.g. well, sick, dead) over a fixed time period, with transition probabilities assigned between states, representing the likelihood of patients moving to different states or staying in the same state (Barton et al., 2004b).

While Markov models readily allow calculations of recurring states, Markov models have no memory of the states individuals were in previously. Therefore, this Markov assumption of homogeneity of individuals within the same state can be overcome through more detailed individual sampling models using Monte Carlo simulation (Barton et al., 2004b). Within these models, individuals are processed one at a time, rather than in cohorts like decision trees or Markov models, allowing for the patient pathways to be reflected within the decision strategy (Briggs et al., 2006). Also, different models for infectious diseases have been implemented to reflect the dynamic nature of the progression of these types of diseases (Brennan et al., 2006; Roberts et al., 2007).

While there are benefits to the more complex individual sampling models and dynamic models, such models are more demanding on the amount of data required, the computational burden and the assessment of uncertainty for such models is much more demanding as a result (Briggs et al., 2006). Therefore, the choice of model for a given economic evaluation needs to be representative of the complexity of the condition, which should aid in deciding how simple or complex the representation of decision strategies are required to make that decision (Barton et al., 2004a).

Outcomes within economic models are employed to capture the changes in health benefits throughout the period at which an intervention will benefit the patient population. While outcomes are preferably based on primary data, it is often the case that the quality part for calculating QALYs is not collected directly for a particular intervention. Therefore, measures which are not preference based are sometimes used to predict values within health related utility instruments such as the EQ-5D (Lin et al., 2013). This process of prediction is referred to as mapping or cross-walking within the literature (Brazier et al., 2010). With the growing recognition of the need for economic evaluation, this practice of mapping between instruments appears to be on a decreasing trend, as new health technology assessments are increasingly collecting HRQoL directly (Tosh et al., 2011). However, such procedures are often necessary for attempting to measure economic outcomes when no other data are available.

2.6 DECISION RULES IN HEALTH ECONOMIC EVALUATIONS

There are a number of decision rules which, in theory, could be used to aid healthcare decision-making. Decision rules are generally based on aiding decision-making as to whether new interventions are worth the additional cost burden to the body in question (e.g. hospital, regional or national provision). For NICE, QALY scores are aggregated for the population under consideration with the costs and benefits combined by calculating a cost effectiveness ratio or cost per QALY. To compare differences between costs and effects for competing interventions, the incremental cost-effectiveness ratio (ICER) is applied to measure the cost per additional QALY gained for the more expensive and/or effective treatments (Drummond et al., 2005). The ICER for a given treatment is then compared with a shadow price of the budget of interest. This is known as the threshold ICER rule (Birch & Gafni, 2006). For new interventions to be recommended by NICE, the willingness to pay for an additional QALY must fall within or below the threshold range of £20,000-£30,000 (NICE, 2013). However, in exceptional circumstances, the willingness to pay for QALY gains is sometimes raised above the £30,000 threshold (NICE, 2013). A recent study has suggested that over four fifths (81%) of NICE decisions can be predicted by the prevailing threshold ICER rule of less than £30,000 per QALY gain (Dakin et al., 2013a).

Another rule with the ICER is the league table rule, where interventions with the lowest ICERs are recommended until no more resources are available (Birch & Gafni, 2006). This type of analysis is better known as cost-effectiveness league tables, which compare outcomes for different patient groups in terms of costs per QALYs and have been applied within the UK previously (Williams, 1985; Maynard, 1991). However, the league tables approach came under heavy scrutiny (Drummond et al., 1993; Gerard & Mooney, 1993), which has led to the

ICER threshold rule as the dominant method for comparing interventions in health economics currently. Members of the DALY team have also attempted to implement the league table decision rule through what they refer to as generalised cost-effectiveness analysis (Tan-Torres Edejer et al., 2003).

One alternative advocated is an incremental approach whereby only improvements over prior treatments could be recommended for treatment (Sendi et al., 2002). This incremental approach has recently been expanded so that the probability of 'bad outcomes' can be quantified through a loss function to aid resource allocation, by incorporating opportunity costs into the function so that the size of these outcomes (good or bad) are accounted for in a transparent manner (Gafni et al., 2013).

2.7 CRITIQUING CURRENT HEALTH ECONOMICS PRACTICE

While Section 2.5 showed the uncertainty in calculating accurate outcomes when modelling in economic evaluation, another concern relates to the three requirements for valuing outcomes (see Section 2.3.1) (Dolan et al., 2009). As was alluded to throughout the chapter, the QALY outcome, based on the extra-welfarism framework, is the primary economic outcome measure of health interventions. Both the theoretical grounding of the extra-welfarism framework and the evaluative space of the HRQoL within QALYs have been challenged from a number of different angles. In this section, the aim is to identify some of the main critiques of the QALY extra-welfarist perspective within health economics, with a particular focus on how alternative frameworks and outcomes have been used to address the issues identified. The section closes with a considered view as to why the QALY has managed to maintain the position of primary economic outcome for measuring changes in health, despite these critiques.

2.7.1 Critiquing the QALY: Evaluation Space

The conventional QALY application in CUA is justified from the extra-welfarist framework developed by Culyer (Culyer, 1989). Culyer's interpretation of extra-welfarism draws from a number of sources, including the capability theory of Amartya Sen and the need for healthcare provision to improve health (Culyer, 1989). This is the main rationale for justifying more than utility. However, in practice the conventional QALY is concerned with HRQoL alone. Preferences for different health states are obtained through the valuation exercises of TTO or SG for generic health instruments. By presenting the population with health states, the general population is presented limited health scenarios with generic HRQoL valuation exercises. This could lead to the so-called "focusing illusion" and may not fully capture the impact of the condition (Ubel et al., 2003).

An additional issue with the focus on health only is the generalisability of QALYs beyond the health service to compare the benefits to society with other public interventions such as education, justice and transport. While there has been some interest in adopting the QALY measure within environmental economics (Hammitt, 2002; Chokshi & Farley, 2012) and crime (Dolan et al., 2005a), a health QALY across the whole public service is not feasible, nor sensible given the limited direct health benefits some important public interventions may have. The consensus within the health economics community is increasingly that the QALY is a measure of health and not well-being more generally (Drummond et al., 2009), therefore limiting the role of the conventional QALY to healthcare only.

A primary concern within the evaluative space of QALYs is the understanding of what health means. Traditionally the instruments which are used to calculate the quality part of the QALY are anchored on a scale of being dead (0) to full health (1) (Drummond et al., 2005). Much of the ethical debate focuses on the lower end of the scale, whether in fact there are health states worse than dead (Rawles, 1989). Another concern is the determination of full health, as the full health for an athlete who is able to complete a marathon is likely to be greater than the “full health” envisaged by the majority of the population when completing valuation tasks used to calculate full health for QALYs, such as the EQ-5D (Sullivan, 2011).

An argument could be made that the welfarist approach is more generalisable across the whole public provision of services than the extra-welfarist QALY approach, as resource allocation decisions are not limited to the health benefits in WTP studies (Hammitt, 2002). However, for this to be unequivocally true, then the QALY would be the sole interpretation of an extra-welfarism framework. As Brouwer and colleagues (2008) show, the conventional QALY approach is just one interpretation of an extra-welfarist methodology. The capability approach, for example, does not exclude utility or life satisfaction from its calculation of a capability set (Sen, 1985), but does have a focus on more than one dimension of space to evaluate societal welfare. The focus on a uni-dimensional space is a critique that can be levelled equally with the QALY extra-welfarist framework as well as that of the WTP welfarist framework (Hurley, 1998).

While in some respects, the focus on maximising health through the health service makes intuitive sense, it is equally arguable that such a focus is not extra-welfarist per se, in the sense that the objective is not welfarist plus something additional. This has led to some health economists referring to the approach as non-welfarist (Birch & Donaldson, 2003; Morris et al., 2007).

2.7.2 Critiquing the QALY: Underlying Assumptions

Another key criticism of the QALY as the basis for aiding decision-making focuses on what the overall objective of healthcare provision should be. The objective of maximising health is embedded within the QALY extra-welfarist approach (Culyer, 1989; Garber et al., 1996). The criterion that a QALY is a QALY is a QALY, irrespective of who are the QALY gainers has caused much of the criticism with the current health economics approach. Challenges have been made as to whether QALY maximisation reflects societal preferences for health maximisation as QALYs infer (Dolan et al., 2005b). There have been many claims on the basis of differing arguments based on ethics, equity and equality which have challenged the QALY assumption that all QALYs are equal and the objective of healthcare is to maximise QALYs.

Equity is an issue because QALY gains are considered equivalent irrespective of age, prior health state or severity of illness (Nord, 1999). Three notable challenges to this basis have been made within the health economics literature, all of which focus on prioritising those in their earlier stages of life. The SAVEs outcome, for instance, was the first example of attempting to incorporate priority of saving the life of a young person as the optimal objective as to which every other health service intervention should be compared (Nord, 1992).

Issues of incorporating age is highlighted by the fair innings approach, in which those who have lived a relatively long lifetime (e.g. 70 years) are given lower priority for health interventions compared with those who had failed to reach their fair innings (Williams, 1997). Some argue that the QALYs indirectly account for this type of argument given that older people are likely to gain fewer QALYs than younger people, as their potential for more QALYs is less (Harris, 1987) makes the role of incorporating this kind of equity argument less compelling.

Incorporating age weights is an issue captured within the DALY outcome, where more weight is given to individuals who are likely to have a caring responsibility for younger and older adults, between the ages of twenty and forty (Fox-Rushby, 2002).

Severity of illness is also seen as a primary concern when measuring the benefits from an intervention. Erik Nord and colleagues have been prominent advocates of this approach. They argue that a fairer way to set resource allocation decisions is by valuing the avoidance of severe conditions, rather than the benefit of treatment from such conditions (Nord et al., 1999; Nord, 1999; Nord et al., 2010). In certain aspects, there are some similarities between the arguments by Nord and colleagues and the rule of rescue, which states that those whose lives are at risk should be prioritised above all others (McKie & Richardson, 2003). A call for a rule of rescue was primarily driven by QALY calculations which appeared to favour minor treatments over life saving action (Hadorn, 1991a).

There are many other aspects where the QALY maximisation principle has been the target of criticism. For instance, special treatment has been proposed for orphan drugs and rare conditions and these are currently more likely to be funded than other treatments within the UK (McCabe et al., 2005a). However, this would appear to be in conflict with those that propose the societal distribution of health which aims to reach as many people as possible (Dolan et al., 2005b). The role of the individual in their own health state also appears to show varying instances of how people would prioritise against those who are responsible for their health state (Dolan & Tsuchiya, 2009). End of life care (Round, 2012), social care (Al-Janabi et al., 2011) and process of care (Brennan & Dixon, 2013) are other areas where the QALY health maximisation objective has been questioned because healthcare provision may not be primarily focused on the maximisation of health over time.

2.7.3 Critiquing the QALY: Decision Rule

Both ICER decision rules discussed in Section 2.6 (i.e. threshold rule and league table rule), which are used to show efficient allocations of QALYs, have come under scrutiny in the health economics literature. The ICER threshold rule relies on divisibility and constant return to scale, which in practice would require an infinite supply of resources and a constant marginal opportunity cost (Sendi et al., 2002). This does not reflect the reality for decision-makers faced with scarce resources for competing interventions (Birch & Gafni, 2003). The league table approach requires perfect information on all interventions to allocate resources efficiently, which is rarely available within a national healthcare system for all interventions (Hutubessy et al., 2003) and there are difficulties of accounting for uncertainty within this framework (Evans et al., 2006). Similarly, mathematical integer programming has been suggested as another alternative (Birch & Donaldson, 1987; Stinnett & Paltiel, 1996), but this approach also requires perfect information on the costs and benefits of all interventions for efficient resource allocation. While the incremental approach has been suggested as an alternative (Sendi et al., 2002), this has rarely been applied in practice and has been criticised because of its inability to account for all potential efficiency gains (Lord et al., 2004). More recent research may make the incremental approach more practical for decision-makers (Gafni et al., 2013).

2.7.4 The Durability of QALYs

Notwithstanding the alternative outcomes and opposition to the QALY, the QALY remains the primary health economic outcome measure within the UK, through NICE, and also in the Canadian (CADTH, 2006) and Australian (Pharmaceutical Benefits Advisory Committee, 2008) health systems for assessing new technologies as well.

There are a number of examples of countries which have rejected the use of QALYs for use in decision-making in health. This includes Germany which has explicitly rejected the use of QALYs in favour of an efficiency frontier approach for assessing new pharmaceuticals (Caro et al., 2010) and France which also does not use QALYs (Holmes, 2013). The role of the QALY in cost-effectiveness league tables, which ranked health treatments against one another in terms of costs per QALY (Maynard, 1991), caused much controversy in decision-making in the United States (Drummond et al., 1993). However, there are signs of a relaxation of opposition to the QALY in the United States to some extent (Neumann & Greenberg, 2009).

Even within some countries where a sole reliance on QALYs has been rejected, they have continued to play a role in the evidence base for making decisions. There are two European examples of this. In the Netherlands, QALYs are used in an approach known as proportional shortfall, which combines the maximisation of QALYs with the equity argument by Williams for a fair innings (Stolk et al., 2004; van de Wetering et al., 2013). In Norway, alongside cost-per-QALY calculations are supplemented by a scale of severity of the condition in the final decision (Nord, 2012). So, even when the QALY decision rule is objected to, there appears to be a willingness to continue to use the outcome.

2.8 CONCLUDING COMMENTS

This chapter has highlighted the main methods of measuring outcomes for assessing health benefits which are applied by health economists in economic evaluations. The extra-welfarist framework using the QALY is elaborated in detail. While there are challenges to the evaluative space and decision rules from which the conventional QALY is applied, the QALY as a measure of health has withstood many of the challenges and remains the primary method for measuring the benefit associated with health interventions in economic evaluations. The QALY as currently formulated, however, is a measure of health and not a broader measure of well-being (Drummond et al., 2009). Excluding the application of WTP in health economics (Frew, 2010), no other outcome measures discussed here have attempted to measure anything other than health within the outcome measure. A broader benefit outcome measure of well-being may address some of the challenges to the QALY methodology in its current form (as outlined in Section 2.7), and it is to that area, with a focus on capability measures, that the thesis now turns.

CHAPTER 3. CAPABILITY, ECONOMICS AND HEALTH

3.1 INTRODUCTION

In the previous chapter, the current theory, evaluation frameworks and outcomes of health economic evaluations were explained in detail. It was shown that there are many challenges in determining the appropriate outcome measure to use and how it should be implemented to aid decision-making across a health service. In this chapter, a possible alternative theoretical base for conducting evaluations for aiding decision making in health is explored.

In Section 3.2 the capability approach proposed by Amartya Sen, Nobel Laureate in Economics for his contribution to welfare economics in 1998, is introduced. Through his many normative works (Sen, 1985; Sen, 1992; Sen, 1993; Sen, 2009) on the problems of using welfare economics as the theoretical basis for economic evaluation, Sen provides an alternative in the capability approach. This is explored in detail in that section. Also considered in Section 3.2 is a comparison of the capability approach and the extra-welfarist basis for health economic evaluations. As extra-welfarism differs notably from standard welfare economic theory (Brouwer et al., 2008) and since Sen's critique is primarily of welfare economic theory, a comparison between the capability approach and extra-welfarism is required to review whether Sen's criticisms of welfarism are also applicable to extra-welfarism as currently practiced within health economics.

In Section 3.3 the focus moves on to those within the capability approach who have specifically conceptualised the approach for health. Both the works of Ruger (2010a) and

Venkatapuram (2011) are discussed in detail here, and critiques of their conceptions of the capability approach to health for practical application within an evaluation format are provided.

An analysis of previous attempts to incorporate the capability approach within health economics is presented in Section 3.4. The focus is directed towards two notable attempts to re-interpret the QALY as an outcome measure that is compatible with the capability approach. These attempts, as well as other suggestions within the health economics literature, are explored here.

In Section 3.5 the focus is on the instruments that have been developed which have attempted to capture capability for use in aiding decision-making and healthcare resource allocation decisions. The three most developed questionnaires for health and social care that have links with the capability approach are of central importance here. One method relies on a capability list to design the appropriate questionnaire. Another interpretation has drawn on the capabilities that are of most relevance to the patient population at hand. The final method involves developing capability questionnaires for social care. Questionnaires related to each approach are explained in detail.

Finally, in Section 3.6, the chapter closes with a summary of the previous sections and areas which require further research so that a capability perspective can be adopted within an appropriate framework to aid decision-making.

3.2 WHAT IS THE CAPABILITY APPROACH?

In this section, the objective is to highlight some of the key arguments typically used against the current approaches to health economic evaluation (i.e. welfarism & extra welfarism) based on the beliefs of an alternative theory as a normative basis for evaluation, which is referred to as the capability approach. The basic concepts and ideals upon which the capability approach is founded are explored in detail in Section 3.2.1. Section 3.2.2 presents an investigation of the differences between the capability approach to evaluation and the current health economic evaluation techniques which are standard practice at present.

3.2.1 Theoretical Argument against Welfare Economics: Amartya Sen and the Capability Approach

Amartya Sen is the philosophical and theoretical inspiration for the capability approach. Throughout his career in economics, he has been a fierce critic of the utilitarianism agenda that has dominated economic policy since the end of World War II (Sen, 1979). From his early mathematical formulations on social choice theory (Sen, 1970) to his most recent writings focusing on social justice (Sen, 2009) a dominating theme of his work has been to expose the limitations of the Pareto/welfare economic approach to evaluation. Since 1979, when he first asked the question “Equality of What?” (Sen, 1980), the capability approach has been constructed and formulated into a normative philosophical basis (Sen, 1985; Sen, 1992; Sen, 1993). Today it provides an alternative way of analysing important issues related to public policy, and particularly relevant to this thesis, healthcare.

The capability approach has been defined as follows:

> *"The capability approach is a broad normative framework for the evaluation and assessment of individual well-being and social arrangements, the design of policies, and proposals about social change in society."* (Robeyns, 2005b, p. 94)

3.2.1.1 A Primer in Capability Terminology

There are three key concepts which dominate the capability literature that need to be defined before exploring the approach any further. Whilst these three definitions are related, in the sense that they all focus on individual well-being, they are important concepts independently.

The three central definitions within the capability approach are:

- Functionings:
  "represent parts of the state of a person – in particular the various things that he or she manages to do or be in leading a life" (Sen 1993, p. 31)
- Capability:
  "the various combinations of functionings (beings and doings) that the person can achieve. Capability is, thus, a set of vectors of functionings, reflecting the person's freedom to lead one type of life or another…to choose from possible livings"
  (Sen 1992, p.40)
- Agency:
  "the goals that a person has reason to adopt, which can inter alia include goals other than the advancement of his or her own well-being" (Sen 1993, p. 35)

The distinction between functionings and capabilities is crucial to understanding how individual welfare through the capability approach is assessed. Examples of functionings range from basic achievements in life such as being well-nourished or avoiding premature mortality, to achievements which vary across different cultures, such as having self-respect (Sen, 1992).

The capability an individual has~~,~~ relates to his or her ability to achieve valuable functionings in his or her life. Functionings for individuals can also include income or utility (in terms of desire fulfilment), but Sen, and others, believe that focusing on one of these issues alone is inadequate for assessing individual or societal welfare (Sen 1985). Instead, Sen states the focus should be on the capability to achieve various valuable functionings. This has been termed as the "capability criterion" when evaluations are made using Sen's theoretical basis (Gasper, 2007). The focus of analysis should not be on well-being as such, but rather on the opportunity to achieve such well-being (Sen, 1985).

To understand the capability approach by the definitions of functionings and capabilities alone, however, does not give a complete picture of the theoretical basis of the approach. What is key and unique within this theory is the role of "agency"[2]. Agency represents the opportunity to achieve well-being, but because an individual, as an agent of their values, can

[2] The numerous meanings of words like "agency" and "capability" are of considerable angst to Sen in his writings. Indeed, Sen (1993, p. 30) wishes to have phrased terms, in particular capability differently, which may have led to less confusion of his theory. "Agency" as defined by Sen is **not** the same as agency usually defined in health economics, which refers to the agency relationship, for example, when a doctor makes decisions on behalf of the patient, acting in the patients best interest (Mooney & Ryan, 1993).

have reasons to value goals which may harm their own well-being, agency can increase while the functionings or capabilities of the individual may decrease.

Sen (1992) states:

> *"A person as an agent need not be guided only by her own well-being."* (Sen 1992, p. 56)

The pursuit of agency expansion, which can damage individual well-being, does not refer to irresponsible acts of social behaviour which an individual may have reason to value (e.g. being drunk and disorderly, consuming illegal drugs etc.). Alternatively, agency in Sen's approach references issues where capabilities, not necessarily one's own, may be limited and where one feels so strongly that action must be taken, be it protesting against political oppression or sacrificing one's career in an advanced economy to help others in the developing part of the world (Sen, 1992).

3.2.1.2 Evaluation Base of the Capability Approach

Now that the three underlying concepts (capabilities, functioning and agency) have been introduced, it is possible to look more closely at the detail behind the key facets of the capability approach. Sen's capability approach focuses on how the three main foci of functionings, capability and agency, affect an individual's well-being, not just in terms of specific functionings, e.g. income, but in a multi-dimensional fashion that influences all relevant spaces for assessing welfare. A capability evaluation, thus, focuses on an individual's capability set, i.e.:

> *"capability is a **set** of such points (representing the alternative functioning n-tuples from which one n-tuple can be chosen)"* (Sen 1993, p. 38).

There are two key distinctions within a capability set: (1) "the promotion of the person's well-being" versus "the pursuit of the person's overall agency goals" and (2) "achievement" versus "freedom to achieve" (Sen, 1993). The first distinction concerns where the pursuit of a person's agency goals may conflict with the promotion of one's well-being because the person feels strongly enough to risk their well-being in pursuit of their agency goals. The second distinction is in relation to what is at the heart of the evaluation process: what a person has achieved versus what a person has the freedom or capability to achieve. "Achievement" in this sense refers to the functionings which affect well-being that have been realised by the individual, whilst the freedom to achieve is related to the choices available (and possible limits to choice) which have an impact on an individual's functioning achievement.

Following on from these two distinctions within the capability set, Sen (1992) pin-points four options for evaluation:

- well-being achievement;
- agency achievement;
- well-being freedom;
- agency freedom.

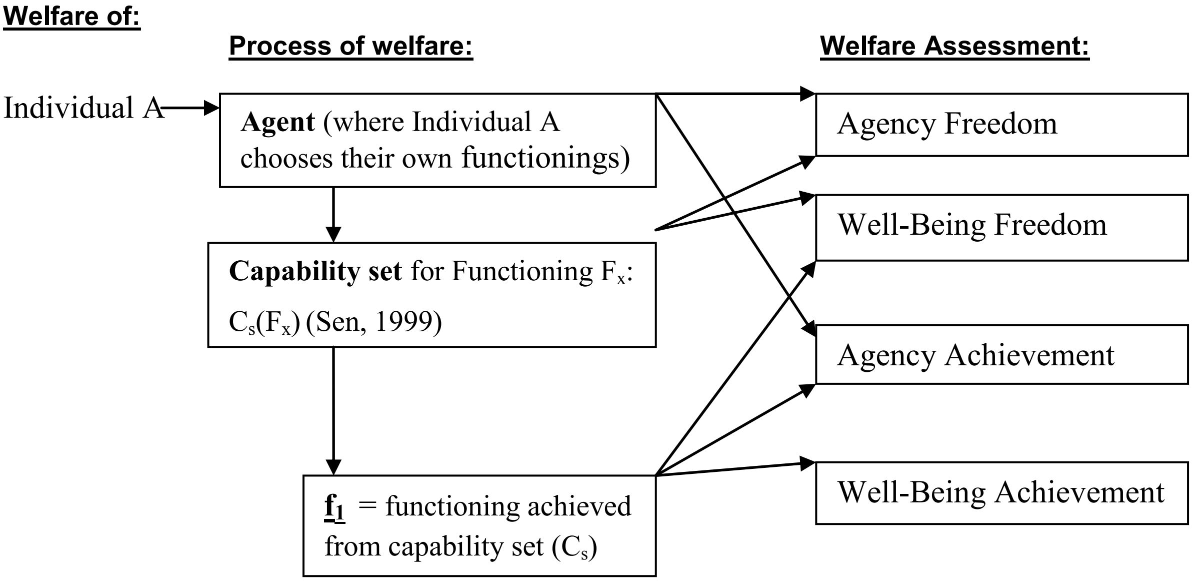
The relationships between well-being and agency, and achievement and freedom to achieve, are the key dynamic interactions that influence how the capability set is measured, which can be done under the four different headings concerning individual "advantage" outlined above. This approach allows for the possibility to measure how the achievements were reached and not just the achievements themselves, which overcomes one of the main critiques of current approaches within welfare economics.

Even though Sen refers to his theory as "capability", what he is really trying to measure is the "freedom" or "advantage" an individual has in life. When Sen refers to freedom, it is not a generic statement that can be applied differently in varying scenarios but relates specifically to what he defines as "effective freedom":

> *"If the levers of control were used by those in charge of them to, say, promote epidemics, rather than to eliminate them, our 'freedom to control' would not be changed (i.e. would continue to be absent) but our 'effective freedom' (in particular, the freedom to live the types of lives we would choose to lead) would be severely compromised"* (Sen, 1992, p.66)

Figure 2 summarises the capability approach in relation to the ways that "effective freedom" can be assessed. The capability approach, in theory, allows for factors generally not considered when measuring individual well-being, such as the agency of a person, and the freedom to choose. However, assessments rarely consider all four forms of welfare assessment simultaneously. For example, capability researchers have argued that in many situations a focus on functioning or well-being achievement is of primary importance for particular analysis (Robeyns, 2005b; Robeyns, 2006).

Figure 2 Options for assessing societal welfare in the capability approach



3.2.1.3 Early Applications of the Capability Approach

Since the capability approach allows for a broader intake of information which has an impact on individual well-being, it is not surprising that the approach has been applied within a wide variety of disciplines. One of the first, and most recognisable applications of the approach was developed by Sen and a Pakistani economist, Mahbub ul Haq, who established the Human Development Reports (HDR) for the United Nations Development Programme (UNDP) in 1990 (ul Haq, 1990). The Human Development Index (HDI), now described annually within these reports, compiles measures relating to health (life expectancy at birth), income (Gross Domestic Product or GDP per capita) and education (combination of adult literacy rate and school enrolment), rather than focusing on a single dimension for measuring a country's progress (e.g. GDP) to better understand *"the real wealth of nations"* (ul Haq, 1990). At the time of the first reporting of the HDI, this represented a significant departure in the way a country's progress could be assessed year on year.

The HDI has shown that a uni-dimensional measure for development may not capture the whole picture of development in terms of societal progress. One example of the frailty in analysing national income data alone is provided by Saudi Arabia, which scores very well in terms of GDP, but when the other issues of measurement in the Human Development Reports are included (including adult literacy rates, child mortality rates), the society is not so well-off as it might appear if the focus was purely on economic progress (Alkire & Deneulin, 2009).

3.2.1.4 Critiques of Applying the Capability Approach Quantitatively

Sen's philosophical theory is complex and is underspecified for practical policy evaluation, so there are a number of potential conceptual difficulties that arise for researchers in practice. These difficulties are highlighted in detail by Robeyns (Robeyns, 2005b).

3.2.1.4.1 Functionings or capabilities?

Firstly, there is a choice about whether the analysis should focus on individual functionings or capabilities. Most of the empirical research within the capability approach, particularly in the early applications, has focused exclusively on functioning achievement (Kuklys & Robeyns, 2005). This has been largely due to the difficulty with measuring capability (Krishnakumar & Ballon, 2008) and agency (Alkire, 2009) using data routinely collected for other purposes. More theoretical arguments against measuring capabilities come in the form of a welfare economics perspective with regard to the expansion of the capability set. The expansion of the capability set alone may not subsequently enhance the person's functioning well-being, and can in fact decrease functioning well-being with the number of choices a person may now have to make over functionings which may have a very small impact on a person's life

(Cohen, 1993). However, Sen was well aware that this misinterpretation could surface, and he tried to counteract this problem by excluding the word choice from most of the literature around the capability approach. Instead he relied on the use of the word freedom more often than not and hoped that this would clear up any misconceptions with regards to choice:

> *"Actually the capability approach recognizes that the goal is not to expand the number of choices – it is to expand the quality of human life"* (Alkire & Deneulin, 2009, p. 34)

3.2.1.4.2 Which capabilities/functionings to measure?

The second issue concerns the choice about which capabilities or functionings to focus on within an evaluation. The breadth of the capability approach is a strength in that it allows a vast array of information to be captured, but it is also regarded as one of the main weaknesses of the framework (Sugden, 1993). The lack of identified specific capabilities that are inherent when applying the approach and that different capabilities may be of more priority in different scenarios was a distinctive criticism of Sen's theory in practice (Robeyns, 2005b). Table 1 gives three examples of capability lists developed within the capability approach, including a list of "instrumental freedoms" from work by Sen (Sen, 1999).

Martha Nussbaum (Nussbaum, 2000; Nussbaum, 2011), a leading scholar and philosopher within the capability approach, has been one of the most notable critics of this problem with the approach and consequently she has compiled a list of ten 'central human capabilities' (Nussbaum, 2000, pp. 78-80) (see Table 1). Nussbaum's list aims to be comprehensive in including all capabilities that matter for human life. However, others have argued that a more deliberate and procedural process to reflect important capabilities in different contexts (Alkire, 2002; Robeyns, 2005b). An example of one of these context specific lists was developed by Robeyns to assess gender inequality in developed societies (Robeyns, 2003) (see Table 1).

Whilst aiming to be comprehensive, Nussbaum's list of ten central human capabilities is not appropriate for application in all areas of evaluation, particularly in context specific evaluations (e.g. "care for other species" would not generally be seen as an appropriate criterion by which to evaluate healthcare interventions). Indeed, Sen argues against the use of a pre-determined list as too constraining (Sen, 2004). Robeyns' (2005b) advice is to be as explicit as possible about the choice included in each given scenario, as well as providing a clear rationale for that choice.

Table 1 Examples of Capability Lists

Developer	**Sen (1999)**	**Nussbaum (2000)**	**Robeyns (2003)**
Description	Instrumental freedoms	"central human capabilities" for a good life	Assessing gender inequality in developed nations
Capability:	• Political freedom • Economic facilities • Social opportunities • Transparency guarantees • Protective security	• Life • Bodily health • Bodily integrity • Senses, imagination and thought • Emotions • Practical reason • Affiliation • Other species • Play • Control over one's environment	• Life and physical health • Mental well-being • Bodily integrity and safety • Social relations • Political empowerment • Education and knowledge • Domestic work and non market care • Paid work and other projects • Shelter and environment • Mobility • Leisure activities • Time autonomy • Respect • Religion

3.2.1.4.3 The individual as the focus of analysis

The third issue of concern relates to the interpretation of individualism in the capability approach and the effect on the aggregation of capabilities across a population. A distinction is made as to what type of individualism is important in the capability approach by Alkire & Deneulin (2009):

> *"The capability approach thus does not defend methodological or ontological individualism. But even if we are highly interested in groups, the capability approach, as initially framed by Sen, takes the normative position of 'ethical individualism' – the view that what ultimately matters is what happens to every single individual in a society."* (Alkire & Deneulin, 2009, p.35)

While this specific kind of ethical individualism takes into account societal structures and constraints by theoretically distinguishing functionings from capabilities (Robeyns, 2005b, p.108), some believe that ethical individualism is not a sufficient requirement in measuring capabilities and that any measurement of individualism should also include ontological individualism, a measurement of:

> *"nothing more than the sum of individuals and their properties"* (Alkire & Deneulin, 2009, p.35)

However, most scholars within the capability literature "embrace ethical individualism" (Robeyns, 2005b, p.109) and the belief remains that the capability approach addresses the issues of importance concerning individual well-being. This offers a challenge to those assessing societal welfare across populations who want to embrace the capability approach, but require a summary measure of some kind to compare between populations. However, as the example of the HDI and the focus on functioning variables within the literature, the theoretical richness of the capability approach has been interpreted for practical empirical applications to add a role for the theory to influence decision-making and public policy within a summary index.

3.2.1.4.4 Critiquing the link between the HDI and the capability approach

There are criticisms against the summation of the capability approach within a single index (Fukuda-Parr, 2003). It has also been argued that including a measure of income within an index, such as the HDI, is counter-intuitive in providing an alternative evaluative base to a uni-dimensional focus on human progress from welfare economic measures such as Gross Domestic Product (GDP) (Gasper, 2007). However, as much as the QALY has put health economics on the policy map, the HDI has helped to show that the capability approach can be applied as an alternative theoretical basis to welfare economics.

3.2.2 How is the capability approach different from health economic theory?

While the capability approach is primarily a critique of the use of welfare economic evaluation methods in assessing human well-being, it is important to note that the foundations of extra-welfarism, in health economics, derive in part from the capability approach (Culyer, 1989; Brouwer et al., 2008).

The underlying theory of welfarism was explained in detail in Chapter 2. The practical examples of this application within the health economics field have focused on the economic evaluative practice of cost-benefit analysis (Mishan, 1988) where the most common outcome measurement has been with the use of the willingness-to-pay method (McIntosh et al., 2010). There have been notable limitations with this method beyond Sen's critique of the welfare economics, including the use of stated preference of individuals for treatment in areas where they may not be able to weigh up alternative options realistically because of a lack of knowledge within the field (Cookson, 2003) and also that such data gathering is likely to favour individuals with the largest income, as they are likely to put a higher value on health treatments than those on modest or low incomes (Hammitt, 2002).

Extra-welfarism, as it is commonly known in health economics, has tried to move away from some of the problems associated with the normative foundations of welfarism to be more useful to healthcare decision-makers (Sugden & Williams, 1978) and to align itself more closely with the capability approach (Culyer, 1989). Under extra-welfarism, health economics has tried to avoid some of the problems associated with cost-benefit analysis in healthcare. The main priority of "extra-welfarism" focuses on the incorporation of information other than utility for consideration in evaluation (Culyer, 1989). While this would be referred to as the "extra" part of the extra-welfarism approach, it is also known as non-welfarism in welfare economics, as it moves away from utility as the sole measurement of individual well-being (Morris et al., 2007).

Although Culyer's (1989) extra-welfarism draws on the capability approach, a number of health economists have highlighted that the full implementation of Sen's concepts has not taken place with extra-welfarism theory as currently practiced (Cookson, 2005b; Coast et al., 2008d) . Many of the key aspects of welfarism theory are still intact, which contradicts Sen's capability theory. While the focus of welfarism is on utility maximisation, extra-welfarism, and in particular cost-utility analysis (using the QALY as the outcome measure) focuses primarily on health maximisation (Coast et al., 2008d). Even though this may be considered by some, or even many, as a superior basis for evaluation within health economics (Williams, 1985; Gold et al., 1996; Lipscomb et al., 2009; Edlin et al., 2013), it does not incorporate some of the main theory which the capability approach is based on, such as the ability to incorporate more than one dimension of individual well-being (Coast et al., 2008c).

Therefore, the "extra-welfarism" approach could favour treatments that are supposedly more beneficial in CUA, but could forgo information of importance to well-being freedom, such as the loss of capability which is not captured when looking at health only. As Sen points out in much of his analysis, utility, in terms of desire fulfilment, is an important functioning for individual well-being but it is not the only factor and as such not sufficient as the sole basis of analysis (Sen, 1992). Similarly, it is argued here that while achieving the best health state may be important to an individual, it is not the only factor which affects well-being freedom for that person.

Another aspect of the capability approach which extra-welfarism does not adopt in its evaluative process is the actual evaluation of capabilities itself (Coast et al., 2008d). However, this limitation has not been restricted to health economics alone, where much capability analysis relies on measuring functioning rather than capability (Robeyns, 2006). The focus of the extra-welfarism approach remains largely on functionings, specifically health functioning. This can be seen quite clearly in the common outcome measurement used under extra-welfarism theory, the QALY, which measures quality of life by the impact a treatment has on improving health function and not broader well-being (Drummond et al., 2009).

A final feature which is not addressed in extra-welfarism is the issue of agency, which was raised in Sen's approach. While there have been attempts to conceptualise how this could be taken into account in relation to health policy (see Section 3.3), there has been difficulty in operationalising such considerations for evaluation. Again, this is not limited to health economics, but is a broader obstacle for researchers wanting to operationalise the capability approach (Alkire 2009).

3.3 CONCEPTUALISING THE CAPABILITY APPROACH FOR ASSESSING HEALTH

While the capability approach can be applied across a wide range of topics and disciplines, the main focus of this section is to analyse the literature which has conceptualised the capability approach within health. The two primary conceptualisations of the approach to date are presented within Sections 3.3.1 and 3.3.2. In Section 3.3.3 a critique of the two conceptualisations is presented from the standpoint of applying the conceptualisations within an evaluation framework.

3.3.1 Ruger: Health and Social Justice

Arguably the most comprehensive attempt to develop the capability approach within the fields of health ethics, policy and law, has been made through numerous studies over a decade by Jennifer Prah Ruger (Ruger, 1998; Ruger, 2004; Ruger, 2006; Ruger, 2010a; Ruger, 2010b; Ruger, 2011; Ruger, 2012). Recently, Ruger has compiled her attempts within a book (Ruger 2010a) to provide a single conceptualisation of the theoretical justification for what she calls the "health capability paradigm" (Ruger 2010a & 2010b).

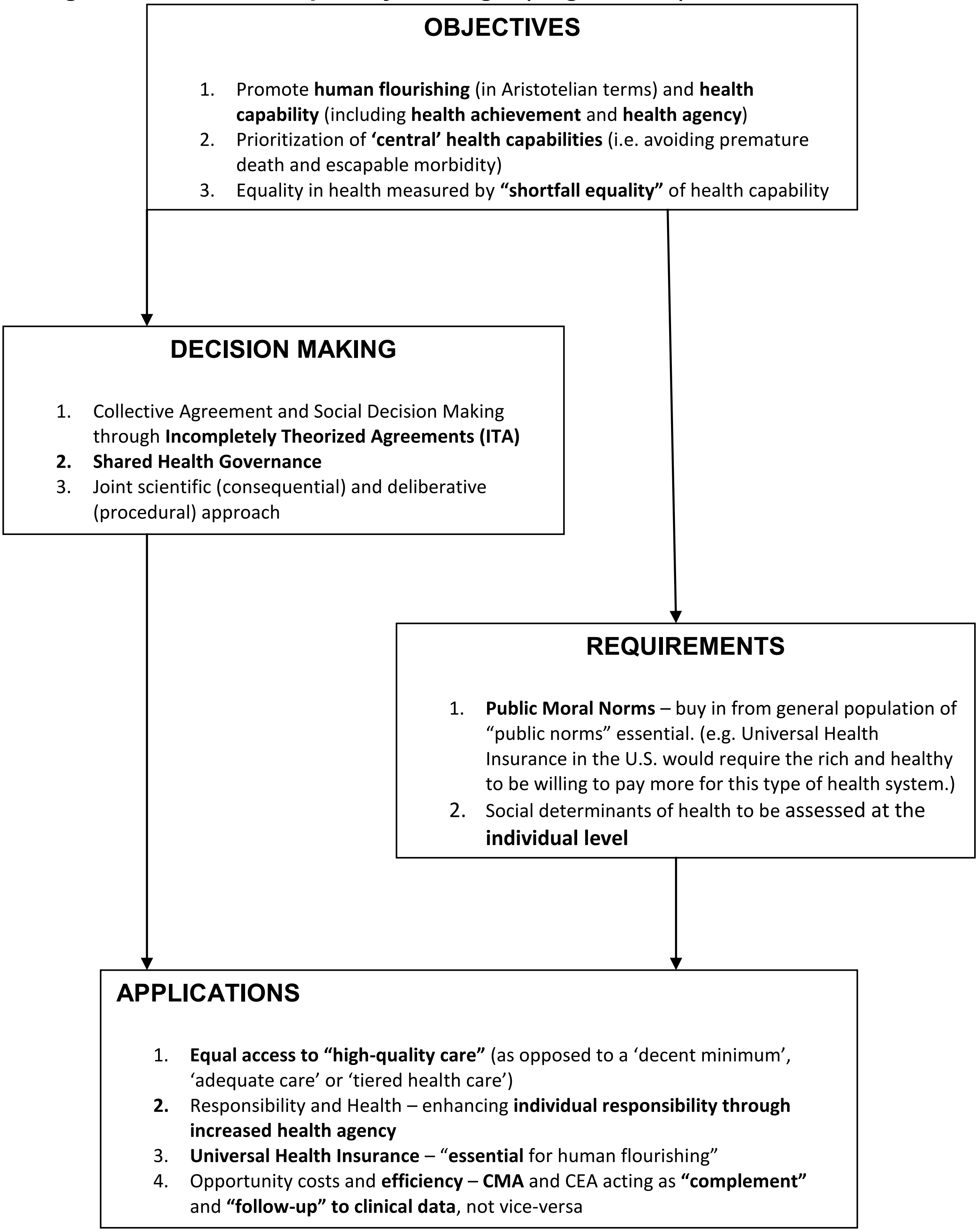
While Ruger's argument for Health Justice appears on first viewing to be a very complex collection of ideas, theory and applications across a wide variety of disciplines that are not easily digestible (Reinhardt, 2011), the health capability paradigm can be broken down into four distinct parts. The health capability paradigm consists of a number of (1) objectives; (2) rules for decision making (3) application requirements and (4) examples of how applications ought to be developed (see Figure 3).

3.3.1.1 Objectives of the Health Capability Paradigm

There are a number of objectives within the health capability paradigm. The primary objective is the promotion of human flourishing, an Aristotelian understanding of the good life, combined with health capabilities, which includes Sen's notions of capability, functioning and agency, specifically applied to health. The prioritisation of 'central' health capabilities is said to take precedence in this paradigm over 'secondary' capabilities. Ruger defines the two central human capabilities as:

> *"the capability to avoid premature mortality and the capability to avoid escapable morbidity"* (Ruger, 2010a, p. 61)

Figure 3 The Health Capability Paradigm (Ruger 2010a)



However, this is not a very helpful objective for allocating resources as the majority if not all interventions are likely to have an impact on one or both of the above “central” health capabilities. The most intriguing objective of Ruger’s framework is her objective for equality in health, which is called “shortfall equality” in health capabilities, i.e.:

> *“the shortfalls of actual achievements from their respective maximal achievements”* (Sen, 1992, p. 90).

This is proposed as an alternative to “attainment equality”, the levels of achievements actually attained (Sen, 1992). Ruger argues where the former may be more beneficial than the latter when evaluating health capabilities:

> *“This (shortfall equality) might be particularly relevant for assessing health capabilities of people with disabilities because it accounts for differences in the maximal potential for health functioning without ‘levelling down’ achievement goals for the entire group”* (Ruger 2010a, p.90)

Such an approach in practice would be interesting when assessing health capabilities. Unfortunately, there is no clear example within Ruger’s (2010a) work as to how it would be applied to prioritise one group of individuals over another.

3.3.1.2 Decision Making Rules in the Health Capability Paradigm

How Ruger plans to see through these objectives requires a number of decision making rules to be operated simultaneously. First, there is a need for a joint social decision-making framework within a society in what Ruger refers to as "shared health governance" (Ruger 2010a; 2011). This system of collective social governance across society requires:

> *"individuals, providers, and institutions work together to empower individuals and create an environment for all to be healthy"* (Ruger, 2010a, p.173)

The shared health governance process is founded upon a joint scientific (or consequential) and deliberative (or procedural) approach to cooperating on the appropriate allocation of resources for a given society (Ruger, 2010a). Shared health governance, as Ruger (2011) describes, combines the rationality of economic game/cooperation theory with the field of ethics, so that responsibility of resource allocation is a part of national to individual governance, to achieve:

> *"the alignment between the common good and self-interest"* (Ruger, 2011, p. 32)

As a method for solving conflicting conceptions on resource allocations, consensus within shared health governance is made through a social choice theory solution known as "incompletely theorized agreements" (Ruger, 2006; Ruger, 2010a). Incompletely theorized agreements (ITA) is a concept primarily used within aiding decision-making in law (Sunstein, 1995). ITA, as detailed by Ruger (2010a), consists of three forms of agreements.

The first type of ITA are “incompletely specified agreements”, which occur when there is agreement on a general high level principle, but disagreement on mid-level associated principles. For example, people may support good health, but disagree on how good health could be achieved (Ruger, 2010a, p. 71).

The second type of ITA, called “incompletely specified and generalized agreements”, involves an agreement on a mid-level principle, but where there is disagreement on how this principle can be achieved. An example is that people may agree on Universal Health Insurance, but not on a theory of equality or public policy about how it should be implemented (Ruger, 2010a, p. 72).

Finally, “incompletely theorized agreements on particular outcomes” is used to aid decision-making when agreement is reached on low-level principles, but which are perhaps not justifiable by a single higher principle. This happens when people might agree on a particular judgement, but for varying reasons (Ruger, 2010a, p. 72.).

Ruger (2010a) believes that the ITA approach could have a place in setting public policy, and particularly for setting policy within the health capability paradigm. This is because the likely interpretation of health capability will be different among individuals within a particular population. The ITA approach also adds flexibility when making decisions where:

> *“different paths lead to the same conclusion”* (Ruger, 2010a, p. 74)

3.3.1.3 Application requirements in the Health Capability Paradigm

For the health capability paradigm objectives and decision making practices to be met in practice, the theoretical argument of the approach needs to be simultaneously equipped with a certain level of general population consensus that such an approach is correct. Ruger's objective of health capability (including health agency and health functioning) for public policy could not be implemented in practice without a general consensus that health capability was generally agreed upon within society as something worth pursuing. Ruger (2010a) refers to this societal consensus as "public moral norms". For example, there would need to be an ethical commitment at a national level to finance a Universal Health Insurance scheme in the United States, a commitment to collective action for public involvement (Ruger, 2010a, p.14).

Additionally, in helping to achieve the objectives of the health capability paradigm, the social determinants of health need to be assessed on an individual basis (Ruger, 2010a). Ruger appears to reject the pursuit of reducing socio-economic inequalities (i.e. health and income) in improving societal welfare, as advanced by others (Daniels et al., 2000). However, it is not entirely clear why this is objected to in the health capability paradigm.

3.3.1.4 Applying the Health Capability Paradigm in practice

No matter how theoretically appealing some may find the health capability paradigm, the ability to apply the theory into practice is essential. Four examples of "applications" as described by Ruger (2010a) are presented here.

While the health capability paradigm is supposedly set within a global context, there is a clear focus on the United States public policy towards health within this work. Ruger (2010a) argues that Universal Health Insurance is "essential for human flourishing", although she points out that Universal Health Insurance could be justified under many normative theories, including a utilitarian or welfare economic theory of justice (Ruger, 2010a, p.220). Within her analysis of the attempt to introduce Universal Health Insurance within the United States in the early 1990s, Ruger states the proposal at that time represented the second type of incompletely theorized agreement (incompletely specified and generalized agreement). Ruger (2010a) argues this type of ITA is unstable to carry out in practice due to a lack of one overall high-level principle or one lower-level principle within Universal Health Insurance (a mid-level principle) that could have brought the change within US healthcare, which the majority of Americans supported at one point in time (Ruger 2010a, p.225). Ruger (2010a) illustrates what she calls a "wedge theory", an attack from opponents of the mid-level principle with numerous high-level and low-level principles upon which there is no general consensus. Ruger (2010a) suggests such attacks can lead to problems in reaching a general consensus through with the second form of ITA.

The second "application" by Ruger (2010a) states that the equity argument of the health capability paradigm must ensure:

> *"equal access to high-quality care, not a 'decent minimum', 'adequate care', or 'tiered health care'"* (Ruger, 2010a, p. 8)

Once more, this application of the capability paradigm appears aimed at economically developed nations like the United States, where a decent minimum at least is expected from public policy for health. Ruger has expanded on this equity argument more recently, where in the United States a state-by-state approach to setting healthcare objectives has been proposed by the health and human services (HHS) agency (Ruger 2012). Ruger labels such an approach "inadequate" as it would give:

> *"potential for discrimination against patients with rare, severe and costly health conditions"* and that *"A better approach would be to establish uniform standards so that all Americans would have access to the same high-quality goods and services"* (Ruger 2012, p. 682)

Third, by focusing on health capability, which includes health agency, the paradigm shifts responsibility of health towards the individual with increased agency for their health choices. Whilst an intriguing proposition, that with added health capability in terms of agency necessitates additional responsibility to make healthy choices, Ruger believes that it is not feasible to assess in practice, given the varying degrees of responsibility people contribute to their own health, combined with their genetic make-up (Ruger, 2010a, pp. 9-10).

The final application presented in Ruger's (2010a) book is the incorporation of opportunity costs within the paradigm, recognising that for resource allocation:

> *"A robust concept on opportunity costs is necessary"* (Ruger, 2010a, p.40).

However, Ruger (2010a) dismisses the idea of using valued health status outcomes like QALYs or societal value outcomes like Saved Young Life Equivalents (SAVEs) (Nord, 1992). Such an approach, Ruger argues is:

> *"monistic - evaluating health programmes or policies by one measure"* (Ruger, 2010a, pp. 23-24)

Ruger (2010a) suggests using cost-minimisation analysis (CMA) as the primary method of assessing between different interventions, although recognising that cost-effectiveness analysis (CEA) can also be useful at times. The role of health economics in aiding decision-making is of small importance within Ruger's paradigm, and should play a secondary role to clinical effectiveness:

> *"Under the health capability paradigm, economic solutions should follow and complement clinical progress, not vice versa"* (Ruger 2010a, p. 195)

3.3.2 Venkatapuram: Health Justice

In "Health Justice: An Argument for the Capabilities Approach" Venkatapuram (2011) makes an argument for "the capability to be healthy" to be the primary basis for a theory of health justice. Venkatapuram rejects standard definitions of health as the absence of disease or illness (Boorse, 1975; Boorse, 1977). Instead, Venkatapuram argues for an alternative theory of health based upon the capability approach, the capability to be healthy. His conception is a combination of two philosophers work. Firstly, he justifies an alternative theory of health based on the work of Lennart Nordenfelt, who reasoned that health could be conceptualised in a more holistic manner, arguing that health is the ability to achieve a set of "vital goals"

(Nordenfelt, 1995; Nordenfelt, 2007). However, Venkatapuram argued that aspects of Nordenfelt's theory were not compatible with the capability approach, arguing against the emphasis on "subjective preferences" and "cultural relativism" (Venkatapuram, 2011). Instead, Venkatapuram favours the use of the list of ten central human capabilities, developed by the eminent capability philosopher Martha Nussbaum (Nussbaum, 2000; Nussbaum, 2011) (see Table 1), as the set of vital goals to be accounted for in Nordenfelt's conception of health. Venkatapuram argues that Nussbaum's list of "pre-political moral entitlements" can be applied across countries as:

> *"a minimum conception of vital goals that is applicable across the human species"* (Venkatapuram 2011, p. 31)

In order to measure Venkatapuram's vital goals, he next turns to the epidemiology literature to examine how his methods could be used in practice. Once more Venkatapuram draws from work linked to the capability field in terms of research on famines and the development of an entitlement theory to understand such phenomenon (Drèze & Sen, 1989). The entitlement theory is used by Venkatapuram to emphasise the need for moving beyond an explanatory model approach within epidemiology. Instead, he argues, epidemiology should focus on four causal factors: individual biology, physical exposures, social conditions and individual agency (Venkatapuram, 2011). He links this approach to measuring causation within the social epidemiology literature, where researchers are keen to emphasise the role the environment plays within both individual and population health over their life course (Kelly et al., 2010). Venkatapuram argues that a social epidemiological approach can be linked within an entitlement analysis, used principally to measure the capability to be adequately nourished previously, to measure his set of vital goals more broadly (Venkatapuram, 2011).

The second part of Venkatapuram's focuses on justifying his conception of health justice within the theory of the capability approach. From his capability literature review, he focuses primarily on the distinction between the writings of Sen and Nussbaum. He uses this review to justify the role of the capability approach in health as the ethical justification for the assessment of the "capability to be healthy" in terms of the vital goals of Nussbaum's list. He argues that his approach is again a combination of two alternative conceptions of the capability approach, by applying Sen's analytical methods with Nussbaum's list to develop a:

> *"Sen-Nussbaum 'hybrid' argument"* (Venkatapuram 2011, p. 34)

In the final part of his work, Venkatapuram contrasts his proposal for the capability to be healthy with five alternative ways in which claims on health have been made. The health equity argument, primarily the work of Margaret Whitehead (Whitehead, 1990), the linkage of health and human rights (Mann et al., 1994), a resource theory based on equality of opportunity (Daniels, 2008) and luck egalitarianism (Segall, 2010). However, of greatest interest for this thesis is his comparison with welfarist claims on health.

While Venkatapuram considers in great detail the links between the resource based equality of opportunity in health and the capability to be healthy, he finds little in common with the welfarist approach. He states:

> *"a welfarist approach to health would probably focus on just maximising a single metric of health across individuals"* (Venkatapuram 2011, p. 184)

He then draws from work by Sen (Sen, 1999) to formulate the welfarist approach as a consequentialist focus on preference satisfaction and sum-ranking across populations (although he is really referring to extra-welfarism as presented in Chapter 2). All three components of the welfare theory are objected to by Venkatapuram. He rejects the singular focus on outcomes in health economics as "myopic" or short sighted (Venkatapuram 2011, p. 185). The role of aggregating across populations for average welfare improvements is also rejected as the "individuals on the tail of the population" need to be accounted for, especially if their disadvantage includes more than health alone (Venkatapuram, p. 188). His primary objection to the welfarist approach is the role of subjective preferences, however, especially when relying on assessment of physical and mental functioning, stating:

> *"The 'happy sick' or 'worried well' both point to the possible perverse results of relying wholly on subjective well-being"* (Venkatapuram 2011, p. 186)

The second section in part 3 of his work asks how the capability to be healthy proposition handles groups rather than individuals. There is an inherent focus on individualism within the capability approach, known as ethical individualism, where individuals are the ultimate units of moral concern (Robeyns, 2005b). As already mentioned in his critique, even with a shift towards capabilities, the maximisation of overall capability should not be traded off at the expense of the minority. He states:

> *"Improving the capabilities of the many does not make up for others not having their minimal or threshold level of capabilities commensurate with human dignity"* (Venkatapuram 2011, p. 212)

Any focus on group capability to be healthy must also account for the individual entitlement of the capability to be healthy across Nussbaum's list, in the argument presented by Venkatapuram. Part 3 closes with an argument of the capability to be healthy as an argument for a global conception of health, which can be applied across national borders and monetarily rich and poor nationalities.

3.3.3 Comparing, contrasting and critiquing conceptualisations of the capability approach for health

Unlike Ruger (2010a) who, to a certain degree, ignores the social determinants of health in her health capability paradigm (Saith, 2011), Venkatapuram (2011) places the social determinants of health at the heart of the capability to be healthy. He describes the capability to be healthy as a:

> *"meta-capability; an overarching capability to achieve a cluster of basic capabilities to do and be things that make up a minimally good human life in the contemporary world"* (Venkatapuram, 2011, p. 20)

While the capability to be healthy can be placed within Nussbaum's conceptualisation of basic capabilities, there are many similarities between Ruger's and Venkatapuram's proposals. Venkatapuram also places the role of personal responsibility within his conception of the capability to be healthy, albeit with a caveat:

> *"From the capabilities approach perspective, individuals become morally responsible for their choices in light of their capabilities, not irrespective of their capabilities"* (Venkatapuram 2011, p. 22)

The pluralist nature of capabilities is emphasised in Venkatapuram's focus on the social determinants of health, with human health and longevity as a primary focus in his conception of the capability to be healthy, as in Ruger's "core" health capabilities. It may be then somewhat surprising that both theories reject the role of summary measures like the QALY and DALY out of hand given they would appear to complement the primary objectives of the theories, albeit with a notable difference on the focus on capabilities rather than health status. However, it is primarily related to their stances on inequality and equity that such outcomes appear to be rejected. Unlike Ruger (2010a) who proposed equal access to high-quality healthcare as the equity criterion for health capability, Venkatapuram draws on work rejected by Ruger, and focuses on a more minimalist objective in reducing health inequalities:

> *"The idea of sufficient and equitable capabilities commensurate with equal human dignity in the modern world aims to capture such multidimensional concern"* (Venkatapuram, 2011, p. 21).

Both Ruger and Venkatapuram argue that their objectives are not compatible with "monist" measures used within health economics currently. Unlike Ruger (2010a), Venkatapuram (2011) does not indicate how the capability to be healthy should be measured beyond improving health and longevity. It may be over-critical of this conception of "health justice" given his theory is notably incomplete, but to criticise health economic outcomes related to welfarist economics and utilitarian ideals, an alternative must be proposed instead. Proposing Nussbaum's list as an alternative leaves many questions about the practical applications of Venkatapuram's theory.

Whilst both Ruger (2010a) and Venkatapuram (2011) offer an illuminating dissection of the multidimensional nature of the role in health in improving societal well-being through the capability perspective, neither proposes a practical method for evaluating choices with healthcare.

Ruger's (2010a) example of incompletely theorized agreements (ITA) in the United States in the early 1990s helps to show the health capability paradigm in practice and explain where the problems arose in the past that could have been handled differently. However, it does not give a clear rationale as to how the introduction of health capability could be achieved, without health capability as a public moral norm high end principle in the ITA framework. Since there is no guarantee that this is the general consensus of the public, it is hard to see how the health capability paradigm can be implemented fully in practice without such an agreement. How realistic her equity argument of "equal access for high-quality care" is, is an additional problem with her proposition.

Whilst Ruger (2010a) offers a critique of current health economic practices, primarily the QALY, there is no alternative suggested as to how resources should be allocated if choices need to be made between two or more interventions. Her primary choice of evaluation, CMA, appears to be a major oversight within her paradigm. It has been noted within the health economics literature that the role of CMA should be diminished because of issues around uncertainty (rather than anything related to outcome measurement) (Briggs & O'Brien, 2001) and more recently it has been argued that the role of CMA within economic evaluation:

> *"should not only be dead but be buried"* (Dakin & Wordsworth, 2013)

This has led to questions of Ruger's knowledge of the scope of work within health economics and how it relates to her paradigm (Lorgelly et al., 2010b; Saith, 2011).

Venkatapuram (2011) has similar qualms about the use of health economic outcome measures, stating that the focus of the capability approach should be on the individual. However, his objection with such outcomes is questionable. He states:

> *"Fundamentally, the difference lies in relying on individual's subjective preferences about states of physical and mental functioning...Yet, despite the external visibility of pain and suffering, people's valuation of their own physiological functioning is not a good indicator of their claims for social support."* (Venkatapuram, 2011, p. 186).

The basis of this critique is drawn from research by Sen on self-reported morbidity in India, where regions with higher life expectancy reported the highest amount of morbidity problems, compared to poorer regions with less life expectancy (Sen, 2002). Where Venkatapuram's analysis falls down is that Sen (2002) also recognises that internal morbidity related to health cannot be reported externally as it can be difficult for an observer to quantify mental illness or chronic pain (Sen 2002). There is no clear rationale as to why a so called "objective" view on health would supposedly be a more accurate description of an individual's health status in this scenario.

Venkatapuram is strident in his opposition to outcome measures such as QALYs and DALYs. However, he ends his critique of welfarism by stating than an extra-welfarist approach is required within health (Venkatapuram 2011, p. 189). He seems unaware that such outcomes used within health would consider themselves "extra-welfarist" too. It would appear that, like Ruger, Venkatapuram offers a critique of economics in health without a complete understanding of what the different approaches within health economics are.

Somewhat more puzzling is Venkatapuram's (2011) distinct bewilderment with economic concepts. While there are many questionable things within the health economics literature, some concepts are taken as a given and are unavoidable. However, this does not appear to hold for Venkatapuram:

> *"They maintain that resources are always finite, all individuals cannot be helped, and therefore, weighting lives is unavoidable and must be tackled head on"*
> (Venkatapuram 2011, p. 186)

Indeed any comparison between individuals seems a non-starter within Venkatapuram's conception of health justice. He seems taken aback by the idea that QALYs could even be extended beyond their current form to all aspects of public policy as suggested by one notable economic philosopher (Broome, 2006).

At the start of his critique of welfarism, Venkatapuram (2011) made a very important point: that all states of distributive justice have a metric (object of justice) and a rule (how to distribute the object) (Anderson, 2010). Throughout both lengthy conceptions of the capability approach for health, it is fair to say that neither his nor Ruger's (2010a) account of distributive justice offers an alternative combination of a metric and decision rule that could be used instead of current outcome measures like QALYs and DALYs. While a substantial effort is made to conceptualise the capability to be healthy as a measure of social justice (Venkatapuram, 2011), the reliance of his vital goals on Nussbaum's list makes a rule based around his metric unoperationalisable within its current form. A similar critique can be levelled at Ruger, whose detailed approach to conceptualisation leaves little in the way of advising on the conduct of prospective evaluation.

3.4 ALIGNING THE CAPABILITY APPROACH WITH TRADITIONAL OUTCOMES

A number of attempts have been made to incorporate the capability approach beyond Culyer's (1989) attempts in extra-welfarism. Section 3.4.1 looks at further attempts to align the capability approach with current health economic outcome measures. The second section presents a critique of such an approach.

3.4.1 Capability and the QALY outcome

The first attempt 'post-Culyer' to incorporate the capability approach within a health economic evaluation format was by Cookson (2005b) (although it had been previously suggested as an alternative to HRQoL measures (Verkerk et al., 2001)). Cookson (2005b) attempts to bring the current QALY outcome in line with the capability approach:

> *"to re-interpret the QALY as a cardinal and interpersonally comparable index of the value of the individual's capability set"* (Cookson, 2005b, p. 818).

In his analysis, Cookson (2005b) suggests rejecting a welfare economic interpretation of a "utility QALY" and moving beyond the general consensus of a "health QALY". Cookson (2005b) instead recommends the continuing use of the QALY, but suggests a reshaping of the current QALY as a capability set, representing a "capability QALY".

Cookson (2005b) argues that, in his view, the QALY outcome is a feasible option for assessing health interventions through the capability approach. He believes that the direct estimation and valuation of capability sets is not feasible and he also rejects alternative preference based measures used in public policy evaluation such as WTP as being "inadequate" for capturing capability as intended by Sen and others. Cookson refers to the use of his capability QALY as a measure of:

> *"capability efficiency alone (i.e. maximising the aggregate value of individual capability sets, ignoring equity considerations)"* (Cookson, 2005b, p. 824).

Cookson advocates that the capability QALY be used as an alternative to the health QALY for a number of reasons. Firstly, the capability QALY captures health and non-health within the same measure (i.e. "non-separability"). Secondly, the "process of care" can be captured with a shift to a focus on non-health functionings within capability sets. Finally, his approach can also account for the value different people could attach to achieving the same level of functioning attainment (Cookson, 2005b).

In a similar theme, Bleichrodt and Quiggin (Bleichrodt & Quiggin, 2013) offer a formulation of the QALY which they argue can be interpreted as:

> *"a local approximation to a ranking over capabilities"* (Bleichrodt & Quiggin, 2013, p.129)

Cookson (2005b) recognises the need for incorporating broader non-health functionings within the QALY, but he feels that the EQ-5D dimension "usual activities" fulfils this role currently (Cookson, 2005a). However, Bleichrodt and Quiggin (2013) argue that such an interpretation is not required. Bleichrodt and Quiggin (2013) model "capabilities as menus" (i.e. a person chooses a capability set from a menu of capabilities, from which they then choose their functioning attainment). They draw upon previous methodology on incorporating a two step choice process into an axiomatic framework (Kreps, 1979) to test their assumptions that the QALY is a compatible measure of their formulation of "capabilities as menus". They prove, under a number of assumptions, which include freedom of choice and preferences for capability sets, that:

> *"any rankings of capabilities gives rise to a 'shadow price' for QALYs"* (Bleichordt & Quiggin, 2013, p.129)

3.4.2 Critiquing the formulation of QALYs with the capability approach

Whilst the above suggestions are impressive for their efforts in their meticulous conceptualisation, there are a number of problems which neither Cookson (2005b) or Bleichrodt and Quiggin (2013) take into consideration. Firstly, Cookson's definition of “capability efficiency” as the maximisation of capability, regardless of equity concerns is at odds with both Ruger’s (2010a) and Venkatapuram’s (2011) conceptualisations of the capability approach for health. Additionally, the objective of maximising health has been a critique of those who support a move towards the capability approach (Coast, 2009; Smith et al., 2012; Payne et al., 2013). Therefore, a uniform interpretation of the purpose of evaluation within health under a capability approach being to maximise absolute levels of capability is not necessarily a good reflection of how the capability approach, as understood more generally, should be employed in practice.

Secondly, whilst the model presented by Bleichrodt and Quiggin (2013) is mathematically precise, it appears distinctly at odds with the capability approach in assuming a capability set with the highest attainable functioning will automatically be chosen. Notably, their model does not take into account the role of agency within a capability set. Therefore, their formulation would appear more appropriate for a justification of a functionings only approach or even the current extra-welfarist health economics framework rather than a capability framework.

Additionally, advocates of the capability approach within health economics believe that adopting such an approach should provide a more encompassing basis for evaluation than that involved within the QALY measure. For example, Coast and colleagues (Coast et al., 2008c) suggest that a broader evaluative space based on capabilities would encapsulate non-health benefits for interventions like public health. This view of “more than health” appears a strong theme that comes from a focus on the capability approach to healthcare rationing (Anand &

Dolan, 2005). Specific areas such as chronic pain (Kinghorn, 2010), public health (Lorgelly et al., 2010a), social care (Grewal et al., 2006; Netten et al., 2012), mental health (Simon et al., 2013) and complex interventions (Payne et al., 2013) have been identified as areas that would benefit from the capability concept of a broader set of benefits than is measured by HRQoL instruments currently.

A final critique of the re-interpretation of the QALY as a capability measure is that it makes no attempt to measure capabilities, focusing instead on the assessment of functionings used to calculate QALYs. This appears a dilution of what the capability approach tries to encompass.

3.5 MEASURING CAPABILITY

Section 3.4 outlined an argument against capturing capabilities directly as unfeasible (Cookson, 2005b). However, a number of researchers have attempted to develop questionnaires representing the capability approach more closely for use in decision-making in healthcare resource allocation. In this section, three attempts to incorporate measures of capability for use in such decision-making are outlined. The three attempts discussed are those that are most advanced and related to the UK health system. However, it is worth noting that other attempts to generate capability outcomes have been made (Greco et al., 2009; Kinghorn, 2010).

3.5.1 Capability Indicators: The OxCAP family of instruments

The first attempt to measure capabilities directly by using existing data from household and panel surveys was developed initially by Anand and colleagues (Anand et al., 2009). Instead of trying to develop measures of capability directly, Anand et al. (2009) proposed that "capability indicators" could be constructed from Nussbaum's list of 10 central human capabilities, similarly to Venkatapuram's theory for the capability to be healthy (see Section 3.3.2). Here, capability is largely inferred through questions of individual achieved functioning. The use of secondary data to infer capabilities from achieved functionings has been a popular method for the application of the capability approach in practice, primarily due to the ease of using data that has already been collected (Kuklys & Robeyns, 2005; Chiappero-Martinetti & Roche, 2009). Subsequent questionnaires developed using Nussbaum's list as compiled by Anand are referred to as questionnaires within the OxCAP (Oxford Capability) family of instruments (even though OCAP is used as abbreviation for some of the questionnaires too).

Originally developed in 2005 (Anand et al., 2005), the OCAP survey contains 64 indicators of capability, drawn from the British Household Panel Survey. The primary goal of Anand et al. (2009) was to test two hypotheses from their capability indicators – firstly, if their indicators are related through a measure of subjective well-being in terms of satisfaction with life and secondly, whether people place different values on capability indicators. These tests are carried out by comparing results from their cross-sectional dataset (n=1000) with a measure of life satisfaction to give a measure of subjective well-being (SWB) or happiness. The 64 capability indicators in OCAP can be seen in Table 2, column 2.

The advantage of incorporating existing questionnaires from a pre-existing panel survey was that similar questions are likely to be collected across a number of countries, which could allow for cross-country comparison of capability indicators (Anand et al. 2009). However, it does have the problem of being limited to data which are already collected and may not capture capability as accurately as developing a questionnaire with this exact purpose. This is particularly a problem for the broad and, in some cases, vague interpretations that can be drawn from Nussbaum's list. Refined lists may be more appropriate for policy specific instruments, such as Robeyns (2003) set of capabilities for assessing gender inequality (see Table 1).

One of the problems of relying on secondary data was shown by Anand and colleagues (2009) as they used their capability indicators to predict life satisfaction, a 7-part Likert scale of subjective well-being (where 1 is completely dissatisfied and 7 is completely satisfied with life). By predicting life satisfaction through their indicators using ordinary least squares (OLS) regression, only 17 of the 64 indicators were found to be statistically significant (see Table 2, column 2). Of greatest concern is that from the ten capability categories, only six have statistically significant indicators of life satisfaction within OCAP. This raises another concern with relying upon secondary data to inform information related to individual capability.

Table 2 The OxCap family of questionnaires

Nussbaum's capability list (2000)	**OCAP (Anand et al. 2009)**	**OCAP-18 (Lorgelly et al. 2008)**	**OxCap-MH (Simon et al. 2013)**
1.Life	Life Expectancy	Life Expectancy	Life expectancy
2.Bodily health	Health limits activities Reproductive health Adequately nourished Adequate shelter* Ability to move home	Health limits activities Adequate shelter	Adequate shelter
3.Bodily integrity	Safe during day Safe during night Previous violent assault Future violent assault Past sexual assault Future sexual assault Past domestic violence* Future domestic violence Sexual satisfaction* Reproduction choice	Safe walking alone near your home Future assault (any)	Safe walking alone near home Future assault (any)
4.Senses, imagination, thought	Education Uses imagination Political expression Exercise religion Enjoys activities	Political and religious expression Uses imagination	Political and religious expression Uses imagination Access to interesting activities (or employment)
5.Emotions	Make friends Family love* Expresses feelings* Lost Sleep Under Strain	Enjoy love and friendship of family and friends Lost sleep	Enjoy love and friendship of family and friends Lost sleep
6.Practical reason	Concept of good life Plan of Life Evaluates Life* Useful role*	Free to decide how to live life	Free to decide how to live life
7.Affiliation	Respects others* Takes holidays* Meets friends Thinks of others Feels worthless* Past Discrimination (question each for race,gender,sexual orientation, religious and age discrimination) Future Discrimination (five questions – same as past discrimination)	Respect others Able to meet people socially Likelihood of discrimination outside of work	Respect others Able to meet people socially Likelihood of discrimination
8.Other species	Appreciates plants, animals, nature	Able to appreciate plants animals, nature	Able to appreciate plants animals, nature
9.Play	Enjoy recreation	Ability to enjoy recreation	Ability to enjoy recreation
10.Control over one's life	Participate in politics Owns home Discrimination (work) Past* and Future* (10 questions, same categories as capability 7, affiliation, discrimination questions) Expect stop and search* Skills used at work* Useful role at work Relate to colleagues Respected by colleagues	Participate in local decisions Owns home Current or future discrimination within work	Participate in local decisions Owns home

*significant predictors of life satisfaction in Anand et al. (2009)

To address the issues raised above and to adjust the OCAP for evaluating health interventions, Lorgelly et al. (2008) developed a refinement of the OCAP questionnaire. Lorgelly and colleagues (2008) aims were not only to further develop and refine the OCAP survey, but to validate the survey so that it could be used to evaluate public health interventions. After a number of phases of focus groups, factor analysis, pilot questionnaires and interviews, 18 questions across Nussbaum's 10 central human capabilities were found to be of most relevance. Additionally, some questions were re-worded so that the capability of an individual and not their functioning levels was captured (Lorgelly et al., 2008).

To develop the measure for economic evaluations, Lorgelly and colleagues (2008) provide an index of capability for their refinement of OCAP (OCAP-18, see Table 2, column 3). All capability indicators (18) are valued equally, with 1 assigned to the highest level for each question, resulting in a scale of 0-18 (binary questions given a 0 or 1 value, while questions with 5 options were given values of 0, 0.25, 0.5, 0.75 and 1). This resulted in an OCAP-18 mean score for a general population of Glasgow (n=166) of 12.44 (ranging from 3-17.75). This capability score correlates strongly with the EQ-5D score (mean 0.757) for the same population (pairwise correlation of 0.576; p-value:<0.001) (Lorgelly et al., 2008).

However, Lorgelly and colleagues (2008) rightly note that their method of valuation means that some of Nussbaum's list will have extra weight if they have more indicators per capability. Additionally, there is no weighting attached when people value the 10 capabilities differently. Therefore, for example, equal value is attached to the capability indicators for "life" and "other species" (see capability 1 and 6 in Table 2), which may not be seen as a sensible basis by decision-makers for use in prioritising health interventions in practice.

While Lorgelly and colleagues' (2008) refinement of the OCAP survey was designed for public health, it has more recently been adjusted again for the purpose of evaluating capabilities for mental health interventions (Simon et al., 2013). The work by Simon and colleagues focused on altering the original OCAP of Anand et al. (2009), but offers a very similar questionnaire to the OCAP-18 of Lorgelly et al. (2008). The primary difference between the OCAP-18 and the OxCap-MH (Oxford Capability Measure for Mental Health) is that only one capability question on discrimination appears in OxCap-MH as opposed to two (work and outside work) indicators of discrimination for OCAP-18. Some of the questions on OxCap-MH were re-worded but intend to capture the same capability principle as on the OCAP-18. The OxCap-MH also includes 18 questions in total, with an additional question related to access to activity (or employment), for senses, imagination and thought (capability four, see Table 2, column 4).

Simon and colleagues (2013) tested their capability instrument within the Oxford Community Treatment Evaluation Trial (OCTET) and developed a capability index (CAPINDEX16). They also adopt a similar valuation approach to Lorgelly et al. (2008) for the OxCap-MH, treating each question equally but placing all questions on a 1 to 5 scale instead of the 0 to 1 scale used for the OCAP-18 questions. 172 patients recorded a response for sixteen questions (CAPINDEX16, which excluded questions on property ownership and life expectancy), resulting in a mean CAPINDEX16 score of 58.40 (range 26-75), which correlated with the

EQ-5D (0.514; p-value: 0.01) and Visual Analogue Scale (0.415; p-value: 0.01) (Simon et al., 2013).

3.5.2 ASCOT measure of Social Care

Another questionnaire which has been developed with an aim to capture capability to be used in economic evaluations is the adult social care outcome toolkit (ASCOT). The ASCOT questionnaire aims to measure social care related quality of life (SCRQoL), which can then be used to measure a social care QALY (SC-QALY) and make comparisons between health related QALY interventions. The developers of ASCOT argue that social care is a particular case where there is no start and end point of treatment, so to measure the gain from social care interventions questions are posed on their current SCRQoL status as well as their expected SCRQoL state if the intervention was not available (Netten et al., 2012).

The ASCOT questionnaire has evolved through four versions to the present version, which attempts to account for Sen's capability theory within the latest version of the questionnaire's development (Netten et al., 2012). Originally the ASCOT instrument was called the Older People's Utility Scale (OPUS), based on five attributes (food and drink; personal care; safety; social participation and involvement; control over daily living) of social care for older people across four levels (no problem, all needs met, low unmet needs , high unmet needs) (Netten et al., 2002). This was further developed to capture social care outcomes for adults less than 65, with the new instrument including three new dimensions (which are employment and occupation, having a caring role, living at home) (Netten et al., 2005).

Version three of the ASCOT was the first attempt to incorporate the theory of Sen's capability approach within the instrument. The rationale for drawing on the capability approach in social care was outlined in a Measuring Outcomes for Public Service Users (MOPSU) study which showed that the OPUS instruments developed focused more on functionings and not on capability (Forder et al., 2008). The ASCOT questionnaire was thus developed further to measure whether social care was at or below what patients wanted and liked, rather than if the social care was good or bad, to account for the measurement of capabilities achievable and constraints on capability within the questionnaire phraseology (Caiels et al., 2010). As well as the rewording of the questionnaire, two more attributes were added between the second and the third version of ASCOT, one category on anxiety and worry and another category for dignity.

Through an extensive qualitative investigation of the attributes to reflect social care outcomes of importance to service users, the most recent ("final ASCOT") questionnaire was developed which included the five categories from the first questionnaire, as well as accommodation, cleanliness and comfort, occupation and dignity (Netten et al., 2012). Unlike the first two versions which focused on needs of the service users, each attribute on the final ASCOT has an emphasis on the highest level of each attribute on the wants and likes of patients to reflect a broader aspect of the questionnaire on capability, with the other three levels reflecting levels of basic functioning (Netten et al., 2012).

The final ASCOT version aims to reflect the SCRQOL of social service users which can be used to compare social care interventions with one another. The final ASCOT also developed a preference weighting of states, such that “1” represents the ideal state of SCRQOL and “0” represents a state equivalent to being dead. Valuation exercises were conducted through a combination of TTO and best-worst scaling (BWS), which is a type of discrete choice model (Flynn et al., 2007). Values for social care states were calculated for the general population and service users, with little difference between the values of both groups. The final value set implemented allowed for the calculation of an SC-QALY which could range from -0.19 to 1. The authors state that this study could allow for comparisons between health QALYs and SC-QALYs once a relationship between a health status instrument like the EQ-5D and the final ASCOT version is established (Netten et al., 2012).

3.5.3 ICECAP Capability Questionnaires

The ICEpop CAPability (ICECAP) measures were developed as an alternative way of capturing individual well-being compared to current measures of preference-based health status indices that are used to generate QALYs. Grewal and colleagues (2006) aimed initially to develop a measure of quality of life for older people that could cross health and social care. They found through qualitative analysis that it was the capability to achieve important functionings that was of most relevance for older people within the UK. Additionally, it wasn’t good health as an end in itself that was the objective for this population group, but it was how that good health allowed them to live a good life:

> *“it was not poor health in itself which was perceived to reduce quality of life, but the* ***influence*** *of that poor health upon each informant’s ability to achieve the attributes of quality of life, that seemed to be particularly important”* (Grewal et al. 2006, p. 1897)

Grewal and colleagues (2006) found many factors that influenced quality of life such as activities, home, family and relationships, health, standard of living and faith. However, it was the role of these factors across a number of key attributes of quality of life that were found to be of primary importance in their analysis. Five key attributes were found by Grewal et al. (2006) in total. They are:

- Attachment – “feelings of love, friendship, affection and companionship”
- Role – “having a purpose that is valued”
- Enjoyment – “notions of pleasure and joy, and a sense of satisfaction”
- Security – “ideas of feeling safe and secure...which include having sufficient finances, sufficient practical and emotional support and sufficient health”
- Control – “being independent and able to make one’s own decisions”

  (all from Grewal et al. 2006, p.1897)

These attributes of quality of life for older people led to further research which involved the development of an index of capability for older people, the ICEPOP (Investigating Choice Experiments for Preferences of Older People) CAPability measure for Older people or ICECAP-O (Coast et al. 2008). The attributes identified by Grewal and colleagues (2006) were used as the basis for developing a short self-report questionnaire, that could be used to aid decision making in health and social care (Coast et al., 2008a). An iterative process was used to test the understanding of questions and make sure that questions were interpreted as meant within the research by Grewal and colleagues (2006). This resulted in five attributes of capability across four levels of well-being, ranging from no capability to full capability for each attribute. The ICECAP-O can be seen in Appendix 1.

Values for the ICECAP-O capability index were generated for a sample of the over 65 UK population through the BWS method of valuation. BWS presents scenarios to participants whereby, for the ICECAP-O, they are asked to state their most and least favoured attribute from the five options presented to them (i.e. one from each attribute). Coast and colleagues (2008a) argue that this method is appropriate for the capability approach as it represents the values of the individuals, and not necessarily their choices or preferences, as there is no trade-off between capabilities with BWS.

The validity of the ICECAP-O has been tested in various countries for different populations. The construct validity of the ICECAP-O was established within the UK older population twice. Firstly, Coast et al. (2008) reported a relationship between the ICECAP-O categories for 315 elderly individuals with age and general well-being (i.e. a single question "are you basically satisfied with your life?") (Coast et al. 2008b, p. 969). The attributes of role, enjoyment and control were related to physical measures of health, whilst the attachment and enjoyment attributes were more closely associated with mental health measures (Coast et al., 2008b). A larger sample size of older people (n=809) was later examined, with similar associations found between ICECAP-O and poor physical and mental health (Flynn et al., 2011). Additionally, living alone, low social contact, feeling unsafe after dark and being without a faith can result in lower capability scores (Flynn et al. 2011).

Validity studies for ICECAP-O have also been conducted in the Netherlands for elderly psycho-geriatric patients (Makai et al., 2012), in Canada for older people who attended a falls prevention clinic (Davis et al., 2013), and also in the context of quality of care transition for older hospital patients in Australia (Couzner et al., 2012). The ICECAP-O was also tested within a general adult population sample in Australia, with lower income and employment being closely associated with poor capability, whilst marriage and cohabitation was a positive sign of capability for those under 65 years old (Couzner et al., 2013a).

With a growing interest in using capability instruments internationally and across a broader range of population groups than the over 65 population, it is not surprising then that this has led the ICECAP research team to develop a capability measure for the general adult population. The ICECAP instrument for adults (ICECAP-A) has recently been developed to provide a measure that would be able to capture the capability of all adults over 18 in a similar way to the ICECAP-O. A similar qualitative process to that used by Grewal et al. (2006) was carried out to capture attributes of capability wellbeing for the general UK population (Al-Janabi et al., 2012a). This resulted in five attributes: “stability”, “attachment”, “achievement”, “autonomy” and “enjoyment” (Al-Janabi et al., 2012a). The questions were then developed to capture capability across four levels for the five attributes.

Similarly to the ICECAP-O construct validity studies (Coast et al. 2008b; Flynn et al. 2011), the first ICECAP-A validation study found capability differences between health and socioeconomic groups (Al-Janabi et al., 2012b). While there was no significant difference found for those classed as deprived and non-deprived, this may be as a result of the small sample size (n=418) within this study, which might make distinctions between deprivation levels within a population difficult to distinguish.

While there are noticeable similarities between some of the attributes in the ICECAP-O and ICECAP-A, the ICECAP-A captures a wider demographic of valuable capabilities. Indeed, further qualitative investigation found that the older population struggle with the "achievement" attribute, given that it has been closely associated with purely work achievement by some individuals (Al-Janabi et al., 2013).

Finally, the ICECAP team has most recently developed an instrument to account for the attributes of capability that are important at the end of life, which is known as the ICECAP supportive care measure (ICECAP-SCM) (Sutton & Coast, 2013). Unlike the ICECAP-O and ICECAP-A which both have five capability attributes, the ICECAP-SCM consists of seven capabilities "autonomy", "love", "physical suffering", "emotional suffering", "dignity", "support", and "preparation". Given that the current focus within health economics is on health maximisation over time, the role of end of life and palliative care are areas which have not been tackled extensively within health economics. Therefore, the ICECAP-SCM offers a departure from current health economic questionnaires for aiding decision-making to assess care at this stage of life.

3.5.4 Choosing a questionnaire for resource allocation

The three approaches to capability questionnaire development are notable advances in the application of the capability approach. With three diverse approaches to questionnaire development, there are a number of similarities between the methods, but also different strengths and weaknesses attached to each approach.

Overall, the OxCAP related research is a welcome addition to the capabilities research field and particularly the refinements for assessing health interventions, indicating that there are many important capability indicators which affect human well-being and are important to consider in decision-making. The diversity of each of the three questionnaire types have a number of different strengths and weaknesses in their design. However, of particular concern would be a lack of a consistent approach to evaluating interventions across a health service. The OxCAP family of instruments do not provide a consistent approach to valuing capabilities or questions to include, which would be a major concern for making resource allocation decisions across different population groups as in the current approach of, for example, NICE.

While the ASCOT claims that the questionnaire is geared towards the capability approach, it is difficult to grasp the influence of the capability approach in practice. The most recent version of the ASCOT was developed using a similar mechanism to that of the ICECAP-O but the actual attributes of the ASCOT questionnaire have not changed much since the first questionnaire was developed, which as the authors rightly suggest is focused on functionings only (Netten et al., 2012). Therefore, the implementation of the capability approach appears an afterthought for justifying the development of the questionnaire, rather than a specific interest in applying the capability approach itself. This argument is based on the fact that the third version developed for the low-level interventions has no reference to any capability literature whatsoever (Caiels et al., 2010).

Additionally, the development of the final ASCOT version to generate a SC-QALY outcome has the aim of being a comparable instrument with the QALY outcome. However, a health QALY and an SC-QALY are measuring two different aspects of well-being. The ASCOT does not claim to be a broader measure than health related quality of life, so the comparison of the questionnaire could be further hampered by the attempt to measure capabilities rather than functionings as instruments such as the EQ-5D does. In reality, the comparison of the different health instruments that produce health QALYs is already questionable given the different areas of health captured by different health utility instruments (Brazier et al., 2004a; Grieve et al., 2009; Whitehurst & Bryan, 2011). Therefore, to be able to say that a SC-QALY is directly comparable with a health QALY, an investigation needs to be carried out in much greater detail.

Nonetheless, there are a number of important innovations within the ASCOT instrument. The recognition that it is unsatisfactory to focus on functionings only, within a social care setting, to capture changes in SCRQoL supports the argument that a capability approach to assessing both health and social care is appropriate. However, whether a measure of SCRQoL such as the final ASCOT could be used to allocate resources across a health and social services together appears unlikely. The role of ASCOT as a technically efficient measure is possible between interventions for the same intervention group or for allocating resources within the social care service only. The mechanism of measuring expected SCRQoL scores in the absence of social care interventions is also a novel method for accounting for absence of an intervention where there is no clear start or end point to a given treatment.

The ICECAP questionnaires have been qualitatively developed to capture key capabilities within the UK population and these are methods of procedure that are highly recommended in the capability approach (Alkire, 2002; Comim et al., 2008). Values within the ICECAP instruments are also an important aspect of public valuation and participation in what capabilities are important to them. While the ICECAP questionnaires attempt to capture a small number of capabilities (5 each for ICECAP-A and ICECAP-O; 7 for ICECAP-SCM), there is a possibility that these questionnaires could be challenged to be less comprehensive compared with the capability indicators on OxCap instruments and less context specific compared to the ASCOT social care instrument.

While there are a number of benefits with all three questionnaire types, the remainder of this thesis will focus on the implementation of the ICECAP questionnaires. The ICECAP questionnaires are the only questionnaires out of the three that have both a generic method of capturing individual capability, which could be applied across the health service, as well as values from the general public attached to the respective ICECAP questionnaires.

3.6 SUMMARY OF CAPABILITY, ECONOMICS AND HEALTH

In this chapter, the key terms and concepts of the capability approach have been explained. What makes the capability approach distinct from both welfarist and extra-welfarist economics (as currently practiced) was an important theme. Conceptualisations of the capability approach for health were outlined and critiqued with a key concern being the lack of emphasis within these conceptualisations, on using the approaches to aid decision-making within healthcare.

In Sections 3.4 and 3.5, attention turned to how health economists have tried to incorporate the capability approach within outcome measures. Some have formulated the QALY as the most suitable estimation of capability but this was argued to be limiting. Others within health economics have challenged the use of the outcome measure from a capability perspective. The most advanced examples of direct attempts to incorporate capability within questionnaires were presented in Section 3.5. Each of the measures has strengths and weaknesses for taking this research forward. The ICECAP measures have the advantage of conducting a valuation exercise with their measure, while the OCAP survey has the benefit of capturing capabilities that may routinely be collected within most country's household surveys. The refinements by Lorgelly et al. (2008) and Simon et al. (2013) are, however, less likely to benefit from this advantage. The ASCOT questionnaire has been developed to capture social care related quality of life to develop an outcome which could be comparable to QALYs. How such an instrument could be used to compare capabilities from interventions outside of social care would face similar difficulties to those of any condition-specific health instrument collected presently. The ICECAP instruments would appear, currently, to be the most appropriate choice for a generic instrument capability which can be applied across health and social care interventions equally.

A number of questions remain if the capability measures are to be used in health evaluations. Firstly, there is a question of whether capability questionnaires can be used in a similar way to HRQoL questionnaires which are currently used to aid decision-making within economic evaluations. Secondly, there is little guidance as to the objective and decision rule that should be used with such measures (i.e. should the capability questionnaires be used to maximise capability over time for a given population?). These issues are tackled in the next 4 chapters.

CHAPTER 4. A REVIEW OF EMPIRICAL APPLICATIONS OF THE CAPABILITY APPROACH ACROSS DISCIPLINES

4.1 INTRODUCTION

In Chapter 3, the theoretical basis of the capability approach was described as an alternative to the extra-welfarist paradigm. Chapter 3 also presented a discussion of the evolution of the capability approach within health economics and debated the merits of the approach (Verkerk et al., 2001; Anand & Dolan, 2005; Coast et al., 2008c). The recent focus of the capability approach in health economics has turned increasingly to the development of measures or indicators of capability as alternatives to HRQoL instruments (Grewal et al., 2006; Anand et al., 2009; Al-Janabi et al., 2012a). However, other authors have suggested that the capability approach could be aligned with the QALY outcome used predominantly in health economic evaluations (Cookson, 2005b; Bleichrodt & Quiggin, 2013). Therefore, no clear consensus exists as to how capability instruments, once developed, should be applied within an evaluation framework. This chapter aims to illustrate the attributes that are captured within capability measures across disciplines, how different attributes are aggregated within capability questionnaires and what is the objective and decision rules applied to capability measures once aggregated.

The chapter is structured as follows: in Section 4.2, previous empirical reviews of the capability literature are assessed. Both the reasoning behind each review and the papers identified are discussed in relation to the primary aims of this chapter. In Section 4.3, an introduction of the methodology for undertaking the literature review in this chapter is presented. This literature review implements a search strategy known as "comprehensive pearl growing" (Schlosser et al., 2006). The rationale for choosing this review type, how it works in practice and a review of the inclusion and exclusion criteria for studies is also documented in this section. In Section 4.4, the results of the review in terms of the summary statistics of the papers included/excluded are presented. Each study is categorised and described. In Section 4.5, an analytical review of the studies is presented, addressing in particular the aims of the chapter. In Section 4.6, the chapter concludes with a discussion on the findings from this chapter and what health economics can learn for future studies that adopt a capability perspective.

4.2 PREVIOUS REVIEWS OF EMPIRICAL CAPABILITY APPLICATIONS

Before undertaking an empirical review of the capability approach, a preliminary search of previous quantitative capability reviews was carried out to establish what is already known within the literature. Three reviews were identified, including one paper and two book chapters which were found in a manual search of the relevant literature. These reviews represent summaries of previous attempts to analyse how the quantification of the capability approach has evolved since the development of the capability theory since the 1980s and early 1990s (Sen, 1980; Sen, 1985; Sen, 1992; Sen, 1993). A summary of the three literature reviews is presented next in chronological order.

4.2.1 Literature Review (1): Kuklys & Robeyns (2005)

The first review of the capability literature was undertaken in a book chapter by Kuklys and Robeyns (2005), which looked to collect information on the early quantitative empirical applications of the capability approach. The focus of the review by these authors was to explore the empirical applications of the capability approach up until 2002. Following the findings from the review, Kuklys furthered the research in the area by developing a monetary-based functioning approach for assessing disability well-being through "equivalence scales" (Kuklys, 2005).

The early empirical applications of the capability approach captured by the review of Kuklys and Robeyns (2005) found that instruments developed were based on functioning achievement rather than the broader conception of capability. Their review identified seventeen studies which attempted to measure individual well-being from a capability perspective. Five of the 17 studies focused on national and regional comparisons of functioning based welfare, while the remaining twelve studies either consisted of micro analysis of functionings within a particular population group (e.g. unemployed, poor, children) or an assessment across a population group without a comparison with another region. Their review aimed to show differences in studies which chose to elicit functionings in comparison to measures based on income when assessing welfare. This change of focus is one of the primary rationales for adopting a capability based approach, so it was important to establish that differences exist when adopting an alternative object of analysis.

Four main methodological themes were used in their review to compare the selected studies:

1. selecting which functionings were of relevance for welfare assessment;
2. the measurement of the chosen functionings at the individual level;
3. the aggregation of functionings at the individual level;
4. the aggregation of individual functionings across a population.

The main finding from their review was that there were considerable distinctions between those who are classed as "functioning-poor" and those who are classified as "income-poor" in many of the studies reviewed. They also acknowledged the challenge of measuring capabilities directly, with all applications in their review focusing on functionings achieved as a proxy for capability measurement (Kuklys & Robeyns, 2005).

4.2.2 Literature Review (2): Robeyns (2006)

The second literature review is the sole peer-reviewed paper identified. Robeyns (2006) presented a scoping exercise, showing the breadth of the practical applications of the capability approach across a wide variety of disciplines. Robeyns (2006) asked what specifications are required for a study to be considered within the remit of the theoretical justification of the capability approach. She concluded that these were (Robeyns, 2006):

1. the choice between focusing on capabilities or functionings;
2. the selection of relevant capabilities to answer a specific question and;
3. the weighting of the capabilities chosen to give an overall assessment or aggregation of welfare.

Her review suggests that a number of different approaches to the above three specifications have been made in different studies (Robeyns, 2006), with no "reference case", comparable to NICE guidance for cost-effectiveness, being established for capability applications. Robeyns' (2006) review suggested that the first two issues have generally been dealt with in relation to the data available, and weights have usually been set arbitrarily as equal across the chosen dimensions. It should be clarified that this practice is not her *recommendation*, but rather reflects the state of practice as observed in her review (Robeyns, 2006).

In terms of empirical capability applications, the most interesting part of the study by Robeyns (2006), for health economics, is to show exactly where the capability approach has been applied in practice. Table 3 presents a summary of the thematic groups identified in Robeyns' (2006) review where the capability approach has been applied in terms of areas of analysis, examples of the objectives within each groups and the studies which were found in each group. In total, Robeyns (2006) identified nine areas where the capability approach had been applied at the time of the review, ranging from cross-country comparisons of well-being to small scale development projects to help alleviate poverty in deprived populations (Robeyns, 2006). Many of the areas identified by Robeyns (2006) focus on particular populations who may be unfairly treated when traditional measurement techniques are used. Such research is primarily aimed towards the poor, the disabled and studies on gender inequality. Some studies identified here aimed to address a particular policy concern, while other studies were more exploratory, looking to understand the extent that household questionnaires are able to capture indicators of capability from existing capability lists. For example Anand and van Hees (2006) examined the British Household Panel Survey (BHPS) questions in relation to Nussbaum's ten central human capabilities (Nussbaum, 2000).

Table 3 Robeyns' (2006) Overview of where the capability approach has been applied

Application Type	Examples or Explanations	Studies
General Assessments of the Human Development of Countries	Human Development Index (HDI)	(Sen, 1985) UNDP Human Development Reports (1990-present) (Fukuda-Parr, 2003)
Assessing Small-scale Development Projects	Poverty reduction projects	(Alkire, 2002)
Identifying the Poor in Developing Countries	Functionings-poor versus income poverty measures	(Ruggeri Laderchi, 1997) (Ruggeri Laderchi, 1999) (Ruggeri Laderchi et al., 2003) (Klasen, 2000) (Qizilbash, 2002) (Asali et al., 2005)
Poverty and Well-Being Assessment in Advanced Economies	Profiles of the poor; well-being trends in household surveys	(Balestrino, 1996) (Phipps, 2002) (Brandolini & D'Alessio, 1998) (Chiappero-Martinetti, 2000) (Anand & van Hees, 2006)
Deprivation of Disabled People	Achievement of functionings compared to non-disabled	(Zaidi & Burchardt, 2005) (Kuklys, 2005)
Assessing Gender Inequalities	Differences between males and females in developing and advanced countries	(Sen, 1985) (Chiappero-Martinetti, 2003) (Robeyns, 2003)
Debating Policies	Discuss and assess government policies (e.g. education, welfare state reform)	(Schokkaert & Van Ootegem, 1990) (Lewis & Giullari, 2005) (Dean et al., 2005)
Critiquing and Assessing Social Norms, Practices and Discourses	Social norms that restrict capability	(Lavaque-Manty, 2001) (Olson, 2002) (Robeyns, 2005a)
Functionings and Capabilities as Concepts in Non-normative Research	Ethnographic research Exploratory analysis	(Arends-Kuenning & Amin, 2001) (Anand et al., 2005)

Robeyns' (2006) review of applied research raises a number of challenges to those advocating the application of the capability approach. One of the most important of these is around the value added of including indicators of capability, compared to the indicators of sociological research and multidimensional well-being, which are already widely developed and used in a number of countries. A second important issue is the reliance on data that are not specifically designed to capture capabilities (Robeyns, 2006).

Given that researchers have previously specified different ways to identify which capabilities are important for specific purposes, Robeyns (2006) argues that the justification for the selection of capabilities needs to be explicit in the analysis, to ensure that the method used is transparent. In summary, Robeyns (2006) recognises the challenges associated with applying the capability approach in practice and emphasises the need for a clear explanation and justification around the three specifications (i.e. 1. the choice between focusing on capabilities or functionings; 2. the selection of relevant capabilities to answer a specific question; 3. the weighting of capabilities chosen to give an overall assessment or aggregation of welfare).

4.2.3 Literature Survey (3): Chiappero-Martinetti & Roche (2009)

In the final review of empirical literature, the book chapter by Chiappero-Martinetti and Roche (2009) summarises the empirical challenges in applying the capability approach and discusses the studies they consider to be notable empirical advances within the capability field. They aim not to give a complete overview of all empirical applications within the field but:

> *"to discuss some basic principles and to review how the most consolidated applied literature dealt with these kinds of issues."* (Chiappero-Martinetti & Roche, p 160)

The review by these authors takes a more analytic approach to evaluating the literature than the previous two capability reviews. Drawing from the specification of capabilities as set out by Robeyns (see Section 4.2.2), Chiappero-Martinetti and Roche (2009) analyse different methods which have been used to select and/or combine attributes within a measure of capability. They look at the strengths and weaknesses of four statistical strategies for compiling and/or analysing outcomes of capability. They are:

1. *Scaling and ranking solutions*

Scaling and ranking solutions is the primary method of aggregation of indicators within the studies they reviewed. These indicators are not necessarily on the same scale as each other. The Human Development Index (HDI) was identified as the most commonly used example of this methodology within the capability literature. The HDI is the main measure of a country's progress used by the United Nations (UN) in their annual Human Development Reports (HDR). The HDI comprises three dimensions (health, education and income), measured across four indicators (life expectancy at birth, mean years of schooling, expected years of schooling, gross national income per capita at purchasing power parity). The HDI is averaged evenly across the three index dimensions and results in a HDI between 0-1 (lowest to highest human development) (Human Development Report, 2013).

2. *Fuzzy set theory*

Chiappero-Martinetti and Roche identified fuzzy set theory, a mathematical framework that allows for the study of non-binary, vague indicators (e.g. age described as "young", "middle aged" and "old"), as a method for combining categorical and continuous variables to measure the "degree of membership" of different indicators for an unobservable outcome (Chiappero-Martinetti & Roche, 2009). This, they argued, allowed for complex and vague indicators to be incorporated within an outcome to indicate capability when it may not be clear whether the level of attainment is a positive or negative result. However, it was also recognised that similar problems of aggregating across different membership functions or dimensions is also a problem. This fuzzy set approach is a popular methodology in poverty where defining someone as poor or non-poor may not be obvious by some levels (Chiappero-Martinetti & Roche, 2009).

3. Multivariate data reduction techniques

Multivariate data reduction techniques such as factor analysis and principal component analysis are used to identify key indicators of capability from a large amount of potential variables which could be included (Chiappero-Martinetti & Roche, 2009). This is conveyed in their review as a popular method due to the belief that capability itself is difficult to measure, so by treating capability or functioning as a latent variable, multivariate data reduction techniques allow for the measurement of unobservable variables from real data (e.g. Krishnakumar & Ballon, 2008).

4. Regression modelling

Regression modelling was shown to be implemented to predict a multidimensional outcome related to the capability approach. The outcome was predicted from variables thought to be most closely related to the capability concept of well-being. Chiappero-Martinetti and Roche identified a number of different types of regression models that have been implemented in capability literature, such as ordinary least squares (OLS), probit, logit and structural equation models (Chiappero-Martinetti & Roche, 2009). The choice of such model depends on the relationship between the dependent and independent variables. The difference between regression modelling and fuzzy set theory is that an observable variable of well-being is required for regression models, whilst fuzzy set theory assumes the dependent variable is unobservable (Chiappero-Martinetti & Roche, 2009).

The review chapter by Chiappero-Martinetti and Roche closes with a description of the empirical contributions to the capability approach which have used these statistical techniques. The description used the specification of a capability analysis as set out by Robeyns (2006) in Section 4.2.2 to compare and contrast the studies in their survey. Ranging from the earliest applications of empirical examples by Sen (Sen, 1985) to more recent econometric models of capabilities as latent variables (Krishnakumar & Ballon, 2008), they show the growth in sophistication of the statistical techniques used to analyse capabilities within the literature. Thirty two studies, over thirty years, are presented to show the numerous methodologies that can be explored within the capability theoretical framework. This review also documented an increasing number of empirical capability studies in the more recent past, suggesting a continued and growing interest in the application of the capability approach.

4.2.4 Summary of literature surveys on empirical capability research

The three reviews of the capability literature presented in this section indicate that a wide range of disciplines and methods for estimating capability across studies interested in operationalising the capability approach quantitatively. However, there are a number of questions left unanswered by this research, which suggests that a further, more systematic, review is required.

Firstly, of the three reviews, none offer a clear rationale as to how studies were included and excluded from each review. This could lead to concerns about the inclusion of the papers within each review, which may not have been subjected to any type of assessment. Given that the aim of two of the reviews was to improve the reporting methods for eliciting and measuring capabilities (Robeyns, 2006; Chiappero-Martinetti & Roche, 2009), it is surprising that they themselves did not transparently report their own review methodology.

Secondly, there appears to be an increase in the quantity of papers included in each review as time has progressed. This would suggest either a high growth of research within the capability approach since the turn of the century, or alternatively a greater knowledge of the most recent advances within the literature by the authors conducting their literature reviews. If it is assumed that the former is true, it is likely that new methods have emerged since the last review conducted by Chiappero-Martinetti and Roche (2009), which included literature up until the start of 2008. It would seem worthwhile in a new review to focus on the most recent applications within the capability approach since the last empirical review was conducted, albeit with a different objective, given the apparent growth in applications in the twenty-first century in the three literature reviews presented in this section.

Additionally, none of the three literature reviews contained detailed evidence relevant to the objectives and decision rules for which such outcomes are then used to inform decision-making, which is a key aim for this chapter. This will help understand how the capability approach has been interpreted as a measure for aiding decision-making and potentially offer guidance as to how capability instruments could be used within health economics.

4.3 METHODS

In this section, the methods used to conduct the review of capability outcomes and decision rules from a capability perspective are presented. First, the search strategy employed to identify relevant papers, known as a "comprehensive pearl growing" literature search strategy, is outlined. Inclusion and exclusion criteria are then presented. Second, the method of identifying studies through key pearls is then explained. Finally, a data extraction sheet is presented to show how data were extracted for the review requirements.

4.3.1 Search Strategy

The search strategy employed in this review is known as a "comprehensive pearl growing" literature search, which is a particularly useful search strategy for interdisciplinary topics (Schlosser et al., 2006). The process of pearl growing commences with the identification of key pearls (i.e., key studies), which can be identified from within the literature as being compatible with the aim of the review (Hartley et al., 1990). Once the key pearls have been identified, the first wave of pearls are produced, which are papers that have cited the key pearls within their reference list. Therefore, this type of literature review uses forward citations from the key pearls, unlike a more expansive forward-backward reference search strategy.

The comprehensive pearl growing literature search is of particular use when a keyword search strategy would return an unmanageable number of papers: for example, searching for terms related to economic evaluation such as "costs" or "benefits". Entering only these key terms within an interdisciplinary database is likely to return results in the tens of thousands which is an unmanageable number of papers to review. This is of particular concern for words which can have a number of different meanings within disciplines, or terms that have more everyday uses, such as ‘capability’.

There is precedent for using this methodology in the existing health economics literature. Examples within health economics include Dolan et al.’s (2005) review of people’s preference for QALY maximisation (Dolan et al., 2005b); Tsuchiya and Dolan (2005) review of people’s preferences for health states and health profiles over time (Tsuchiya & Dolan, 2005); and more recently, Stafinski and colleagues (2010 & 2011) use of the pearl growing method to qualitatively assess the different decision making processes for funding health technology across countries (Stafinski et al., 2010; Stafinski et al., 2011). A review of empirical applications of the capability approach seems equally well suited to this method.

The literature search was undertaken through the Institute of Scientific Information (ISI) Web of Knowledge citation search online facility. The ISI Web of Knowledge covers a number of databases including Web of Science (which covers the sciences, social sciences, arts and humanities) and MEDLINE (topics regarding biomedicine and health sciences), which made it an appropriate database for searching capability literature across a wide variety of disciplines.

Only papers published between January 1st 2006 and December 1st 2012 were included. The rationale for this choice was so that this review would focus on the most recent advances in the operationalisation of the capability approach, given that earlier studies were already captured through the reviews discussed in Section 4.2.

4.3.2 Inclusion criteria and Paper Categorisation

To be included in this review, studies identified needed to address at least one of these two main objectives of this review:

1. the aggregation of capability at an individual level and/or across populations, and/or
2. an objective or decision-rule as to how such outcomes could be then used to aid decision-making (e.g. cost per QALY threshold rule in health economics)

Based on the above criteria, titles and abstracts for the studies were sorted through keyword searching:

Keyword searching through title and abstract was structured as follows:

Capability OR Capabilities OR Functioning(s) OR Agency

AND

Measure OR Outcome OR Empirical OR Index OR Operationalisation

This review followed a two stage process of study categorisation. This follows from previous reviews which have used this categorisation process to identify the studies of most relevance to the research question at hand (Roberts et al., 2002).

4.3.2.1 Stage I – Initial Categorisation of studies

The studies identified using the previously outlined search strategy were then sorted into three categories based on the title and abstract.

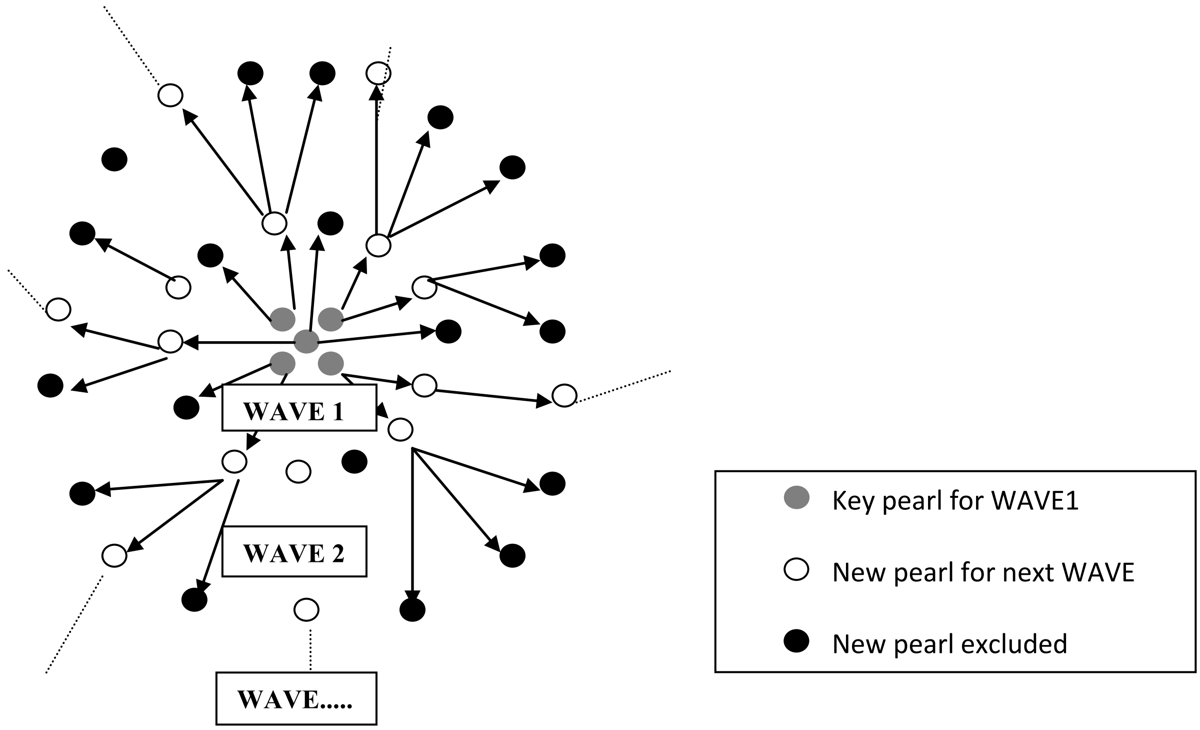
Category A, studies mentioned at least one from the two key search word lines (i.e. capability and operationalisation is Category A, capability only is not Category A).

Category B, studies that could be potentially relevant to the review, but require more information than the title and abstract alone. If the study contained the keywords capability or functioning or agency, but none of the other keywords, the study was examined for a quantitative results section, which could potentially indicate an attempt to measure capability outcomes. If any of the other key words were found in the title and abstract but did not include capability or functioning or agency, the reference list of the study was searched for citations of the key capability writings by either Amartya Sen or Martha Nussbaum, as a means of eliciting whether the study was based upon capability theory.

Category C, studies were excluded from the review. The studies either did not include any of the keywords for inclusion, or did not meet the criteria for Category B. Category C also included non-English publications. Conference abstracts and presentations were also excluded from this review, as it was not feasible to carry out a consistent quality assessment for this review, given the variety of topics expected to be covered within the review.

The studies identified in the first wave from the initial pearls (see Section 4.3.3) are then classified, according to Stage I categorisation. Therefore, the search terms are not used to initially to identify studies to review, but to classify studies once identified through the pearl growing method. Studies identified from the first wave which are classified as Category A or B are then employed to carry out a further wave search. Studies which have cited these new pearls are then categorised in the same manner as the first wave. This process of wave searching continues until a time where no new relevant studies are found. An illustration of the pearl growing method is presented in Figure 4.

Figure 4 An illustration of the Comprehensive Pearl Growing Method



4.3.2.2 Stage II – Further Categorisation of studies

Following the completion of the pearl search, studies categorised as either A or B were further classified after being read in full. Four categories helped to identify the final papers for inclusion.

1. Study uses capability related outcome and discussed decision rule
2. Study uses capability related outcome but does *not* discuss decision rule
3. Study discusses decision rule but does *not* use capability related outcome
4. Study does not use capability related outcome nor discuss decision rule

Papers which are classified within the first three categories are included in stage II classification in the final analysis of this review to help answer the three primary aims of the review. Papers classified as category 4 in stage II classification are excluded from further analysis.

4.3.3 Selection of Key Pearls

As a starting point for the identification of key pearls, from which the first wave of studies are identified within this review, the research already identified in the three previous reviews, presented in Section 4.2, was considered. The broadest disciplinary focus of the three previous reviews was the study by Robeyns (2006).

In Table 3, the nine areas (identified by Robeyns (2006)) where the capability approach has been applied are outlined. From each of Robeyns' nine groups, at least one study was chosen for the initial key pearl list to commence the literature review. One group (non-normative research, i.e. qualitative research) was excluded from the selection for the pearl list, as the studies were not relevant for this review focus of quantitative empirical applications. From the remaining eight groups in Table 3, nine studies from this review were included as the key pearls. The review by Robeyns (2006) was also included as a key pearl. Overall ten key pearls were included as detailed in Table 4.

Robeyns (2006) application groups (see Table 3) are used as a starting point for grouping and analysing all the studies included following the pearl search; where studies fell outside these groups, additional groups for new themes were generated where necessary.

Table 4 Key Pearl References for identifying studies in WAVE 1

Author	Study Title	Journal/Book	Year
Fukuda-Parr	The Human Development Paradigm	Feminist Economics	2003
Alkire	Valuing Freedoms	Book	2002
Ruggeri Laderchi et al.	Does it matter that we do not agree on the definition of poverty?	Oxford Development Studies	2003
Chiappero-Martinetti	A multidimensional assessment of well-being based on Sen's functioning theory	Rivesta Internazionale di Scienze Sociale	2000
Anand & Van Hees	Capabilities and achievements: an empirical study	Journal of Socio-Economics	2006
Zaidi & Burchardt	Comparing incomes when needs differ: equivilization for the extra costs of disability in the UK	Review of Income and Wealth	2005
Kuklys	Amartya Sen's capability approach: Theoretical insights and empirical applications	Book	2005
Robeyns	Sen's capability approach and gender inequality: selecting relevant capabilities	Feminist Economics	2003
Lewis & Giullari	The adult worker model family, gender equality and care: the search for new policy principles and the possibilities and problems of a capabilities approach	Economy and Society	2005
Robeyns	The capability approach in practice	The Journal of Political Philosophy	2006

4.3.4 Data extraction

Data were extracted from all included studies using a standardised data extraction form. The information used to inform data extraction is presented in Appendix 2. This follows protocol for data extraction from systematic reviews in health (CRD, 2009).

4.4 RESULTS

The summary statistics of the comprehensive pearl growing strategy, as well as the second stage review categorisation process are presented in this section.

4.4.1 Summary of Pearl Search

Overall, four waves of searching resulted in the identification of 494 unique studies. The findings are summarised in Figure 5. From Wave 1, 309 studies were flagged to have cited the key pearls. Forty-six (14.89%) of these papers were classified within Category A. From the remaining citations which addressed one of the two keyword searches, 25 of the 83 studies were classified as Category B on closer inspection. The other 58 studies were classified as Category C studies and excluded from the review. One hundred and eighty (58.25%) research studies were classified within Category C based on the title and abstract alone and were excluded from the review, with a total of 238 of 309 (77%) studies excluded in Wave 1.

From the 71 studies classified in Category A (46) or Category B (25) in the first wave, these studies pinpointed a further 182 unique references in wave 2. Twelve (6.60%) of these papers were classified in Category A based on abstract and title only, with 128 studies excluded as Category C at this stage. From the remaining 42 studies that were neither Category A or C on title and abstract only, 8 studies were classified as Category B on closer inspection and included in the review.

Figure 5 Summary Statistics of Comprehensive Pearl Growing Review Results

Key pearls: 10
(see Table 4)

WAVE 1

309 unique studies found
71 new pearls (Cat A/B)

WAVE 2

182 unique studies found
20 new pearls (Cat A/B)

WAVE 3

3 unique studies found
1 new pearl (Cat A/B)

WAVE 4

No new pearls found
494 studies identified in total
92 studies identified as relevant

From the 20 studies of relevance found in WAVE 2 which were used as pearls for a third wave of searching, three new studies were identified (WAVE 3), with no paper meeting Category A or C on title and abstract only. One of the three studies was classified in Category B on closer analysis with two being included in Category C. This sole Category B study was used to run the fourth wave of literature searches, but no further studies were identified which fell within the timeframe of this search. Therefore, this signalled the conclusion of the search.

4.4.2 Final classification of studies

Upon Stage II classification of the remaining studies, almost half of the studies (45/92) were found to address the two key aims of this review (Category 1). 26 addressed one of the aims only (Category 2 or 3). Twenty studies did not measure capabilities directly or discuss a decision rule or objective for such a measurement, so these studies were excluded from further analysis (Category 4). Appendix 3 presents the 72 papers included in this review. Appendix 4 lists the papers excluded at the stage II classification.

4.5 ANALYSIS OF REVIEW

This section begins with summary data on the studies included for full review. This is followed by a description of the primary focus of studies within the review and whether this reflects findings from the previous review which are presented in Section 4.2. Section 4.5.3 then looks at the type of attributes included in the studies which developed or employed capability related measures. Section 4.5.4 explores whether there are any themes within the review for aggregating such measures to form an index. Finally, the objectives and decision rules used within health economics are compared and contrasted with the rules applied for capability studies more broadly in Section 4.5.5. Section 4.5.6 addresses a number of outstanding issues related to capability application.

4.5.1 Summary Data on studies included in review

Figure 6 shows summary information about the locations of where the studies were based on. The highest proportion of studies took place within the UK and the rest of Europe (26 studies out of 72 or 36%).

Figure 6 Population where the capability studies identified were applied

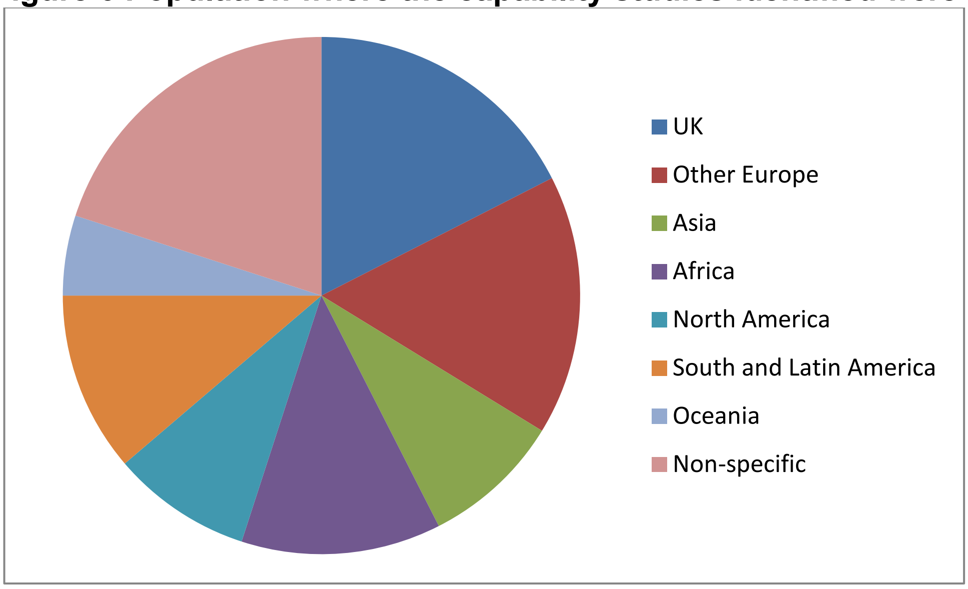
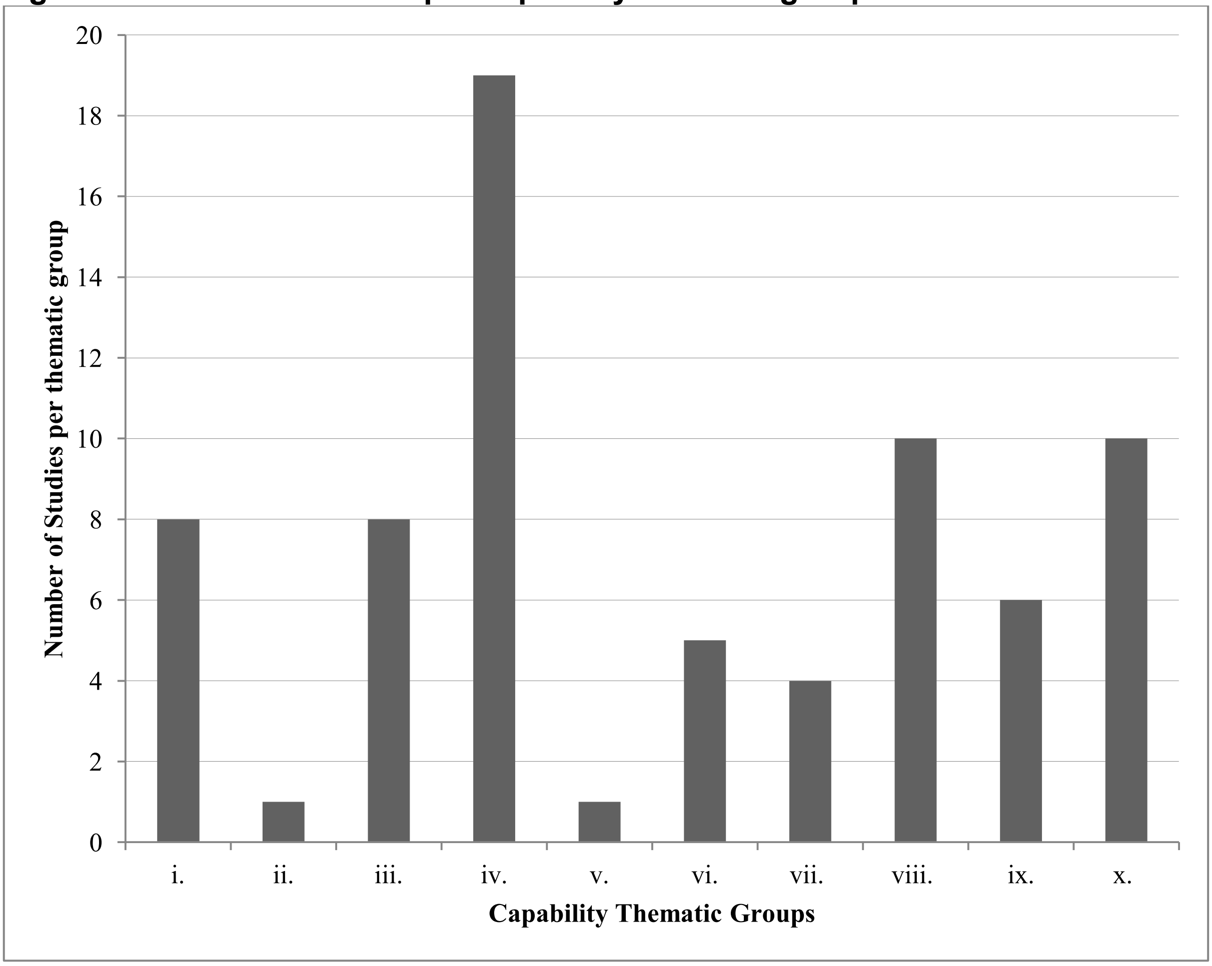


Figure 7 shows the spread of studies across seven capability thematic groups identified by Robeyns' (2006) (i.e. group i.-group vii.) and three new themes which emerged from this review (i.e. group viii.-group x.). Group iv. (assessing poverty and well-being assessment in advanced economies) has the highest proportion of studies identified out of the 10 groups with 19 studies. The three new groups of education (group viii), technology (group ix.) and health (group x.) account for 26 of the 72 studies identified, showing a growing interest in capability applications in these three groups in particular.

Figure 7 Number of studies per capability thematic group



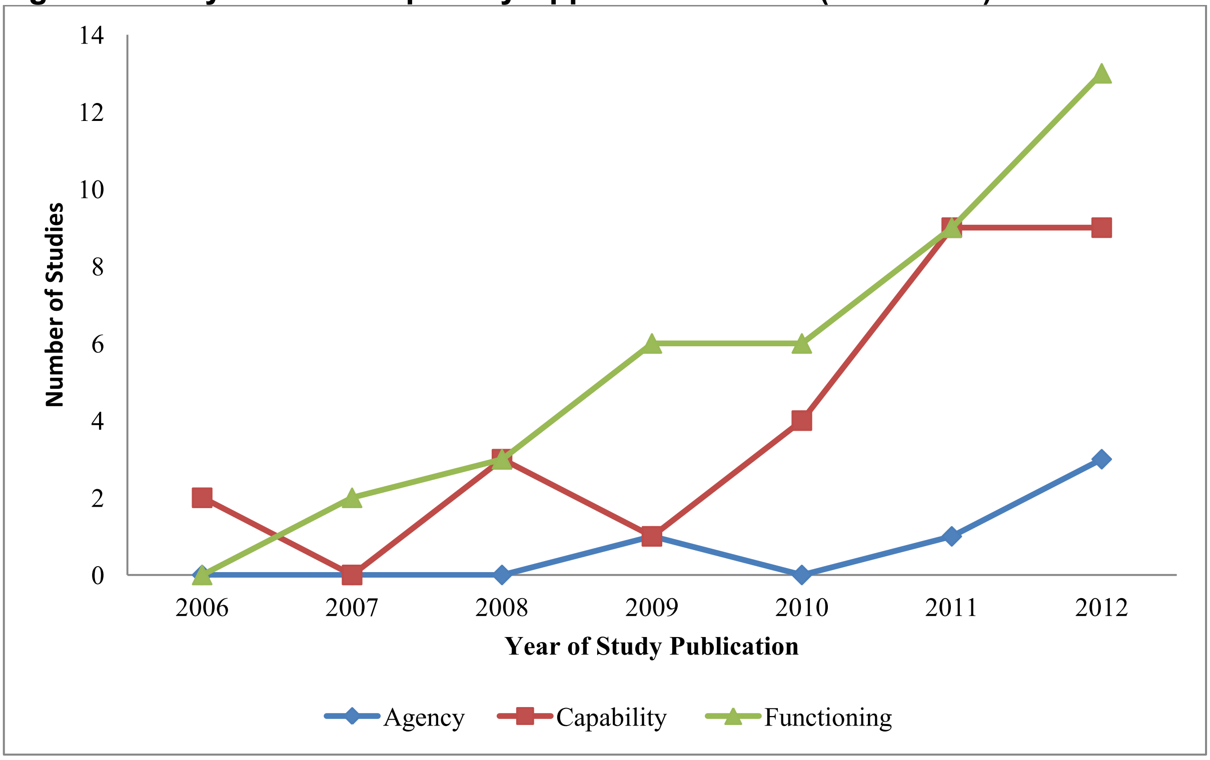
Thematic groups: i. General assessment of human development; ii. Assessing small scale development projects; iii. Identifying the poor in developing countries; iv. Poverty and well-being assessment in advanced economics; v. Deprivation of disabled people; vi. assessing gender inequalities; vii. Debating policies; viii. Education; ix. Technology; x. Health

4.5.2 Focus of study: Capability, Functioning or Agency?

In Figure 8, the focus of the studies included in the review, in terms of their main interest in measuring capabilities, functionings or agency (as defined in Chapter 3), is outlined. Where studies had more than one focus (e.g. capability and agency), the predominant focus of the study was chosen as assessed by this author.

Out of the 72 studies included in the final review, 38 studies had an applied focus on functionings, 29 studies on capability and 5 papers on agency. While the results of this review would confirm the findings of previous reviews of capability applications which are mainly concentrated on functionings (Kuklys & Robeyns, 2005; Robeyns, 2006; Chiappero-Martinetti & Roche, 2009), from Figure 8 there appears to be an increasing interest in the applied focus on capability measures in more recent times.

Figure 8 Study focus of Capability Application review (2006-2012)



4.5.3 Attributes included in capability related measures

From the 72 studies included in the review at this stage, 61 (85%) of them developed or applied an instrument of some description related to the capability approach. Each thematic group is discussed in terms of capability attribute inclusion. The attributes for each study are detailed in Appendix 3.

4.5.3.1 Thematic Group i. General assessment of human development

Eight studies implemented attributes assessing human development (group i.). Distaso developed a multidimensional sustainability well-being index across 10 dimensions and applied the measure within Italy (Distaso, 2007). Gardoni and Murphy (2010) developed a Disasters Impact Index (DII) to measure the consequences of natural disasters upon people's capabilities (Gardoni & Murphy, 2010). A welfare index was developed to measure capabilities within Argentina across 10 dimensions (Anand et al., 2011), although this index was not generated from Nussbaum's list of 10 central human capabilities like previous work from Anand and colleagues (2009) with the OCAP, as described in Chapter 3. Notten & Roelen (2012) collected thirteen indicators of deprivation for four dimensions (housing, neighbourhood, basic services, financial resources) to assess the most appropriate method of measuring childhood deprivation across the European Union (Notten & Roelen, 2012). Nussbaumer (2012) developed five dimensions (across six indicators) to assess energy poverty in Africa (cooking, lighting, services provided by household appliances, entertainment/education, communication). Alkire & Foster (2011a) assessed multidimensional poverty across four attributes in the United States (health status, health insurance, education and income) and across three dimensions in Indonesia (expenditure, health in terms of body mass index and years of schooling). Krishnakumar & Ballon (2008) predicted two basic capabilities (knowledge and living conditions) from six indicators of capability in Bolivia. Nguefack-Tsague et al. (2011) considered the methods for valuing the three attributes on the HDI. This is discussed at greater length in the section on aggregating attributes (Section 4.5.4). All but one study (Anand et al., 2011) in this thematic group developed attributes with a functionings assessment of well-being. Anand and colleagues (2011) developed a questionnaire to capture capabilities directly.

4.5.3.2 Thematic Group ii. Assessing small scale development projects

Only one study was identified for assessing small scale developments (group 2). Peris and colleagues (2012) examined the role of El-Almanario, a small grants development programme (SGP) tool aimed at enhancing community empowerment in a rural town in Guatemala (Peris et al., 2012). Community residents are asked to fill out a questionnaire, where they were asked to identify the problems within the communities and then decide which of these problems should take priority in being addressed by the community (Peris et al., 2012).

4.5.3.3 Thematic Group iii. Identifying the poor in developing countries

Seven studies were found for the group concerned with identifying the poor in developing countries (group iii.). The majority of these studies included attributes based on context specific poverty to the country of origin, such as a child poverty measure developed specifically for Vietnam (Roelen et al., 2010), well-being in Thai provinces (Parks, 2012), poverty in a Brazilian shantytown (Kerstenetzky & Santos, 2009), core poverty in South Africa (Clark & Qizilbash, 2008), or protecting capability in Chile (Barrientos, 2010). Other studies have focused more on the methodology for identifying the poor, in relation to measuring child poverty (Arndt et al., 2012), and defining poverty based on adequate shelter, which is broadly defined as somewhere that an individual can address their basic needs (hygiene and rest) and are able to store their personal belongings securely (Norcia et al., 2012). All seven of these studies were based on capturing functioning, with health and housing among the main attributes used to distinguish the poor from the non-poor.

4.5.3.4 Thematic Group iv. Poverty and well-being assessment in advanced economies

Assessing well-being and poverty in advanced economies (group iv.) had the highest number of studies out of the ten thematic groups (13 studies altogether). Wagle produced two studies which looked at indicators of poverty in the United States, focusing on education, income and type of employment (Wagle, 2008; Wagle, 2009). Capability indicators are a dominant theme within this category. Jordan and colleagues (2010) explored Cape York Institute for Policy and Leadership 'capability indicators' which included eleven dimensions to assess differences between indigenous and non-indigenous Australians (Jordan et al., 2010). Anand's OCAP survey of capability indicators, developed in the UK, was also identified but was already discussed in detail in the previous chapter (Anand et al., 2009). Another UK study focused on "basic functionings" to achieve what Binder and Broekel refer to as "conversion efficiency" (Binder & Broekel, 2011). Two studies looked at whether government measures of human rights (Burchardt & Vizard, 2011) and poverty captured in German wealth reports (Arndt & Volkert, 2011) are compatible indicators with the capability approach. Van Ootegem and Verhofstadt developed a questionnaire of capabilities, which was tested on University students in Belgium across seven domains that were thought to be important for University students (Van Ootegem & Verhofstadt, 2012). Researchers in Australia developed a freedom poverty measure across three dimensions (income, health and education) and have suggested that it can be used as an alternative measure of poverty in Australia (Callander et al., 2012b), to assess regional differences in poverty in Australia (Callander et al., 2012a) and as a broader assessment of child well-being and their ability to work once finished schooling (Callander et al., 2012c). The same three dimensions of education, health and income were used to assess regional development within the UK (Perrons, 2012). Lastly, four dimensions (education/leisure, health, social participation, income poverty) across eight indicators were used to assess childhood deprivation in Germany (Wust & Volkert, 2012). The study by Van Ootegem and Verhofstadt (2012) was the only study to develop capability attributes, with the other 12 studies based on functionings achieved as indicators of capability.

4.5.3.5 Thematic Group v. Deprivation of disabled people

The deprivation of disabled people (group 5) consisted of one study. Rosano and colleagues (2009) applied Kuklys' (2005) equivalence scales for disabled people within Italy and found that disabled Italians require twice as much income to achieve the same functioning as the fully able (Rosano et al., 2009). In Kuklys original study, the disabled in the UK required 50% more income on average to reach equivalent functioning levels (Kuklys, 2005).

4.5.3.6 Thematic Group vi. Assessing gender inequalities

Five unique measures were developed for assessing gender inequality (Group vi.). Anand and Santos (2007) used the OCAP survey to explore three capabilities related to violent crime and the propensity of females to fare worse in terms of their fear of, and actual experiences of sexual assault, domestic violence and violent assaults (Anand & Santos, 2007). Addabbo and colleagues (2010) developed a gender budgeting approach as a method to audit the impact across genders from new government policies (Addabbo et al., 2010). Bérenger and Verdier-Chouchane developed a Relative Womens Disadvantage index to compare women's disadvantage in terms of health, education and labour participation internationally (Bérenger & Verdier–Chouchane, 2011). To assess the quality of life of home based workers in Thailand, three dimensions based on income, work intensity and education, are employed to assess the time-use dimensions of wellbeing (Floro & Pichetpongsa, 2010). Finally, Di Tommaso and colleagues (2009) focused on two of Nussbaum's two central human capabilities (bodily health and bodily integrity) to explore the well-being of women trafficked for sexual exploitation in the Balkans and Eastern Europe (Di Tommaso et al., 2009). All studies in this category focused on functionings.

4.5.3.7 Thematic Group vii. Debating policies

The debating policies group (group vii.) had four studies. Two of the papers are in contrast with the gender inequality group (group vi.) by focusing primarily on the male gender disadvantage when it comes to achieving work-life balance. Within these studies, Hobson and colleagues assess differences across Europe by drawing from European data on parents' work life balance and whether parents felt they had control in achieving a good WLB (Hobson & Fahlén, 2009; Hobson et al., 2011). Renouard argues why "relational capability" within organisations is important for achieving corporate social responsibility by identifying four key components (network integration, individual commitment to work within groups, having attachment with others, valuing others objectives simultaneously) (Renouard, 2011). Reitinger and colleagues (2011) debate the merits of conceptualising Area of Protection over consumption and production processes within a given society for analysing environmental life cycle assessment through a capability lens. They identify eight key dimensions (Reitinger et al., 2011). Two of these studies focuses on capability (Renouard, 2011; Reitinger et al., 2011), one on functioning (Hobson & Fahlén, 2009) and one on agency (Hobson et al., 2011).

4.5.3.8 Thematic Group viii. Education

The group focusing on education (group viii.) had seven studies which developed alternative measures. The majority of these studies focused on expanding capabilities within Africa ranging from basic capabilities in the quality of education (Smith & Barrett, 2011; Tikly & Barrett, 2011), and a more general education capabilities list (Walker, 2006), to a Public-Good Professional education index, which aimed to promote the higher education of those individuals who would expand societal capabilities (McLean & Walker, 2012). A list of functional capabilities was also developed for higher education (Walker, 2008), while in the US, the capabilities gained from teaching arts in schools was captured by the Arts Education Pathway Model (Maguire et al., 2012). A list of five basic capabilities from learning was also developed (Young, 2009). The majority (five) of these studies focused on capability attributes, with only two (McLean & Walker, 2012; Smith & Barrett, 2011) based on functioning attributes.

4.5.3.9 Thematic Group ix. Technology

Six studies were identified on the topic of technology (group ix.). Three of these studies show the development of the Choice Framework, primarily by Kleine (Kleine, 2010; Kleine, 2011; Kleine et al., 2012). The Choice Framework was used to assess issues such as the role of telecentres (i.e. free public access to computers and the internet) in Chile and Fair Tracing of products in the UK (i.e. allowing customers and producers to digitally follow the supply chain of goods), which helped to show the improved benefits gained from individual capability through technological advances. The role of information and communications technology for development (ICT4D) was also explored in the other three studies in this category, with the assessment of the role of internet use (Hatakka & Lagsten, 2011), technology allowing for improvement of instrumental freedoms in Uganda (Kivunike et al., 2011), and in Cambodia through a Capability, Empowerment and sustainability virtuous spiral model (Grunfeld et al., 2011). Three studies included capability attributes (Kleine, 2010; Kleine, 2011; Grunfeld et al., 2011), two studies on functioning attributes (Kivunike et al., 2011; Hatakka & Lagsten, 2011) and one study on agency (Kleine et al., 2012).

4.5.3.10 Thematic Group x. Health

The health group (group x.) had nine studies which included capability attributes. Four of these studies relate to two of the three types of capability questionnaires already discussed in detail in Chapter 3 (Grewal et al., 2006; Coast et al., 2008a; Al-Janabi et al., 2012a; Netten et al., 2012). Three of the studies are focused within the United States, with attention paid to chronic conditions (Ferrer & Carrasco, 2010), as well as the ability of the built urban environment to give more capability for people to participate in healthy activities and reduce obesity levels (Lewis, 2012a; Lewis, 2012b). The remaining two studies were from Africa, with both studies interested in the control and empowerment of women's decision making in accessing healthcare in Burkina Faso (Nikiema et al., 2012) and reducing shortfalls in health functioning in Ethiopia (Mabsout, 2011). While the role of health is an important indicator of capability within most of the other thematic groups in this review, an indicator of health was included in the minority of studies within this section, indicating that focusing on health alone is not adequate from a capability perspective.

4.5.4 Aggregating capability related attributes within an index

The majority of studies discussed in the previous section could not be easily grouped within a pre-existing index, with the majority of such studies developing new outcomes to address particular problems with capability within a specific context or country. However, some of the studies identified did show trends in the development of an index once the attributes of the instrument had been decided. In this section, the two major methodologies that are developed for compiling capability indicators into an index are detailed.

4.5.4.1 Alkire and Foster Multidimensional Poverty Methods

The primary methodological contribution to index development found in this review appears to be the work of Alkire and Foster on multidimensional poverty (Alkire & Foster, 2011a). The Alkire Foster (AF) measures propose a methodology for measuring the multi-faceted nature of the social determinants of poverty. Alkire and Foster (2011a) do not focus on a single indicator of poverty such as income poverty. They show in their seminal paper on multidimensional poverty measurement how such a focus can be misleading in describing the true levels of poverty within a given society. Instead, the AF measures allow other factors to be considered in determining whether or not poverty is present.

Alkire and Foster (2011a) follow the axiomatic approach to measuring multidimensional poverty as detailed by Sen (Sen, 1976). The AF measures focus on two key issues of poverty measurement. Firstly the “identification method”, which considers how an individual is identified as being poor or not poor and secondly, the “aggregation method” which follows from the identification step of determining who is poor, by defining the appropriate measure of the deprivation suffered by the poor (Alkire & Foster, 2011a).

Within a unidimensional poverty measure, there is a relatively simple process involved in defining whether a person is in poverty. It is a case of determining the threshold on that one dimension below which a person is considered to be in poverty. For an example, the World Bank defines "extreme poverty" as individuals living on less than $1.25 a day, adjusted for purchasing power parity (PPP), with a higher level of $2 for middle income economies (The World Bank, 2010).

The "identification method" becomes more complex in multidimensional poverty measurement. Here, the classification of an individual as poor additionally requires a decision about the number of dimensions in which a person has to fall below the threshold to be classified in this way. Atkinson (2003) outlines two common identification approaches in poverty assessment. Firstly, the "union approach", whereby a person is classified as poor if they fall below the threshold on any dimension. Secondly, the "intersection approach", whereby a person is poor only if they fall below the threshold of all included dimensions (Atkinson, 2003).

Alkire and Foster (2011a) found flaws with both identification methods and developed an alternative "in between" method for AF measures, referred to as the "dual cutoff" method. The dual cutoff method operates by first identifying a cutoff for each dimension below which a person is classed as deprived in that dimension, and second determining the number of dimensions in which an individual must be deprived to be classified as poor. These cutoffs can vary with context, enabling flexibility for specific purposes, whether it be a cross-national comparison of multidimensional poverty or a more refined policy question (Alkire & Foster, 2011b).

The "aggregation method" relies on a "censoring" step, whereby those who do not meet the criteria for poverty (i.e. individuals not deprived in the required number of dimensions) are censored from the poverty measurement exercise. The aggregation process then relies on a number of different methods, depending on the complexity of the poverty measurement required. These methods are captured in four different AF measures of multidimensional poverty, which consider not only whether or not a person is poor, but also on how many dimensions they are poor in (accounting for the breadth of poverty over dimensions considered), how far away an individual is from the threshold on each dimension in which they are deprived (accounting for poverty depth within a dimension) and whether different weights across dimensions are attached to the same levels below the poverty thresholds on dimensions (accounting for the severity of poverty across dimensions) (Alkire & Foster, 2011a).

The four AF measures are outlined below:

1. Headcount ratio (H): p/P
   (where p = population who are poor; P = total population under consideration)
   This measure is essentially a counting exercise whereby the proportion of poor individuals are measured as a percentage of the total population under consideration.

2. Adjusted headcount ratio (M_0) = Headcount ratio*(d_p/D)
(where d_p = dimensions which individual is poor, D = number of dimensions for poor population, p, any individual can be poor in)
This measure shows on how many dimensions a person is classified as poor, with a proportionally higher deprivation experienced by people deprived in more dimensions.

3. Adjusted poverty gap (M_1) = M0*(l_p/L)
(where l_p = levels below poverty threshold where poor population, p responded, L = total levels below threshold across dimensions for poor population, p)
This measure takes into consideration the breadth of deprivation amongst the poor, as well as the depth of poverty suffered, which accounts for higher poverty levels the further an individual is from the cutoff threshold.

4. Adjusted Foster-Greer-Thorbecke(FGT) (Foster et al., 1984) measure (M_2) = M0*(v_p/V)
(where v_p = values attached to levels below threshold where poor population, p responded, V = total value between lowest levels on all dimensions and threshold for poor population, p)
The most sophisticated AF measure, allowing values to be attached to different response levels on dimensions. It allows for the measurement of the frequency, breadth, depth and severity of deprivation suffered by the poor.

Alkire and Foster (2011a) illustrated these AF measures in the United States, by measuring multidimensional poverty between three ethnic groups (African-American, Hispanic and White) across four dimensions (income, health status, health insurance, education level), with a cutoff of deprivation in two dimensions for a person to be considered poor. They found that whilst the African-American population in their sample was the most impoverished when analysing income only (i.e. using the Headcount ratio H), using their dual cutoff approach for the M0 measure considered the Hispanic population group to be most deprived overall, as the health insurance and education attainment indicators were considerably worse for Hispanics (Alkire & Foster, 2011a). This methodology has been adopted by the United Nations and since 2010, it reports a Multidimensional Poverty Index (MPI) annually in its Human Development Reports (HDR) (Klugman, 2010).

Eight studies in this literature review, including the methodology paper by Alkire and Foster (2011a), are connected by the use of this multidimensional approach to measuring capability. Three of the studies are related to developing a Freedom Poverty Measure in Australia, which looked at accounting for three dimensions of poverty (income, education and health) (Callander et al., 2012b), including measuring freedom between Australian regions (Callander et al., 2012a) and for assessing childhood development (Callander et al., 2012c). Two further studies also focused on measuring childhood deprivation: one study in Germany (Wust & Volkert, 2012) and another study in Vietnam (Roelen et al., 2010) using the AF methodology. One other study focused on measuring childhood deprivation but this time by comparing deprivation across four European countries (Notten & Roelen, 2012). Finally, Nussbaumer and colleagues (2012) developed a multidimensional energy poverty index (MEPI) and applied it across the African continent (Nussbaumer et al., 2012).

4.5.4.2 Human Development Index and related measures

The Human Development Index (HDI) is the traditional outcome measure that is associated with the United Nations Development Programme (UNDP) and the capability approach. Since the UNDP launched the annual Human Development Reports (HDR), the HDI has been included for cross-country comparison of the “real wealth of nations” (ul Haq, 1990). The construction of the HDI index was explained previously in Section 4.2.3.

Only four studies within this literature review used aggregation based on the HDI methodology. Nguefack-Tsague and colleagues (2011) provide a statistical justification for the equal weighting of the three dimensions captured within the HDI (Nguefack-Tsague et al., 2011). Perrons (2012) used the three dimensions of HDI to develop a regional development index (RDI) to assess the development of regions within the UK. Comparing RDI with GDP showed that London was not as developed as other UK regions when the additional dimensions were accounted for in the RDI (Perrons, 2012). Two other measures were developed outside of the UK. One study developed an individual well-being index (WBI) across three dimensions (personal income index, work intensity index, education attainment index) to measure the well-being of home-based workers in Thailand (Floro & Pichetpongsa, 2010). The other study developed capability indices (CI) to assess two basic capabilities (knowledge and living conditions) across a number of indicators to compare different regions in Bolivia (Krishnakumar & Ballon, 2008).

4.5.5 Exploring the objectives and decision rules within the capability approach

In this section, the third aim of this review is explored in detail. Firstly, the objectives and decision rules within health specifically that were found in this review are examined. Within the second sub-section (Section 4.5.5.2), objectives and decision rules outside of health are analysed to see whether compatible or alternative decision rules are being used and justified based upon capability theory. Given the breadth of studies and objectives across disciplines which are included within this review the form of a narrative review was considered the most suitable to explore this aim.

4.5.5.1 Is the QALY objective (health maximisation over time) compatible with the capability approach?

Ten out of the 72 studies identified had a specific focus on health (group x.). There are different interpretations of how the capability approach can be used within an evaluation framework for health based on these studies. The work based upon the ICECAP indices, which were discussed in detail in Chapter 3, suggests that one possibility for the capability approach is to measure a concept broader than health maximisation:

> *"One option would be to use such an index essentially as a QALY replacement – as a means of retaining a single measure of outcome for such evaluations, but with that measure covering more than just health."* (Coast et al., 2008a, p. 882)

Others have argued that while the QALY objective can be useful in certain circumstances, it is not a useful tool for assessing interventions where health is not a relevant outcome:

> *"It is possible to view healthcare interventions as being on a spectrum between those that purely maximise health (e.g. use of insulin to control diabetes) and those that purely maximise empowerment (e.g. counselling interventions to improve ability to make informed health decisions in people living with diabetes). However, it is possible to conceive that the latter, if done well, could contribute significantly to the former, and indeed this is the basis of most self-management interventions"* (McAllister et al., 2012 p. 6)

Accounting for empowerment within an evaluation framework applies as much in developed countries as it does in developing countries, as emphasised by another study which looked at how women could overcome hurdles in accessing healthcare in Burkina Faso (Nikiema et al., 2012).

However, there is no overall consensus on the separation of the QALY objective from the capability approach. Netten and colleagues (Netten et al., 2012) argue that the measure of social care ASCOT, which they link to Sen's notion of capabilities, can be framed within the QALY framework in what they refer to as a social care related quality of life (SCRQOL) measure:

> *"The revised ASCOT measure developed through the study shows considerable potential, providing a first estimate of a social care equivalent to the QALY, and can be used in a range of circumstances, including cost-effectiveness, cost-utility and policy evaluation."* (Netten et al., 2012, p. 95)

An alternative to separating QALYs and capabilities would be to measure health alongside a number of indicators of well-being. Such an approach has been developed in Australia, where a "freedom poverty" measure has been conceptualised with indicators of health, education and income to assess whether a person is deprived (Callander et al., 2012b). However, Callander and colleagues (2012b) have yet to define what would be considered good or bad health and what measure should be used to calculate good or bad health. Therefore, further research is required in this area.

Barring the example of Netten and colleagues (2012), who used the time-trade off technique to generate a social care QALY (see Chapter 3 for more), there has been little emphasis on incorporating time alongside capability or empowerment measures, which has, to date, been an important part of health economic evaluations. The ICECAP indices provide the most developed attempt to capture capability directly. However, there is no guidance as to how such indices could be used to aid decision-making. The next section assesses information from studies outside of health in the literature review to see if there is any other guidance available on the overall objective of an evaluation based upon capability theory.

4.5.5.2 Capability Objectives and Decision Rules beyond health

While the capability approach was developed as an alternative to the traditional utilitarian approach in welfare economic assessment, there are some who claim that capabilities can fall within a similar maximisation framework. One example of this is Renouard (2011), whose study suggests that corporate social responsibility within private enterprise should account for what they term as "relational capability", drawing upon research within anthropology and Sen and Nussbaum's research, to look beyond utility maximisation of company stakeholders but rather achieve the:

> *"maximisation of the relational capability of people impacted by the activities of companies"* (Renouard 2011, p. 85)

This concept of maximising an absolute level of capabilities is not limited to the above example, with Tikly and Barrett (2011) stating that the capability approach of "maximising choice" is a more appropriate assessment of welfare than the standard rational choice theory of economics within education of low income countries:

> *"Here the assumption is that individuals act on the basis of the maximisation of their own utility and that efficiency within the public welfare system is best served through maximising 'choice'"* (Tikly & Barrett, 2011, p. 8)

However, the objective of maximising capabilities in some form or another as an absolute aim is not a reflection of the majority of work related to the papers found in this review. As an example of this, Anand et al. (2009) states:

> *"they (people) do not wish to maximize total social welfare for a variety of reasons, not least of which is that they are concerned about distributional issues too"* (Anand et al. 2009, p. 127)

Many papers focus on the maximisation of something less than optimum levels as a priority, such as the maximisation of basic capabilities (Krishnakumar & Ballon, 2008) or by measuring poverty as:

> *"insufficiency in basic capabilities"* (Kerstenetzky & Santos, p. 189).

Other conceptualisations of the capability approach have developed within more advanced economies. Binder and Brockel (2011) develop their concept of "conversion efficiency" as an alternative to traditional well-being assessment:

> *"The idea of relative efficiency means we are evaluating individuals' efficiency not with a theoretically derived maximum, but to the maximum of functioning achievement observed in the data given a certain level of resources"* (Binder & Brockel 2011, p. 261)

Binder and Brockel demonstrated their measure within Great Britain and showed that conversion efficiency is improved within this sample, by age, self-employment, marriage, no health problems and living in London and the surrounding boroughs (Binder & Broekel, 2011). Murphy and Gardoni (2010) developed a two-stage process for assessing individual capability within a risk analysis, such that:

> *"for defined groups, the goal should be to maximise variability of non-basic capabilities and minimise variability within sub-vectors of basic capabilities and among defined groups of those with similar boundary conditions"* (Murphy & Gardoni 2010, p. 145)

Another alternative to average welfare maximisation in a narrow space comes from the field of education. Callander et al. (2012c) argue that increasing educational opportunities for youths is not an adequate pre-requisite to future labour force participation (Callander et al., 2012c). Instead they develop a measure drawn from the multidimensional poverty literature (Alkire & Foster, 2011a) to assess health alongside education, which they argue is also likely to have an impact of the probability of labour force participation in the future:

> *"efforts to increase children's future labour participation rates as a means of improving their living standards should also focus on improving childhood health, as well as education."* (Callander et al. 2012c, p.179)

4.5.6 Outstanding issues

The quantitative studies analysed in this review most commonly apply the capability approach to compare capabilities across populations. The majority of population comparisons deal with cross-sectional data only, with little attention paid to how capability is affected over time within population groups. The reference case of current health economic evaluations for groups like NICE recommends accounting for changes in quality over time within outcomes (NICE, 2013), so this is an important consideration for using the capability approach in health evaluations.

One notable attempt identified in this review to measure well-being over time was carried out by Clark and Hulme (2010). The aim of their study was to bring together two separate parts of poverty measurement, by measuring core poverty (at a point in time) and chronic poverty (over time) simultaneously (Clark & Hulme, 2010). In this study they extend the vagueness poverty framework (Qizilbash, 2003), which designates people within three categories based upon their reported well-being in a given dimension (core poor, vulnerable and non-poor)

(Clark & Qizilbash, 2008). When combined with the research within measuring prolonged or "chronic poverty" research (Hulme & McKay, 2008), even more definitions are introduced when measuring poverty over time. Such definitions include transitory vulnerable people (at risk of poverty in future or just coming out of a definite poverty state of being), transitory core poor (people who go in and out of poverty over time) and chronically core poor (i.e. persistently poor over time) (Clark & Hulme, 2010).

While this research is a welcome improvement in addressing capability longitudinally, it remains less clear of the value of different states of poverty within Clark and Hulme's (2010) approach (e.g. should all resources be channelled at those in core poverty before other levels of poverty should be addressed?). An alternative study states that their "fuzzy" approach to defining different poverty levels could be used longitudinally (Betti & Verma, 2008). However, the question of which poverty states (or capability states) are prioritised over another remains to be addressed before this work can be applied within an evaluation framework.

From this review, there does not appear to be a method for combining a measure of capability with the cost of an intervention, even though studies have developed outcomes as alternatives to measuring benefits monetarily in a cost-benefit analysis (Beyazit, 2010; Gardoni & Murphy, 2010). It is unlikely, then, that health economics can learn much about developing a comparable cost-effectiveness ratio for capability outcomes from the capability literature.

4.6 DISCUSSION

The analysis of the comprehensive pearl review has shown that research in a wide variety of disciplines has attempted to measure and apply the capability approach quantitatively. This chapter reports the most recent advances within the applied capability literature. It specifically explores the application of the capability in health, and whether applications in other disciplines are appropriate when evaluating health interventions too. The areas of “Health”, “Education” and “Technology” are among the highest growing disciplines where the capability approach is being most frequently applied. This resulted in three themes being added to Robeyns’ (2006) initial thematic groups developed in 2006. Additionally, previous reviews have shown an overwhelming majority of studies have focused on functioning based attributes. While functioning attributes remain the most prevalent method in this review too, Figure 8 suggests that this review has identified an increasing trend of using capability based attributes.

The primary results show that there is no overall consensus as to how the capability approach should be applied either in health or elsewhere to aid decision-making. The majority of studies in this review suggest that an objective other than the maximisation of health when applying the capability approach is likely to be more appropriate for an application which corresponds with the underlying theory. Rather, it seems that a more consistent theme through the majority of the studies reviewed is the idea of achieving “basic capabilities”(Young, 2009) or “minimum level of capabilities attainment” (Murphy & Gardoni, 2008). This threshold approach has also been referred to within regions as a “sufficiency economy” (Parks, 2012) or within adult literacy as a “sufficient” level of learning (Maddox & Esposito, 2011). Within the majority of cases, this review highlights these concepts as the primary quantitative methods for applying the capability approach. This is reflected in the main way attributes were aggregated within an index through the Alkire and Foster methods of multidimensional poverty, with the purpose of allowing multiple determinants of “unfreedom” to be calculated simultaneously (Alkire & Foster, 2011a).

Outcomes within the capability approach explored in this review offer little as to how resource allocation could be informed by such indices. The primary objective of studies was an attempt to highlight areas where inequalities or poor performance in specific contexts can be improved, with little guidance as to how such improvement should be made. This is not useful within a health economic evaluation context, where choices between alternative interventions are required. Additionally, capability over time is a sparsely explored topic, so a method for the combination of a measure of capability with time cannot be drawn from studies found within this review.

The quality of the studies included within the final review was not assessed given the wide variety of areas of application and the desire to explore possible methods. Papers were chosen in a defined period of time based on a defined number of relevant original pearl cites; although the entirety of the literature is not reported, earlier capability applications were considered through the summary of the three reviews prior to the original pearl search undertaken here.

The strengths of this literature review in comparison to previous capability literature reviews are that the search strategy methodology was established and made explicit, whilst previous reviews of applied capability work have been less transparent. Given the growth of areas where the capability approach is applied, a less systematic approach is likely to be subject to bias arising from researcher disciplinary knowledge and is likely to miss new research developments in different contexts.

There are a number of limitations associated with this review. The selection of key pearls from a previous review of themes in the capability approach was an attempt to include a wide variety of studies from different disciplines. However, the selection of those key pearls may not have been cited by all the most recent applications of the quantitative application of the capability approach. Therefore, not all relevant papers may have been included in this review. This is the trade-off associated with using the comprehensive pearl growing literature search strategy to identify a manageable amount of studies to review.

This review suggests that those wishing to apply the capability approach should take into account objectives other than maximisation of an ideal state, for example health maximisation. The prioritisation of basic capabilities or a threshold of sufficient capability to have a good life would appear to be a constant theme throughout the majority of research examined here and should be explored further when applying the approach to health economics.

The comprehensive pearl review collected papers from a range of disciplines to answer specific questions related to health. Studies identified within this review will be of interest to researchers in other disciplines in the application of the capability approach, in terms of types of capability attributes, methods of aggregation of attributes and decision rules applied in such studies.

It remains to be seen how the capability approach can be combined with duration to assess between alternatives for health interventions. Some attempts have been made to connect different poverty literature to account for this (Clark & Hulme, 2010), but more research is required to understand the relationship in greater detail. This could be an area where the capability approach learns from the health economics outcomes (as detailed in Chapter 2), which combine health-related quality of life and time.

CHAPTER 5. CASE STUDY SELECTION AND MAPPING TERMINOLOGY

5.1 INTRODUCTION

In the first three chapters of this thesis, the impact of health economics on decision-making over the past half century has been detailed (Chapter 2), as well as introducing the capability approach to evaluating societal welfare (Chapter 3). The capability approach has been interpreted by a number of health economists as providing an alternative to the present theoretical basis and objectives of health economic evaluations (Verkerk et al., 2001; Anand & Dolan, 2005; Coast et al., 2008c; Ruger, 2010a; Smith et al., 2012). The development of measures of capability well-being designed to be used in assessing health and social care interventions was also explained in Chapter 3. The ICECAP questionnaires were identified for further use in Chapter 3 as the most developed approach for generating economic outcomes.

The first task in this chapter is to identify a relevant case study for this analysis. A study selection process is outlined and reasoning for the choice of the final case study chosen is justified. The reasons for exclusion of other potential case studies are also discussed.

From the case study selection process, it became clear that the lack of directly collected ICECAP data would make the measurement of capability over time difficult to account for with primary data. This chapter outlines an alternative approach to incorporating measures within economic models which is pursued in this thesis. This is known as mapping.

Individual well-being in health economics is frequently assessed through health related quality of life (HRQoL) questionnaires. When such questionnaires are combined with population preferences, they are often referred to as health utility instruments or preference based measures (Brazier et al., 2007). However, not all clinical studies collect HRQoL data on a frequent basis, with many trials focusing instead on condition-specific health status questionnaires, which aim to capture specific changes in quality of life and are favoured by clinicians as more likely to show changes associated with a given disease/illness. However, to comply with guidance from the National Institute for Health and Care Excellence (NICE) for economic evaluations within the UK, health utilities are required to be estimated so that Quality Adjusted Life Years (QALYs) can be calculated (NICE, 2013). QALYs, as discussed in Chapter 2, allow for comparisons across different interventions and patient groups across a health service.

Mapping from condition-specific and generic non-preference based questionnaires onto HRQoL to generate health utilities has become a popular method to generate QALYs when insufficient data on health utilities are available for the population under consideration. Mapping consists of generating a relationship between two instruments, allowing for the generation of QALYs when studies have not directly collected preference based measures (Brazier et al., 2010).

In this chapter, the first aim is to identify where an ICECAP questionnaire could be implemented in a case study to show how relevant capability outcomes can be generated from existing data in a modelling exercise. The second aim of this chapter is to explore current advice available to researchers on the methodology of applying mapping to generate health economic outcomes. This chapter lays the foundations as to what recommendations are currently available when undertaking a mapping study, and provides guidance for subsequent mapping research which is developed in later chapters of this thesis.

The chapter is structured as follows. Section 5.2 presents the case study selection process for exploring a capability measure in an economic model. Section 5.3 introduces the terminology associated with the mapping literature and the method of incorporating capability questionnaires within economic models, which is referred to frequently in the remainder of the thesis. Section 5.4 reviews the primary characteristics of a mapping study by appraising two recent sources of mapping research, to clarify what should be specified when formulating a mapping algorithm. Section 5.5 provides a summary of the case study selection process and the guidance that is currently available for mapping studies.

5.2 CASE STUDY SELECTION

In this section the case study selection process is chronicled step by step.

The data needed to generate capability outcomes required the collection of capability instruments, which could then be applied to make relevant comparisons between interventions. The two capability instruments considered for generating capability outcomes were the ICECAP-A for the general adult population (Al-Janabi et al., 2012a) and the ICECAP-O for the over 65 UK population (Grewal et al., 2006; Coast et al., 2008a). The ICECAP-A and ICECAP-O instruments were discussed in detail in Chapter 3.

It has been identified in previous research that capability outcomes are likely to be advantageous in areas where benefits usually are broader than health alone, for example in public health (Lorgelly et al., 2010a) and complex interventions (Payne et al., 2013). Furthermore, it has already been recognised by NICE that the benefits from public health go beyond health alone and therefore evaluations should take non-health benefits into consideration where possible (NICE, 2009b). Therefore, case studies which are from a public health perspective or another background that are likely to have broader benefits than health alone are likely to be particularly relevant for this research.

Health economists and modellers at the University of Birmingham were approached in person or by e-mail in October 2010, and were provided with a brief synopsis of this research project. They were asked to provide a list of potential models which they had worked with in the past decade or work-in-progress at that time point for consideration as a potential case study. Additionally, trials that had collected ICECAP questionnaires or were known to be in the process of collecting ICECAP data were also explored. This was done so that all possible options could be examined as to how best to compile capability outcomes in economic evaluations with the data currently available. Resource constraints of the doctoral funding restricted the possibility for primary data collection. An additional concern was the difficulty of assessing models from other developers outside of the University. Therefore, neither of these two options were considered any further in this research.

Table 5 presents the 18 topics that were considered for case study selection. The primary reasons that 12 topics (67%) were discarded in the case study selection process was the lack of readily available ICECAP data or the inability to collect the data in a reasonable time period for this study (Connock et al., 2007b; Barton & Andronis, 2009; Barton et al., 2011; Andronis et al., 2009; Chen et al., 2009; Auguste et al., 2011; Roberts et al., 2011; Roberts et al., 2007; Lovibond et al., 2011; Fletcher et al., 2010). This initial criterion was required as capability outcomes could not be generated without such data. The smoking prevention topic (Jit et al., 2009) was discarded as the model population were primary school children. The ICECAP questionnaires have so far not been developed to capture children's capabilities.

Three topics (topic 16 (Dowswell et al., 2012), topic 17 (Rouse et al., 2011) and topic 18 (Foster et al., 2010) in Table 5) which had ICECAP-A data available were subsequently ruled out due not having developed a model for the studies and not capturing health utility measures to be able to draw comparisons with the capability instrument. The final topic that was excluded was maintenance therapy for opioid dependents (Connock et al., 2007a; Holland et al., 2012). Even though ICECAP-O data could have been obtained from this topic, given the age group of the majority of opioid dependents is between 15-64 in the UK (Connock et al., 2007a), the ICECAP-O capability instrument, which is designed for the older over 65 UK population, was deemed inappropriate for the population under consideration and ICECAP-A data were not available for this population.

The chosen case study was therefore rheumatoid arthritis patients in the Birmingham Rheumatoid Arthritis Model (BRAM) (Jobanputra et al., 2002; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011). Even though the ICECAP-O was not collected for previous BRAM studies, a dataset of arthritis patients was obtained and could be applied to previous BRAM models as condition-specific questions were collected alongside the ICECAP-O (Pollard et al., 2009). Whilst not a public health topic *per se*, there is recognition that relevant benefits and costs are often unaccounted for within economic evaluations given the NHS and personal and social services perspective in the majority of UK based evaluations, as a health service perspective is required, according to current guidance by NICE (NICE, 2013). The treatment for patients in BRAM involves complex drug treatment strategies, which is another area where capability measures of benefit are likely to be fruitful because complex lengthy treatments could be at a disadvantage if wider benefits over a longer period are unaccounted for in an evaluation (Payne et al., 2013). Furthermore, the use of ICECAP-O as the capability measure with the BRAM population is more appropriate than for the opioid dependence option, given that the majority of the population are aged 55 and older (Malottki et al., 2011). A particular advantage of this model was the possibility for drawing on the advice from the principal model developer, Dr. Pelham Barton. Given this additional benefit, as well as the BRAM model meeting the main criteria for this case study, it was decided that BRAM was the most appropriate option for this case study. The methods and results of the BRAM case study are described in Chapters 8 and 9.

Table 5 Model Case Study Selection

	Area	Primary contact	Existing model?	Main economic outcome	ICECAP data	Public Health?	Model experience within team?	Reason for omission
1	Rheumatoid Arthritis - Birmingham Rheumatoid Arthritis Model (Jobanputra et al., 2002; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011)	P Barton	Y	QALYs	√		√	
2	Oral Anticoagulation (Connock et al., 2007b)	S Jowett	Y	QALYs				i.
3	Cardiovascular Disease Prevention (Barton & Andronis, 2009; Barton et al., 2011)	P Barton	Y	QALYs		√	√	i.
4	Skin Cancer (Andronis et al., 2009)	P Barton	Y	QALYs		√	√	i.
5	Pulmonary arterial hypertension (Chen et al., 2009)	S Jowett	Y	QALYs			√	i.
6	Smoking cessation (Jit et al., 2009)	P Barton	Y	QALYs		√	√	ii.
7	Breast Cancer (Auguste et al., 2011)	T Roberts	Y	QALYs			√	i.
8	Menorrhagia (Roberts et al., 2011)	T Roberts	Y	QALYs			√	i.
9	Chlamydia (Roberts et al., 2007)	T Roberts	Y	MOA		√	√	i.
10	Opioid dependency (Connock et al., 2007a; Holland et al., 2012)	S Jowett	Y	QALYs	√	√		iii.
11	High Blood Pressure Diagnosis (Lovibond et al., 2011)	S Jowett	N	n/a*				i.
12	Past Blood Pressure & TIA (Fletcher et al., 2010)	S Jowett	N	n/a*				i.
13	CLAHRC Stroke - Decision Tree Model - WIP	S Jowett	N	n/a*				i.
14	CLAHRC Transient Ischemic Attack - Discrete Event Simulation - WIP	S Jowett	N	n/a*				i.
15	Polypill for Cardiovascular Disease Prevention - Markov Modelling - WIP	S Jowett	N	n/a*				i.
16	Diet and Physical Activity to Prevent Recurrence of High Risk Adenomas: A Feasibility Study (Dowswell et al., 2012)	T Keeley	N	n/a	√			iv.
17	Motivation to exercise for at risk Cardiovascular Disease population (Rouse et al., 2011)	T Keeley	N	n/a	√			iv.
18	Benefits of Effective Exercise of Knee Pain (BEEP trial) (Foster et al., 2010)	T Keeley	N	n/a	√			iv.

QALYs, Quality Adjusted Life Years; MOA, Major Outcomes Averted; n/a, Not Available; WIP, Work-In-Progress

* Not available when selecting model. Could be available now.

i. No ICECAP data available
ii. ICECAP questionnaires not applicable for population under consideration
iii. Inappropriate capability measure for study population
iv. Time constraints in developing new model from beginning

To incorporate the ICECAP-O questionnaire into the BRAM model, a statistical association with a questionnaire collected within BRAM is required to produce capability outcomes. This requires what is referred to as mapping between instruments. The terminology and guidance for mapping studies is the focus of the remainder of this chapter.

5.3 MAPPING TERMINOLOGY

'Mapping', sometimes also referred to as 'cross-walking', is defined by Longworth and Rowen (2011) as:

> *"the development and use of an algorithm (or algorithms) to predict health-state utility values using data on other indicators or measures of health"* (Longworth & Rowen 2011, p. 9)

Mapping has grown in prevalence and statistical complexity over the past twenty years. In the UK in particular, this growth has coincided with the establishment of NICE in 1999, which stipulated the need for QALY outcomes to be generated for economic evaluations of new health technologies (NICE, 2004; NICE, 2013). Mapping allows for the generation of such data, even when they have not been collected within the primary study.

There are distinct phrases associated with the mapping literature. The terminology is primarily attributed to the work of Brazier and colleagues (2010). There are at least two measures/questionnaires/instruments required for a mapping study to be undertaken.

According to Brazier et al. (2010) these are:

- *Starting measure*: the primary measure of change in patients' health status (either condition-specific or generic) which was collected within the study sample.

- *Target measure*: the measure which is needed to generate required outcomes (i.e. QALYs from EQ-5D). The target measure is not collected within the study sample. In this case, the target measure is the ICECAP-O.

There are two stages to mapping which require two datasets to be available. According to Brazier et al. (2010) these are:

- *Estimation dataset*: In this dataset, starting and target measures must be completed by the sample population. The estimation dataset is where the predictive relationship between the starting measure and target measure is estimated to generate the mapping algorithm.

- *Study dataset*: In this dataset, only the starting measure is required to be collected within the dataset as the mapping algorithm into the target measure has already been generated from the estimation dataset. The study dataset is where the target measure is predicted from the starting measure.

The mapping algorithms generated from the estimation dataset are assessed using predictive error statistics to determine the most appropriate representation of the relationship between the starting and target measure. Lower prediction errors suggest that a mapping algorithm most accurately predicts the relationship between the starting and target measure.

Two predictive error statistics are most frequently used in mapping studies:

- *Mean absolute error (MAE)*: this statistic represents the average absolute distance from the observed to predicted values, which is non-negative regardless of the sign of the individual errors.

- *Root mean squared error (RMSE)*: this statistic is calculated from the squares of individual errors through the mean squared error (MSE) before finding the square root of this overall statistic. Higher individual errors from observed to predicted will make RMSE larger.

Predictive error statistics (MAE and RMSE) are generated by validating the mapping algorithms from the estimation dataset. Once again, two options are possible for choosing a validation dataset (Brazier et al., 2010):

- *External validation*: Mapping algorithms are tested on an external dataset which should have similar population characteristics to the estimation and study dataset.

- *Internal validation*: Also referred to as a "within-sampling" approach. The estimation dataset is split in two, with one split used to generate the mapping predictions, whilst the other part of the dataset is used to validate the algorithms.

5.4 GUIDANCE FOR MAPPING RESEARCH

To establish the most up to date theory and techniques applied for mappings in health economic evaluations, this section focuses on a recent review of mapping studies, as well as the current guidance offered by NICE on mapping methodology within health economics. These studies were identified through a manual search of the mapping literature.

5.4.1 Academic Guidance: Brazier et al. (2010)

In 2010, Brazier and colleagues undertook a review of the methods for generating QALYs through mapping or cross-walking from non-preference based measures of health in clinical studies onto generic preference-based measures. In their search, they found thirty studies mapping from non-preference to preference based measures of health, which between them developed 119 different mapping algorithms (Brazier et al., 2010).

Brazier and others (2010) present a table, which was adapted from an earlier discussion paper, that compiles the type of mapping functions that can be assembled (Tsuchiya et al., 2002). Six mapping functions were identified in total. The first four models predict the overall target measure score from: the starting measure overall score (1); dimension scores (2), item levels as continuous variables (3) or item levels as discrete variables (4). For the final two models, dimension levels for the target measure as continuous variable (5); or more appropriately as discrete variables (6), can be predicted by any of the previously method employed for the first four models from the starting measure. Additionally, other information such as demographic and clinical measurements as well as squared terms can be implemented as controls alongside the starting measure to predict the target measure.

The systematic review conducted by Brazier and colleagues (2010) extracted data on issues including the starting and target measures used, how the prediction models were specified, the type of regression analysis implemented and goodness-of-fit of the mapping models alongside predictive accuracy statistics. These data were then used to establish the validity of mapping models developed for economic evaluations and to generate guidance from these studies for future mapping research.

Brazier and colleagues (2010) found that from the thirty papers identified half of the studies mapped onto the EQ-5D (Brooks, 1996) and eight studies mapped onto the Health Utilities Index (HUI) (Feeny et al., 1995). The SF-6D (Brazier et al., 2002) was the third most prevalent target measure with five studies, whilst the remaining four mapped on to less commonly used preference measures. Two studies included mappings to more than one measure (Brazier et al., 2010).

The two most frequently applied starting measures were the SF-36 (Ware, Jr. & Sherbourne, 1992) and the refined version of the SF-36, the SF-12 (Ware et al., 1995), with the remaining studies primarily focusing on condition-specific questionnaires. The number of respondents for the mapping papers varied from very small numbers with the lowest sample of n=68 (Bosch & Hunink, 1996) to very large samples from national panel survey datasets, with the highest sample of n=23,467 (Sullivan & Ghushchyan, 2006).

Across the paper 119 different mapping models were tested, ranging from simple linear prediction from summary score to summary score for both measures, to more complex models, including interaction terms, dimension and item scores (Brazier et al., 2010). In the majority of cases, added complexity produced diminishing improvements in terms of model goodness-of-fit. Only one published study analysed in their review mapped onto individuals EQ-5D dimensions (Gray et al., 2006), with prediction of preference based total scores being the usual approach. Ordinary Least Squares (OLS) was the most popular regression estimation method, although other methods such as censored least absolute derivations (Kaambwa et al., 2013), generalized linear models (GLM) (Kaambwa et al., 2006) and Tobit models (Sullivan & Ghushchyan, 2006) were also attempted.

Model performance from generic health non-utility based to generic health utility instruments recorded average R^2, a routine measure of goodness-of-fit, of 0.5. Higher variation of goodness-of-fit was associated with condition-specific to generic measures, with R^2 ranging from 0.17 (Roberts et al., 2005) to 0.51 (Brazier et al., 2004b). In assessing the predictive ability of the model, however, the R^2 statistic is less important than prediction error statistics such as MAE and RMSE, with lower scores for these statistics representing better model predictions. MAE for all thirty studies ranged from 0.0011 to 0.19 and for RMSE, which gives more weight to less predictive accuracy (i.e. higher lower bound), going from 0.084 to 0.2 (Brazier et al., 2010). Within-sample testing (i.e., internal validation), was most commonly applied, although a number of studies attempted out-of-sample tests (i.e., external validation) (Brazier et al., 2010).

In summary, Brazier and colleagues (2010) suggest that the validity of the model depends on a number of details. A degree of overlap between the starting measure and target measure is essential, as the existence of dimensions in one measure that are not captured by the other measure is likely to weaken such model predictions. If minimal important differences between measures are a guide to whether a mapping model can be used, high RMSE is likely to undermine such a model to be used in economic evaluations. Nonetheless, the most important message from Brazier and colleagues' (2010) review is that mapping models should accurately account for the population as to where the algorithms will be mapped onto (i.e. the study dataset). Model specification is also important, as additional complexity can be added with little extra computational burden (Brazier et al., 2010).

5.4.2 NICE Guidance: Longworth and Rowen (2011)

Following on from Brazier et al.'s review of the mapping literature (Brazier et al., 2010), Longworth and Rowen (2011) produced a technical support document showing how mapping methods should be used to produce health state utilities for NICE. This technical support document was subsequently published in an academic journal (Longworth & Rowen, 2013).

As discussed in detail in Chapter 2, QALYs are the recommended outcome by NICE for economic evaluations in health. For a study to produce QALYs, health state utilities are required to be estimated for the population under consideration. The NICE methods guide states that when health utilities are unavailable, mapping or cross-walking between two instruments can be applied if a suitable mapping function can be established and validated (NICE, 2013). Furthermore, unlike early attempts to estimate utilities through expert opinion or “reprocessing” measures to produce QALYs (Gudex & Kind, 1988; Coast, 1992), mappings for NICE since 2008 require the implementation of a statistical mapping approach to estimating relationships between questionnaires, necessitating the use of empirical data (NICE, 2008).

Longworth and Rowen (2011) addressed a review specifically related to how utility values have been produced for health technology assessments (HTAs) in the UK (Tosh et al., 2011). Whilst mapping has been used in economic evaluations since the origins of NICE in 1999, the production of mappings became widespread following the first methods guide for technology appraisal in 2004 (Longworth & Rowen, 2011). The review by Tosh and colleagues (2011) of HTAs between the production of the first (NICE, 2004) and second (NICE, 2008) NICE methods guidance, found that out of the 71 HTA submissions in the UK using the first guide, more than 1 in 4 (n=19) produced mapping algorithms, with 16 of these mappings relying on empirical data, and the other three using expert opinion or reprocessing. However, since the second methods guide recognised explicitly that mapping was an inferior method to directly collected health utility values (NICE, 2008), the number of mapping studies for technology assessments has decreased since the period between the first and second guidance compared to the time up to the end of the review by Tosh et al. (2011).

The primary aim of the technical support document by Longworth and Rowen (2011) was to improve standards in the reporting of statistical performance of the mapping functions, which had been inconsistent and, in some cases, poorly handled within HTA submissions to NICE. Longworth and Rowen (2011) agree with the NICE methods guide that mapping should only be carried out through statistical association, where both the estimation and validation stages of the mapping process are clearly stated. They recognise that the estimation sample needs to be a fair reflection of the study dataset population characteristics. Statistical techniques should be implemented prior to model selection (e.g. correlation analysis) and post model selection (i.e. prediction error statistics), so that the appropriateness and performance of the mapping models can be assessed in full (Longworth & Rowen, 2011).

Longworth and Rowen (2011) also recommend that prediction errors for EQ-5D sub-groups, in terms of different ranges of EQ-5D health states, should be reported, as lower EQ-5D scores have been associated with higher prediction error scores (Gray et al., 2006). Longworth and Rowen (2011) also recommend that mapping models should preferably be validated by an external dataset from the model estimation sample, although a split within-sample is sufficient when an external dataset is unavailable. Finally, it was noted that the application of mapping models will not solve the issue of appropriateness of health utility measures for certain populations when such measures are not sensitive to capture all condition-specific changes for particular interventions (Longworth & Rowen, 2011)

5.4.3 Summary of Guidance available at present

The above two examples have attempted to improve the reporting of mapping studies for predicting outcomes in health economic evaluations. This will be further helped with the recent publication of the NICE guidance within an academic journal (Longworth & Rowen, 2013). However, there remains a gap within the health economics literature to guide researchers that are new to mapping on the processes involved in conducting a mapping study.

Indeed, one of the primary complaints within the mapping review was the inconsistency of reporting of the key steps involved within the mapping process (Brazier et al., 2010). The consistency of the methods used within mapping studies could be improved with clearer guidance as what steps are required to conduct a valid mapping between two instruments.

5.5 CONCLUDING COMMENTS

In this chapter, the process of case study selection for exploring the generation of capability outcomes from economic models was explained. Only one case study, the Birmingham Rheumatoid Arthritis Model (BRAM), was found to be eligible for this thesis investigation. This chapter has also explained the role of mapping within health economics. The guidance that is currently available to researchers interested in undertaking a mapping study, was described.

CHAPTER 6. EXPLORING THE RELATIONSHIP BETWEEN HEALTH STATUS AND CAPABILITY: MAPPING FROM THE WOMAC OSTEOARTHRITIS INDEX TO THE ICECAP-O CAPABILITY INDEX

6.1 INTRODUCTION

In the previous chapter, the process of the development of an algorithm between two measures, generally referred to as mapping, was documented in terms of how it is commonly applied to generate health utility data in studies that have not collected such data directly. Available academic and advisory guidance was reviewed to incorporate the ICECAP-O into the Birmingham Rheumatoid Arthritis Model (BRAM). However, before the case study can be dealt with in greater detail, a number of issues need to be addressed. First, it must be determined whether a capability instrument, such as ICECAP-O, has a relationship with health status questionnaires in order to be able to establish whether capability can be predicted when the capability questionnaires have not been collected directly. Second, the objective of a capability measure like ICECAP-O in an economic evaluation must be established. In this chapter, the first of these two research questions is addressed.

The aim of this chapter is to assess whether the mapping process used to predict health utilities can also be utilised to predict capability well-being for patient populations. Osteoarthritis (OA) patients requiring hip and knee replacement who have completed both the ICECAP-O capability index and the Western Ontario and McMasters Osteoarthritis Index (WOMAC) at three different time points are used to assess whether it is feasible for a questionnaire of capability well-being to be predicted from a condition-specific health status measure. This is the first attempt to map between health and capability and is investigated in terms of the feasibility and reliability of using the mapping process for predicting a capability instrument from a measure of health.

Section 6.2 presents the dataset applied to carry out this research. The questionnaires used in this case study, the WOMAC Osteoarthritis Index and ICECAP-O capability index, are then described. Section 6.3 presents literature which has mapped from WOMAC onto a preference based HRQoL questionnaire previously, so that QALYs can be obtained. These mapping studies provide a comparison for the mapping between WOMAC and ICECAP-O that will be carried out here. Section 6.4 tests the feasibility of mapping from health status to capability. This section describes the methods employed to estimate the prediction mapping models between WOMAC and the ICECAP-O capability questionnaire. The mode of validating the mapping models is also presented in this section, before reporting the results of the case study. Additionally, as a secondary analysis, a closer examination of the relationship between health and capability through the preferred mapping model coefficients is also presented in this section. Section 6.5 concludes this chapter with a discussion of the implications of the findings shown in the case study presented in Section 6.4.

Contents of this chapter have been previously published as: Mitchell, PM. et al. (2013) Predicting the ICECAP-O Capability Index from the WOMAC Osteoarthritis Index: Is Mapping onto Capability from Condition-Specific Health Status Questionnaires Feasible? *Medical Decision Making, 33*, 547-557.

6.2 CASE STUDY SELECTION

To predict statistical association between two measures and develop an algorithm for mapping, the target measure (ICECAP-O) and starting measure (WOMAC) must be collected in the same dataset (Brazier et al., 2010). The dataset used was a subset of the Tayside Joint Replacement cohort, where ICECAP-O was collected alongside more established health status instruments such as WOMAC (Pollard et al., 2009). The Tayside cohort was the only dataset available to this author that contained follow-up data on both ICECAP-O and a condition-specific instrument of health. This allowed the mapping models between the two instruments to be internally validated. The two questionnaires used to undertake the objectives in this chapter are explained next.

6.2.1 Western Ontario and McMaster (WOMAC) Osteoarthritis Index

The WOMAC index is a condition-specific health status questionnaire, which aims to measure problems for OA patients in relation to the pain, stiffness and physical function of their affected joints (Bellamy et al., 1988). The WOMAC questionnaire consists of twenty four questions and three categories (see Appendix 5). The first five questions are about pain, followed by two questions on stiffness, with the remaining seventeen questions concerned with the limits of physical function from the affected joint(s). Each question is asked on a five part Likert scale, ranging from no problems to extreme problems. Individual responses to each question produce a score between 0-4 (no problems – extreme problems). Each response is then summed within category and the three categories are combined to form a WOMAC score ranging from 0-96 (Bellamy et al., 1988). When items are missing, standard mean imputation is commonly applied for category completion, although expectation maximisation imputation, which relies on probability based imputation, has recently been suggested as an alternative (Ghomrawi et al., 2011).

6.2.2 ICECAP-O capability questionnaire

The ICECAP-O capability index, already reported in Chapter 3, is the target measure in this study. The ICECAP-O consists of five attributes of capability well-being. These are attachment, security, role, enjoyment and control and are measured across four levels (see Appendix 1 for full questionnaire). Capability well-being is anchored to a 1-0 scale, where 1 is equivalent to full capability on all attributes (44444) and 0 to having no capability on any of the five attributes (11111). Values for each of the five capability wellbeing attributes were obtained in a separate valuation study (Coast et al., 2008a).

6.3 MAPPING FROM WOMAC TO HEALTH UTILITY

Before assessing the ability to predict capability from health, it is important to assess how well health utility instruments have been captured from similar starting measure questionnaires. A recent paper identified three mapping papers where health utility has been predicted from WOMAC up until May 2011 (Lin et al., 2013). Similar search strategies were adapted by this author to update the review by Lin et al. (2013) until October 2012 and no new mapping papers were discovered. One paper predicted the Health Utilities Index Mark III (HUI3) (Grootendorst et al., 2007) and two papers predicted the EQ-5D (Barton et al., 2008; Xie et al., 2010).

6.3.1 Mapping from the WOMAC to the Health Utility Index Mark III (HUI3)

The Health Utilities Index Mark III (HUI3) is a HRQoL preference based measure, used predominantly in North America and can be used to generate QALYs (Furlong et al., 2001). The HUI3 consists of eight dimensions (vision, hearing, speech, ambulation, dexterity, emotion, cognition and pain) with five to six levels for each dimension ranging from severe limitations to normal (Feeny et al., 1995). The HUI3 is anchored on a 0-1 being dead to full health scale, but HUI3 scores range from -0.36 to 1 in actual practice (Feeny et al., 2002). Therefore, scores which can be considered worse than death can be implicitly inferred from patient responses.

6.3.1.1 Grootendorst et al. (2007)

The first paper to map the WOMAC onto a health utility measure mapped WOMAC onto the HUI3 (Grootendorst et al., 2007). The sample consists of 255 patients suffering from mild to moderate knee OA in a Canadian province. The mapping models in this study were validated from a within sample split in the estimation dataset for model prediction (two thirds of the population) and validation (one third of the population).

The prediction of overall HUI3 score by WOMAC was tested in five different mapping models. In the first model, WOMAC item responses were entered as dummy variables, resulting in 96 explanatory variables for the HUI3 score, with no problems as the base category for each WOMAC item. In model two, mild and moderate were combined as one category, as were severe and extreme responses, which reduced the potential explanatory variables by half to 48. Model three included three independent variables in terms of the

WOMAC category scores of pain, stiffness and physical function. Interaction terms between the category scores as well as the squared category responses to account for a potential non-linear relationship between the two measures were added to model three to make up model four. Finally, model five tested overall WOMAC score and the squared WOMAC category score. All five models were tested with and without the demographic variables age and sex, and without the clinical characteristics which were years since onset of OA and Kellgran radiographic scale. All models were tested using OLS and random effects regressions. The best model was assessed through the predictive error statistic MAE, with RMSE also reported.

The OLS regression outperformed the random effects regression for all model specifications. Model four which included age, gender and years since onset of OA produced the lowest MAE score and was therefore considered the best prediction model. When the Kellgran radiographic scale is included, goodness-of-fit of model four improved from explaining 39% to 40% of the HUI3 scores. Without the Kellgran scale, which is not collected as commonly for OA as the other measures included in model 4, MAE (0.1628) and RMSE (0.2065) scores are recorded.

HUI3 score differences of 0.03 are regarded as clinically important (Horsman et al., 2003), but only 10% of the best model predictions fell within this range of individual error predictions. Group mean errors were assessed by bootstrapping the estimation sample with replacement for varying sample sizes. Lower HUI3 prediction errors were reported with higher bootstrapping sample sizes. The authors of the mapping study note that the lack of a crossover between particular dimensions on HUI3, such as vision, emotion and hearing, with WOMAC categories is the likely source for the higher prediction error at the individual level, as well as the moderate goodness-of-fit results (Grootendorst et al., 2007).

Grootendorst et al.'s (2007) recommended model for predicting HUI3 from WOMAC is as follows:

$$
\begin{aligned}
\text{Predicted HUI3 score} = {} & 0.5274776 + [0.0079676 \times \text{Pain}] + [0.006511 \times \text{Stiffness}] \\
& -[0.0059571 \times \text{Function}] + [0.0019928 \times \text{Pain*Stiffness}] \\
& +[0.0010734 \times \text{Pain*Function}] + [0.0001018 \times \text{Stiffness*Function}] \\
& -[0.0030813 \times \text{Pain*Pain}] - [0.0016583 \times \text{Stiffness*Stiffness}] \\
& -[0.000243 \times \text{Function*Function}] + [0.0113565 \times \text{Age in Years}] \\
& -[0.0000961 \times \text{Age in Years*Age in Years}] - [0.0172294 \times \text{Female}] \\
& -[0.0057865 \times \text{Years since onset of knee OA}] \\
& +[0.0001609 \times \text{Years since onset of knee OA*Years since onset of knee OA}]
\end{aligned}
$$

6.3.2 Mapping from the WOMAC to the EQ-5D

The EuroQol 5 Dimension (EQ-5D) has already been discussed in Chapter 2. The EQ-5D consists of five HRQoL dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) (Brooks, 1996). Preferences for health states are valued on a 0-1 death-full health scale, with UK values for health states ranging from -0.59 to 1 (Dolan, 1997). The mappings undertaken previously with the EQ-5D and WOMAC were with the three level (EQ-5D-3L) version (Brooks, 1996) and not the more recently developed five level (EQ-5D-5L) version (Herdman et al., 2011).

6.3.2.1 Barton et al. (2008)

Barton and colleagues (2008) explored whether EQ-5D scores estimated from WOMAC differed substantially from directly observed EQ-5D scores to generate QALYs for economic evaluations. Three hundred and eighty nine individuals participated in a Lifestyle Intervention for Knee Pain (LIKP) study, with EQ-5D and WOMAC collected at four time points over a two year period. The LIKP study explored four advisory interventions on reducing knee pain. Similar to the study by Grootendorst and colleagues (2007), the estimation dataset was split in two for estimation and validation purposes. However, whilst it was unclear in the previous mapping study how the dataset was divided, Barton et al. (2008) split the dataset such that baseline responses were used to estimate the mapping models (n=348), with questionnaires completed at the three follow-up time points (n=259×3) used to validate the mapping models.

Five mapping models examining the predictive relationship between WOMAC and EQ-5D were explored. The dependent variable in all five models was overall EQ-5D score. In model A, only the total WOMAC score was used to predict EQ-5D. Model B predicted EQ-5D by the three WOMAC category scores. Model C added WOMAC squared to Model A, with the interactions and squares of WOMAC category scores added to Model B to Model D. Model E added age and sex variables to the best performing models of the four previous models described. MAE, RMSE and adjusted R^2 were estimated for all models (Barton et al., 2008).

Out of the first four models, Model C, which predicted EQ-5D from the WOMAC score and WOMAC score squared (WOMAC*WOMAC), produced the lowest prediction error estimates. When this was combined with age and sex in model E, MAE was lowest of all five models at 0.129, RMSE was 0.180 and an adjusted R^2 of 0.313, meaning that a little less than one third of the overall EQ-5D score was explained by the explanatory variables included in Model E.

Cost per QALY gains for the preferred intervention almost doubled when using actual data from the 259 individuals who completed all time points versus the mapping prediction estimates from the preferred Model C (£13,154 versus £6,086). While these results would not alter a decision in the UK, where current willingness-to-pay for additional QALYs is estimated at less than £20,000 to £30,000 per QALY gain (NICE, 2013), it nonetheless suggests that care should be taken when inferring results from QALYs generated from mapping algorithms. If at all possible, Barton and colleagues (2008) recommend that primary data should be used to generate QALYs, which concurs with advice produced by NICE in the same year (NICE, 2008).

The recommended model (Model E) for predicting EQ-5D from WOMAC by Barton et al. (2008) is presented below:

$$
\begin{aligned}
\text{Predicted EQ-5D score} = {} & -0.3474012785 + [-0.0005977709 \times \text{total WOMAC}] \\
& +[-0.0001081560 \times \text{total WOMAC*total WOMAC}] \\
& +[0.0326027536 \times \text{age}] \\
& +[-0.0002352456 \times \text{age*age}] \\
& +[0.0475889687 \times \text{Female}]
\end{aligned}
$$

6.3.2.2 Xie et al. (2010)

The most recent of the studies mapping from WOMAC to health utility is a paper by Xie and colleagues (2010), which looked at mapping for WOMAC to EQ-5D for knee OA patients (Xie et al., 2010). They identified overlap between the two measures given that both questionnaires addressed issues of pain and physical function/mobility explicitly. Two hundred and fifty eight individuals from Singapore completed the questionnaires and were randomly split in two for the estimation and validation of the proposed models. EQ-5D values were generated from a Japanese population as no existing valuation dataset was available for the Singapore population at the time of this research.

Models were estimated using OLS and censored least absolute derivations. Four models examined the relationship for both regression methods. Overall EQ-5D score was predicted by WOMAC score (Model i.) and WOMAC category scores (Model ii.). Model iii. represented the same input variables as Model ii. plus interaction and category scores squared, while Model iv. accounted for WOMAC item scores that were identified to be significant after a stepwise regression procedure for OLS to eliminate non-significant explanatory variables. Demographics such as age and sex were not included in the mapping models (Xie et al., 2010).

Model ii. (EQ-5D predicted from WOMAC category scores) using OLS produced the lowest predictive errors in terms of MAE (0.0736) and RMSE (0.0947). However, these results are not directly comparable with individual reported errors in the two previous mapping studies, as the above statistics represent group mean error using bootstrapping, rather than differences in individuals' observed versus predicted scores. When the preferred model was re-estimated with the full sample, an adjusted R^2 of 0.449 was recorded, which is considerably higher than the two previous studies. Additional problems with comparing this research with the other two studies also arise due to the different valuation sets used for EQ-5D as well as the sensitivity of knee pain for the different mapping populations.

The preferred mapping model of Xie and colleagues (2010) is presented below:

Predicted EQ-5D score = 0.83414 - [0.00166 × WOMAC pain score]

-[0.00092 × WOMAC stiffness score]

-[0.00330 × WOMAC function score]

6.3.3 Summary of Mappings from WOMAC to health utility

Of the three previous studies that mapped from WOMAC to health utility, it was the preferred model by Xie et al. (2010) that produced the lowest MAE by a considerable distance (0.074 compared to 0.142 to 0.163). However, different methods, patient groups, utility instruments, interventions and follow-up data were applied in each study so comparisons between studies should be treated with caution. Notwithstanding this, it is worth noting that higher R^2 in the preferred models did not necessarily lead to more accurate predictions of utility, as Barton et al. (2008) have lower R^2 (0.30 versus 0.39) but also lower MAE than Grootendorst and colleagues (2007) (0.142 versus 0.163).

6.4 MAPPING FROM WOMAC TO ICECAP-O: A CASE STUDY

In this section, the chosen case study of mapping from the WOMAC Osteoarthritis Index to the ICECAP-O capability index is detailed, with the methods and results presented.

6.4.1 Methods

6.4.1.1 Dataset

The dataset used for this study was the Tayside Joint Replacement cohort, where ICECAP-O and WOMAC were collected simultaneously (Pollard et al., 2009). This dataset consisted of 107 arthritis patients who were about to undergo primary joint replacement surgery at Ninewells Hospital, Dundee, UK, between September 2006 and June 2007. Prior to treatment, patients completed both the ICECAP-O and WOMAC. Pre-operative data were used to establish statistical relationships between the questionnaires through mapping models in the prediction dataset. Replicating the method used by Barton et al. (Barton et al., 2008), follow-up data at 1 and 3 years post-operation were used to validate the prediction models.

6.4.1.2 Measures

The WOMAC index has been described previously in this chapter (Section 6.2.1) and is the starting measure in this study. WOMAC has been previously mapped onto two health utility instruments since 2007 (Section 6.3). The ICECAP-O capability index has also been described earlier in this chapter (Section 6.2.2) and is the target measure in this study. This is the first time that the ICECAP-O has been used in a mapping study.

6.4.2 Statistical Analysis

The first step in the statistical analysis involved the generation of descriptive statistics for all possible dependent and independent variables. Three explanatory variables (overall WOMAC score, age, sex) were explored for predictive significance. Overall ICECAP-O score (continuous variable) and ICECAP-O dimensions (categorical variables) were the dependent variables considered. Scatter-graphs were used initially to explore the association between ICECAP-O scores and each of the potential explanatory variables. Box-plots for the five attributes of capability well-being on ICECAP-O were employed to ascertain the relationship between each attribute level and overall WOMAC scores.

Since this study is the first mapping attempt from a condition-specific health measure to a measure of capability well-being, there was no *a priori* position on what was the most appropriate model, so a process of model specification was required, following methods from a previous mapping study (Kaambwa et al., 2013).

6.4.2.1 WOMAC Imputation

Where all questions on the WOMAC were not completed in full, WOMAC categories and overall scores were completed using standard mean imputation, the most commonly used technique to complete WOMAC scores with missing data (Ghomrawi et al., 2011). Patient responses were excluded if no questions on the pain and stiffness category were answered, or if less than four physical function questions were completed. When not all questions were completed, the average of the completed score for the respective category is calculated, rounded to the nearest whole number and assigned to the missing response. This follows the recommended guidance for imputing WOMAC category scores (Bellamy, 2004).

6.4.2.2 Regression Specification

Two model specifications were considered for further analysis.

6.4.2.2.1 Ordinary Least Squares (OLS) and ICECAP-O score

ICECAP-O as continuous dependent variable. OLS regression was the first regression chosen due to the prevalent use of this approach in mapping studies (Brazier et al., 2010), particularly for arthritis mapping studies (Marra et al., 2011). There have been notable limitations when using OLS to predict EQ-5D scores in previous studies, particularly due to ceiling effects, when a high proportion of scores are observed at one end of the scale (Gray et al., 2006). However, there are no ceiling effects with the ICECAP-O in our current dataset, as only one person recorded the highest capability score achievable in the prediction dataset. Therefore, models that have used alternative specifications to deal with ceiling effects such as censored least absolute derivations (Kaambwa et al., 2013), Tobit (Sullivan & Ghushchyan, 2006), generalized linear models (Dakin et al., 2013b) and two part models (Dakin et al., 2013b) were not tested here. The three explanatory variables considered for the first mapping model consisted of two continuous variables (WOMAC score, age) and one discrete variable (sex). Stepwise regression, a process of finding the best model fit for a regression when no *a priori* knowledge of the appropriate inclusion of explanatory variables is available, was used to test the significance of the three explanatory variables. A number of stepwise regression techniques are used to specify models such as forward selection and backward elimination. Backward elimination stepwise regression includes all variables in the original model and removes the least significant variable for each model run, with the process finalised when only significant explanatory variables remain (Draper & Smith, 1998). Backward elimination is the stepwise regression process employed here.

6.4.2.2.2 Multinomial Logistic (ML) Regression and ICECAP-O dimensions

ICECAP-O attributes (5) as categorical dependent variables. There were two model options available for use with categorical dependent variables: ordinal logistic (OL) or multinomial logistic (ML) regression. While each of the five attributes of ICECAP-O are ordered, tests were required on the assumption of proportional odds, which is required when using the OL regression approach. These assumptions have been violated when applied in a previous mapping study (Gray et al., 2006). Alternatively, if the proportional odds assumption is violated, ML regression can be used. ML regression does not recognise the order of categories in the same way as OL, but assigns a probability to the likelihood of a person having a particular response level on an attribute given the explanatory variable results for an individual (Draper & Smith, 1998).

Three ML methods considered for this study are:

- Expected-Utility Method: The average probability across levels for each category to predict the overall dependent variable score (i.e. ICECAP-O score) (Le & Doctor, 2011)
- Most-Likely Probability: The highest probability level for each category used to predict overall dependent variable score (Le & Doctor, 2011)

- Monte Carlo Simulation: Using repeated simulations to generate random numbers on a uniform distribution. ICECAP-O responses for each ICECAP-O attribute are predicted by the probability of a response level, which are then combined with the other ICECAP-O categories to predict the overall dependent variable score (Gray et al., 2006). Given the small sample size here, 1000 simulations were carried out in the analysis for each observed individual

In total 10 mapping models predicting ICECAP-O scores or ICECAP-O dimensions were tested; the primary characteristics of each model are displayed in Table 6. Seven models (Models 1-7) predicted overall ICECAP-O scores as a continuous variable through Ordinary Least Squares (OLS) regression. Model 1 considered WOMAC score, age and sex as explanatory variables of overall ICECAP-O scores. The reduced equation of WOMAC as the sole predictor of ICECAP-O is illustrated in Model 2. In Model 3, overall ICECAP-O scores are predicted by the WOMAC score squared. Model 4 tested the three category scores for WOMAC (pain, stiffness, physical function) on ICECAP-O; while Model 5 is the reduced version of Model 4 where only significant predictors of capability are included. The same process is used for Model 7, which includes the significant items of WOMAC from all 24 WOMAC items, which were tested in Model 6. Models 8-10 explored ICECAP-O dimensions as dependent variables. Three models (Expected Utility Method – Model 8, Most-Likely Probability – Model 9, Monte Carlo Simulation – Model 10) were analysed for their ability to predict ICECAP-O attributes with WOMAC as the sole explanatory variable.

Table 6 Mapping models for ICECAP-O prediction from WOMAC

Model number	Dependent variable(s)	Independent Variable(s)
1	ICECAP-O score	WOMAC score; age; sex
2	ICECAP-O score	WOMAC score
3	ICECAP-O score	WOMAC score squared
4	ICECAP-O score	WOMAC category scores
5	ICECAP-O score	WOMAC physical function category
6	ICECAP-O score	WOMAC items (24)
7	ICECAP-O score	Significant WOMAC items from model 6*
8	ICECAP-O attributes	WOMAC score - Expected Utility method
9	ICECAP-O attributes	WOMAC score - Most-Likely probability
10	ICECAP-O attributes	WOMAC score – Monte Carlo simulation

Ordinary Least Squares (OLS) regression carried out on models 1-7; Multinomial Logistic (ML) regression for 8-10.
*3 items on WOMAC were significantly related to ICECAP-O score. They were item 9 (difficulty going up stairs, item 15 (difficulty when shopping) and item 22 (difficulty getting on/off toilet)

6.4.2.3 Prediction Accuracy

While common measures of goodness-of-fit of regression models such as R^2 play an important role in showing the explanatory power of a model for the dependent variable, the primary interest here is in the ability to accurately predict the dependent variable from the explanatory variables in the model. Two common measures have been prominently used in mapping studies: mean absolute error (MAE) and root mean squared error (RMSE). RMSE gives a higher error score for larger errors from the observed score to the predicted score than MAE and both are generally reported in mapping studies (Brazier et al., 2010). All models are tested for these two measures of prediction error, with lower prediction error scores indicating a better model for prediction. R^2 of all models are reported to test the goodness-of-fit between capability and condition-specific health status. The relationship between the preferred model attributes are examined in a secondary analysis to help understand the relationship between health and capability in greater detail. All models were also tested for normality and heteroscedasticity. Analysis was carried out using STATA Version 10.1 and Microsoft Excel 2007.

6.4.3 Results

6.4.3.1 Demographics of Dataset

Table 7 shows the demographic information for the sample population. Two patients did not complete either the ICECAP-O (n=1) or WOMAC (n=1) to a level where overall scores for both instruments could be calculated, so they were excluded from further analysis. The mean ICECAP-O score for the remaining 105 patients was 0.772, which is lower than the average UK over 65s population score of 0.832 (Flynn et al., 2011). Respondents had a mean age of approximately seventy years. The mean WOMAC score was 50.628. Figure 9 shows the distribution of the ICECAP-O scores for the prediction dataset at baseline. It shows that ICECAP-O scores are negatively skewed, with the majority of scores closest to the higher end of the scale.

Figure 9 Distribution of ICECAP-O in prediction dataset

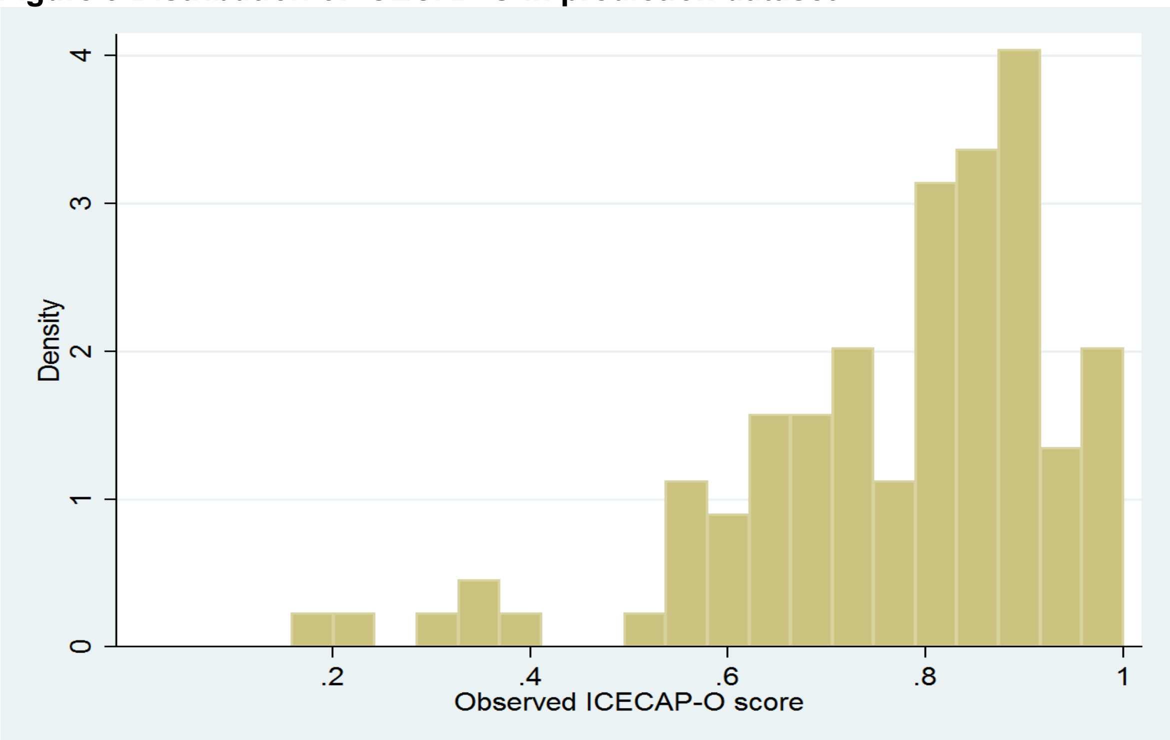


Table 7 Descriptive statistics for prediction and validation dataset

	N	n	Mean	Standard Deviation	Min	Max
Sample size	107					
Missing data	107	2				
Males	105	54				
Employed	105	13				
Living Alone	105	28				
Age (mean)	105	105	69.738	8.894	48	89
ICECAP-O (base)	105	105	0.772	0.168	0.159	1
ICECAP-O (1 year)	107	56	0.861	0.134	0.514	1
ICECAP-O (3 year)	107	54	0.836	0.135	0.481	1
WOMAC (base)	105	105	50.628	17.052	18	91
Pain		105	10.857	3.817	2	20
Stiffness		105	4.714	1.517	1	8
Physical Function		105	35.057	12.711	7	64
WOMAC (1 year)	107	56	16.018	14.633	0	68
Pain		56	2.589	3.561	0	13
Stiffness		56	1.768	1.513	0	8
Physical Function		56	11.661	10.786	0	47
WOMAC (3 year)	107	54	20.667	19.443	0	72
Pain		54	3.111	3.78	0	14
Stiffness		54	1.741	1.604	0	6
Physical Function		54	15.815	15.471	0	52

N,total population; n,sub population from N; min, minimum; max, maximum; ICECAP-O range 0-1; WOMAC range 0-96.

6.4.3.2 WOMAC Imputation

For the 105 patients included at baseline, WOMAC was completed in full by 81 of the patients prior to their operation. Four patients did not complete the pain category fully, with 3 patients not responding to one question and one patient not responding to two questions. Twenty-six patients did not complete all seventeen physical function questions. Fifteen only failed to complete one question, with question 20 (difficulty getting in or out of the bathtub), the most problematic question for this sample, with 17 patients failing to give a response. No imputation was necessary for the stiffness category at baseline.

For the patients who completed sufficient questions on both the ICECAP-O and WOMAC at 1 year follow up (n=56), nine patients did not complete all the pain questions. Seven patients failed to complete only one of the pain questions, with question 5 (pain while standing) causing the most non-completions (six) for this category. Fifteen patients failed to complete all of the physical function questions, with eight patients failing to complete only one question. Question 20 again had the lowest response rate with ten patients not answering. Once more, all stiffness questions were completed in full.

At three years follow up, fifty four patients completed the ICECAP-O and WOMAC to a standard where overall scores could be calculated. Two patients did not complete one of the pain questions (question 1 pain when walking on a flat surface). All other patients completed all five pain category questions. Eight patients did not complete all seventeen physical function questions, with five patients failing to complete only one question. None of these eight patients answered question 20, once again highlighting the low levels of response to problems getting in and out of the bath question. Both stiffness questions were completed by all patients once again.

6.4.3.3 Relationship between ICECAP-O and WOMAC score

Figure 10 shows the relationship between ICECAP-O and WOMAC in the prediction dataset. Figure 10 indicates a trend towards higher ICECAP-O scores. Table 8 presents the relationship between ICECAP-O responses and WOMAC scores from the prediction dataset. For all but two capability attributes (attachment and security), there is an increase in WOMAC score as capability responses decrease. For the security attribute, there is little difference between WOMAC scores for 'a lot of' and 'full' capability, whilst increases in

WOMAC scores for the lower levels of capability follow similar patterns to the ICECAP-O's role, enjoyment and control attributes. There is no clear pattern for WOMAC scores with the attachment capability attribute. These results are intuitive with the symptoms of OA, as it would not be expected to have an impact on psychological well-being (i.e. attachment) as much as physical well-being.

Figure 10 Relationship between ICECAP-O and WOMAC scores (baseline)

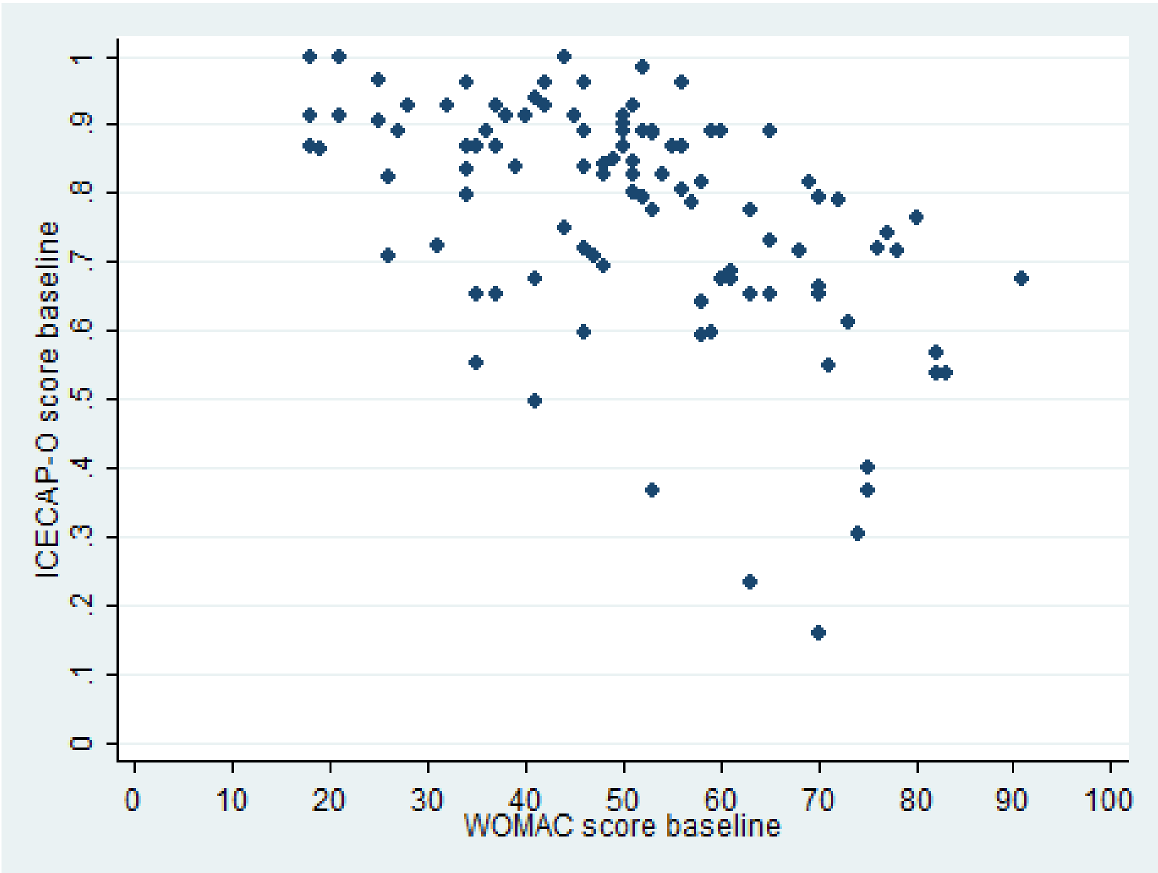


Table 8 Prediction Dataset - ICECAP-O responses and WOMAC scores

ICECAP-O dimension	Freq. (n)	WOMAC score (SD)
Attachment (Love and Friendship)		
(4) I can have all of the love and friendship that I want	58	52.3(17.9)
(3) I can have a lot of the love and friendship that I want	36	47.3(15.4)
(2) I can have a little of the love and friendship that I want	9	49.6(17.7)
(1) I cannot have any of the love and friendship that I want	2	66.5(04.9)
Security (Thinking about the future)		
(4) I can think about the future without any concern	15	45.9(17.3)
(3) I can think about the future with only a little concern	49	45.1(15.7)
(2) I can only think about the future with some concern	29	56.1(15.3)
(1) I can only think about the future with a lot of concern	12	65.7(14.4)
Role (Doing things that make me feel valued)		
(4) I am able to do all of the things that make me feel valued	14	39.4(14.2)
(3) I am able to do many of the things that make me feel valued	46	46.0(15.8)
(2) I am able to do a few of the things that make me feel valued	38	57.1(15.8)
(1) I am unable to do any of the things that make me feel valued	7	68.7(08.1)
Enjoyment (Enjoyment and Pleasure)		
(4) I can have all of the enjoyment and pleasure that I want	12	39.5(15.3)
(3) I can have a lot of the enjoyment and pleasure that I want	46	45.3(14.3)
(2) I can have a little of the enjoyment and pleasure that I want	37	56.5(16.7)
(1)I cannot have any of the enjoyment and pleasure that I want	10	66.7(13.8)
Control (Independence)		
(4) I am able to be completely independent	26	38.6(14.8)
(3) I am able to be independent in many things	46	47.3(12.9)
(2) I am able to be independent in a few things	27	64.4(14.0)
(1) I am unable to be at all independent	6	66.5(15.9)

freq.; frequency; n, population; SD, standard deviation

6.4.3.4 Model Specification

Stepwise regression was applied for both ICECAP-O score and ICECAP-O attributes Mapping Models (see Table 6). Age and sex were not explored beyond Model 1 as they were omitted as non-significant variables for predicting ICECAP-O from the backwards elimination stepwise regression process. Variable transformation was applied in Model 3 to tackle the issue of non-normally distributed data for OLS regressions, which can be seen in

Figure 9. The assumption of proportional odds between categorical levels did not hold for ICECAP-O dimensions using the Wald test, so ML regression was used for Models 8-10.

6.4.3.5 Internal Validation Results

1 year and 3 year post operation data were combined to validate the Mapping Models described in Table 6. Table 9 presents the results of the validation. The majority of models relied on all individuals where ICECAP-O and WOMAC scores could be calculated (responses at two follow-up periods =110 individual responses in total). However, for models 6 and 7, all items on WOMAC had to be completed, which reduced the sample size for both of these model validations. The Monte Carlo Simulation model (Model 10) produced the lowest absolute difference from predicted to observed ICECAP-O mean score at 0.0233. Model 10 also produced the highest variety in ICECAP-O responses, although this is in part due to the Monte Carlo simulation process employed in this model. All OLS models which were not transformed to address the non-normally distributed ICECAP-O data resulted in higher ICECAP-O scores than are feasible in reality, as full capability is equal to 1. This may explain why all predicted ICECAP-O scores are above the observed ICECAP-O scores. In terms of prediction error statistics, Model 4, which predicted ICECAP-O from the three WOMAC category scores, produced the lowest mean absolute error (MAE = 0.0832) and root mean squared error (RMSE = 0.1193). Model 4 also produced the highest R^2 at 0.3976. It is therefore recommended that out of the ten models tested here, predicting capability by category scores of pain, stiffness and physical function produces the most accurate results for arthritis patients requiring knee and hip replacements.

Table 9 Internal Validation: Mapping Model Results

Model Number	n	Observed ICECAP-O	Predicted ICECAP-O	Min Predicted	Max Predicted	Absolute Difference*	MAE(SD)	RMSE	Overall R^2
1	110	0.8486	0.9344	0.6695	1.0372	0.0858	0.1000(0.099)	0.1406	0.3198
2	110	0.8486	0.9410	0.6608	1.0365	0.0924	0.1046(0.102)	0.1454	0.3125
3	110	0.8486	0.8871	0.6516	0.9196	0.0385	0.0865(0.089)	0.1240	0.2394
4	110	0.8486	0.9041	0.5709	1.0516	0.0556	0.0832(0.086)	0.1193	0.3976
5	110	0.8486	0.9283	0.6498	1.0279	0.0797	0.0970(0.094)	0.1344	0.3637
6	83	0.8562	0.8841	0.4232	1.0603	0.0279	0.0966(0.084)	0.1276	0.2326
7	103	0.8520	0.9362	0.6362	1.0108	0.0842	0.1026(0.100)	0.1429	0.2495
8	110	0.8486	0.8737	0.6432	0.9010	0.0251	0.0874(0.087)	0.1229	0.2131
9	110	0.8486	0.8897	0.6753	0.9136	0.0411	0.0965(0.099)	0.1379	0.0429
10	110×1000	0.8486	0.8719	0.1588	1.0000	0.0233	0.0874(0.093)	0.1226	0.2108

n, validation sample size; min, minimum; max, maximum; MAE, mean absolute error; RMSE, root mean squared error; SD, standard deviation; R^2, square of the Pearson product-moment correlation coefficient between observed and predicted ICECAP-O scores; * Absolute Difference between the observed and predicted ICECAP-O scores

6.4.3.6 Recommended Mapping Model

In Table 10, the preferred Mapping Model (Model 4) which can be used to predict ICECAP-O from WOMAC category scores is presented. 95% confidence intervals are given around the explanatory variables, which can be used in sensitivity analysis to account for uncertainty around the model algorithm.

Table 10 Best Performing Mapping Algorithm:
Predicting ICECAP-O scores from WOMAC category scores (Model 4)

Variable	Coefficient(SE)	t	P>|t|	95% Confidence Interval
Pain	0.0019(0.007)	0.28	0.780	[-0.0114,0.0152]
Stiffness	0.0141(0.013)	1.13	0.263	[-0.0108,0.0391]
Physical Function	-0.0088(0.002)	-4.51	0.000	[-0.0127,-0.0049]
Constant	0.9950(0.048)	20.89	0.000	[0.9005,1.0896]

SE; standard error.

6.4.3.7 What aspects of ICECAP-O are WOMAC categories predicting?

Given the validation results of the previous section, it was also possible to examine exactly which attributes of capability were being captured by changes in the WOMAC category scores for pain, stiffness and physical function score respectively. This was carried out by analysing the predictive ability of each capability dimension on ICECAP-O, by individually examining the relationship with each WOMAC category score from the prediction dataset through ML regression. Table 11 presents the findings from this research.

The prediction of the ICECAP-O dimension 'control' from the physical function category on WOMAC produced the highest significance in terms of R^2 at 0.2143, with significant p-values for all control response levels at the 1 percent level. The WOMAC physical function category also predicts the two lower levels of capability responses for three of the other four

ICECAP-O dimensions (security, role and enjoyment). Stiffness is best able to predict the security dimension, in particular the two lower levels of security responses. Pain is able to predict the two lower levels from full capability on three dimensions (role, enjoyment and control). There is no apparent relationship between any of pain, stiffness and physical function on the one hand, and the attachment ICECAP-O attribute on the other.

Table 11 Prediction of ICECAP-O dimensions from WOMAC categories: multinomial logistic (ML) regression p-values

Category (base case = level 4 or full capability)	Pain	Stiffness	Physical Function
Attachment			
a lot of capability (level 3)	0.041	0.023	0.359
a little capability (level 2)	0.335	0.884	0.757
No capability (level 1)	0.448	0.613	0.218
Pseudo R^2	0.0275	0.0304	0.0143
Security			
a lot of capability (level 3)	0.532	0.146	0.840
a little capability (level 2)	0.152	0.014	0.051
no capability (level 1)	0.026	0.001	0.003
Pseudo R^2	0.0562	0.0617	0.0800
Role			
a lot of capability (level 3)	0.143	0.951	0.156
a little capability (level 2)	0.004	0.203	0.001
no capability (level 1)	0.001	0.217	0.000
Pseudo R^2	0.0739	0.0184	0.1139
Enjoyment			
a lot of capability (level 3)	0.181	0.209	0.332
a little capability (level 2)	0.004	0.062	0.003
No capability (level 1)	0.003	0.084	0.001
Pseudo R^2	0.0649	0.0179	0.0985
Control			
a lot of capability (level 3)	0.219	0.564	0.006
a little capability (level 2)	0.000	0.054	0.000
No capability (level 1)	0.002	0.091	0.000
Pseudo R^2	0.1030	0.0238	0.2143

6.5 DISCUSSION

This chapter explored the predictive ability of an instrument capturing capability well-being (ICECAP-O) from a condition-specific health status questionnaire (WOMAC) for arthritis patients requiring surgery on their affected knee or hip. The mapping results produced in Table 9 shows that WOMAC categories (pain, stiffness and physical function) are the best predictors of overall ICECAP-O score (Model 4), with the lowest prediction error statistics (MAE = 0.0832 and RMSE = 0.1193) as well as the highest goodness-of-fit statistic (R^2 = 0.3976). The secondary analysis investigated the prediction of ICECAP-O dimension responses from the WOMAC categories and found that the control dimension and the physical function score produced the highest goodness-of-fit (R^2=0.2143). Physical function was able to predict the lower two levels of four of the five capability attributes (p-value<0.05), when using full capability as the base case in ML regressions. The stiffness category was most closely related to the security dimension on ICECAP-O, while pain was able to predict lower levels of role, enjoyment and control (p-value<0.05). The attachment dimension on ICECAP-O had no significant relationship with any WOMAC category.

This is the first time the prediction of capability well-being from a condition-specific health questionnaire has been explored. For OA patients, it shows that all categories on the WOMAC index are related to ICECAP-O dimensions as shown in the secondary analysis (see Table 11). The preferred mapping model (Model 4, Table 10) allows ICECAP-O scores to be predicted from WOMAC category scores, when ICECAP-O has not been collected within a study.

There are limitations associated with this research. First, the dataset employed here is relatively small for mapping studies, although smaller datasets have been applied previously (Brazier et al., 2010). The mapping model validations were limited to internal validation using follow-up data. This has been the approach for similar mapping work between WOMAC and health utility in the past (Barton et al., 2008). However, individual responses over time were not nested like more advanced mapping methods (Rivero-Arias et al., 2010), so error predictions are likely to be slightly underestimated. Additionally, the internal validation dataset applied was not large enough to capture differences between item scores for both instruments. When a larger sample size is available, it would be of interest to explore the "probabilistic mapping" approach which has been shown to produce lower prediction errors for both OLS and ML regressions in a recent mapping study between the SF-12 and EQ-5D (Le & Doctor, 2011). Finally, the importance of external validation models has recently been highlighted again in the mapping literature (Dakin et al., 2013b). No external validation dataset was available here. However, the importance of assessing external validity when such information is available is recognised.

No other study has attempted to predict a statistical relationship between capability and a condition-specific health status questionnaire. Other mapping studies between WOMAC and health utility instruments have been undertaken, which were detailed in Section 6.3.2. While these studies are not directly comparable to the results obtained here, the R^2 of 0.3976 from the preferred model (Model 4) has similar explanatory power as those other studies that predicted from condition-specific to health utility previously. ICECAP-O research has primarily focused on the validation of the questionnaire in different countries and population groups (Coast et al., 2008b; Flynn et al., 2011; Makai et al., 2012; Couzner et al., 2013a). One study has assessed the relationship between EQ-5D and ICECAP-O for quality of care transition patients and found a positive correlation between the health utility instrument and the measure of capability well-being (Couzner et al., 2012). Another study aimed to address whether ICECAP-O could be used as a substitute or complement for health utility measures, specifically the EQ-5D, for elderly individuals attending a falls prevention clinic (Davis et al., 2013). Whilst the research presented here contradicts their overall finding that ICECAP-O only provides complementary information to health utility, as it has been shown that there is a relationship between change in health and capability (see Table 11), a better test, as Davis and colleagues (2013) have suggested, would be to track longitudinal changes in both capability and health utility instruments simultaneously.

For clinicians and policymakers interested in measuring broader individual well-being rather than disease-specific or generic health questionnaires, based on this research the ICECAP-O can capture the WOMAC categories of pain, stiffness and physical function. A mapping algorithm is provided to generate ICECAP-O scores from WOMAC category scores for OA patients requiring joint replacement. However, given the relatively small sample size employed here, caution is recommended in interpreting capability outcomes solely from this algorithm. Whilst mapping is useful when direct data are unavailable, it remains a second-best option for capability questionnaires. Given the multi-faceted influences on individual well-being and the fact that some elements of the ICECAP-O, the attachment attribute in particular, were not related to WOMAC categories, ICECAP-O data should ideally be collected directly.

CHAPTER 7. SUFFICIENT CAPABILITY

7.1 INTRODUCTION

Previously this thesis has dealt with the role of the capability approach (Chapter 3) in comparison to standard practice within health economics (Chapter 2), as well as exploring the relationship between condition-specific health status and capability for osteoarthritis (OA) patients by mapping from the WOMAC Osteoarthritis index to the ICECAP-O Capability index (Chapter 6). Much of the work within the capability approach in health has focused on theoretical justification (Anand & Dolan, 2005; Coast et al., 2008c; Ruger, 2010a; Smith et al., 2012) and development of capability instruments (Grewal et al., 2006; Anand et al., 2009; Al-Janabi et al., 2012a). Less attention has been paid to date to the application of such instruments. The literature review of empirical capability applications was presented in Chapter 4 to explore the use of decision making rules when using capability outcomes. In this chapter, aspects of the capability empirical literature review are drawn upon to develop a specific methodology for assessing capability outcomes in health economic evaluations.

Two aspects of the capability approach may influence the theoretical framework for economic evaluations. The first is to use a broader definition of individual well-being focusing on capability such that benefits of interventions that go beyond health alone are included in the evaluation. This question was, in part, addressed in Chapter 6 where it was found that condition-specific categories of pain, stiffness and physical function were able to predict the majority of the ICECAP-O attributes (excluding the attachment attribute). As already discussed in Chapter 3 and 4, a number of areas have been identified where adopting the capability approach for health economics would be potentially worthwhile (Kinghorn, 2010;

Lorgelly et al., 2010a; McAllister et al., 2012; Netten et al., 2012; Payne et al., 2013; Simon et al., 2013). The aim of adopting a broader measurement of individual well-being to assist in resource allocation in healthcare decision-making should focus on a fair assessment across the health and related services, so that adopting the capability perspective should not disadvantage interventions which significantly improve individual well-being through improved health.

The second aspect of the practical applications of the capability approach that may influence the theoretical framework for economic evaluations is to consider alternative decision rules that give greater weight to distributional concerns in relation to the relief of (capability) deprivation. In terms of resource allocation, the question most associated with the capability approach and Amartya Sen's "*Equality of what?*" is not the only consideration for health economics. Of equal importance in an evaluation setting is the further question, also recognised by Sen:

> *"Corresponding to 'equality of what?', there is, in fact, also the question: 'efficiency of what?'"* (Sen, 1993, p. 50)

In this chapter the focus is on the second aspect of the practical applications of the capability approach, exploring the use of alternative decision rules derived from work within the capability approach on multidimensional poverty. More specifically, the aim is to develop a method which answers the question "Efficiency of What?" that is in line with the theoretical underpinnings of the capability approach. It was discovered in the capability literature review in Chapter 4 that capability maximisation was rarely the rationale for studies. Alternatively, a threshold approach to poverty reduction, that is, the concern with a sufficient level of well-being to live a valuable life, was more commonly implemented. Therefore, this chapter focuses on incorporating a capability instrument that is developed for health and social evaluations and sits within a threshold approach to reducing capability deprivation. This chapter draws on both health economics (Chapter 2) and capability (Chapter 3 and 4) methodology to develop a new method for aiding resource allocation decisions, which will from here onwards be referred to as the "sufficient capability approach".

The remainder of the chapter is structured as follows. The capability questionnaire and the dataset used in this chapter are first detailed. Then, the method for aggregating capability data utilised by a number of the capability applications in Chapter 4, known as the Alkire-Foster (AF) methods of multidimensional poverty, is further emphasised. The AF methods are the main source for applying the sufficient capability approach. How sufficient capability can be calculated, as well as combining a sufficient capability score with time are then elaborated. Section 7.3 reports the AF methods and sufficient capability results from the dataset used to illustrate the calculations in practice. The chapter concludes with a discussion of the methods presented.

7.2 METHODS

7.2.1 Measure of Capability: ICECAP-O

The ICECAP-O is a five part questionnaire, aimed to capture capability well-being for the older (65 and over) population. The measure has been already described in detail in Chapter 3 (see Appendix 1 for the ICECAP-O questionnaire).

7.2.2 Dataset

A small dataset from the clinical orthopaedic area of joint replacement is used to illustrate the potential of the sufficient capability approach to be used within a clinical context. The dataset is a subset of the Tayside Joint Replacement cohort (Pollard et al., 2009). Between September 2006 and June 2007, 107 patients about to undergo primary joint replacement surgery at Ninewells Hospital, Dundee, UK, were asked to complete the ICECAP-O questionnaire. Follow-up post-operative data were collected at both one and three years after baseline. Although a relatively small dataset, it is the first context in which pre-intervention and post-intervention data are available for any of the ICECAP questionnaires and will provide adequate information to illustrate the sufficient capability approach. This was the same dataset used for the mapping conducted in Chapter 6.

7.2.3 Alkire-Foster method of multidimensional poverty measures

A recent development within the capability literature is the use of multidimensional poverty indices (MPIs) within the fields of human development and international poverty assessment. The MPIs are taken from a methodology called the Alkire-Foster (AF) measures (Alkire & Foster, 2011a). The AF methodology was explained in detail in Chapter 4.

Since 2010, MPIs have been collected for cross country comparisons of multidimensional poverty by the United Nations (UN) and are reported in their human development reports, by measuring three dimensions (health, education and living standards) across ten indicators of states of poverty (Klugman, 2010). This notion of broadening the assessment of poverty, rather than relying on a single indicator (e.g. income measured by Gross Domestic Product

(GDP)), follows from the approach used in constructing the Human Development Index (HDI). This multidimensional approach is not unique to the UN, as the Organisation for Economic Co-operation and Development (OECD) Better Life Initiative also draws on multiple (eleven) indicators, including health, to develop a better life index as another alternative to GDP (Stiglitz et al., 2009).

Although the AF methods are proving popular within the human development literature, they have, to date, only been analysed using cross-sectional data for cross-country comparisons (Klugman, 2010). Additionally, new areas were identified in the empirical capability review in Chapter 4 where the AF methods are used such as freedom poverty (Callander et al., 2012b), child poverty (Roelen et al., 2010), and energy poverty indices (Nussbaumer et al., 2012). Once more, these new applications of the AF methods rely on cross-sectional data. In Figure 11, a simple example of how the AF measures are used in practice is presented. To be able to utilise this approach within health to help guide resource allocation decisions, an additional refinement is required.

7.2.4 Threshold of Sufficient Capability

One possible approach to measuring capability poverty using the ICECAP-O instrument is to follow the AF methodology (Alkire & Foster, 2011a) and capability objectives as set out in the previous section. This sets a minimum threshold level of capability that a person must achieve to be considered to have a "sufficient" level of capability wellbeing. This minimum level here is called the threshold of sufficient capability (TSC) and is defined as the level of capability at or above which a person's level of capability wellbeing is no longer a concern for policy.

Figure 11 Example of Alkire-Foster measures of multidimensional poverty

Take 3 individuals (X,Y & Z) assessed across 4 dimensions (D1,D2,D3,D4) which indicate poverty.
All 4 dimensions are categorical with 5 (1-5) responses possible for each dimension.

	Dimension 1	Dimension 2	Dimension 3	Dimension 4
Individual X	2	3	4	5
Individual Y	3	3	3	3
Individual Z	5	1	4	2

Let **poverty line** across 4 dimensions fall at **response level 3**.
Let **cutoff** for an individual to be classified as poor fall below poverty line on **any one dimension**.
Let **values** for levels below poverty line for **level 1 = 1; level 2 = 0.3.**

(1) **Headcount Ratio (H) = p/P**
p = individual X (poor on dimension 1) & individual Z (poor on dimension 2 & 4) =2
P = total sample size = 3
H = 2/3 = 0.667

(2) **Adjusted Headcount Ratio (M_0) = H×(d_p/D)**
Individual X d_p = 1 out of 4 dimensions
Individual Z d_p = 2 out of 4 dimensions
d_p = 3 out of 8 dimensions
M_0 = 0.667×(3/8) = 0.250

(3) **Adjusted Poverty Gap (M_1) = M_0×(l_p/L)**
Individual X l_p = 1 out of 8 levels below poverty line
Individual Z l_p = 3 out of 8 levels below poverty line
l_p = 4 out of 16 levels below poverty line
M_1 = 0.250×(4/16) = 0.063

(4) **Adjusted FGT (M_2) = M_0×v_p/V**
Individual X v_p = 0.3 out of 4 for lowest value attached across all dimensions
Individual Z v_p = 1.3 (1+0.3) out of 4 for lowest value attached across all dimensions
v_p = 1.6 out of 8 values attached to lowest dimensions for poor, p
M2 = 0.250×(1.6/8) = 0.050

The threshold for sufficient capability (TSC) is defined as:

$$TSC = td_1+td_2+...td_x=>k, \text{ where } k = [td_1...td_n] \tag{7.1}$$

Here td is the threshold level for each dimension that is "sufficient" which depends on k, the cutoff number of dimensions to be considered poor, must fall within the range of the total number of dimensions measured [$td_1...td_n$].

A person who has reached a level of capability across all attributes which is deemed 'sufficient' will have reached an adequate level of wellbeing, indicating that further allocation of resources to this individual is no longer a priority. A person who does not reach the sufficient level on the attributes examined falls below the overall threshold depending on the cutoff in dimensions (k) to be considered in poverty. To improve the level of sufficient capability within a population, the aim of using the threshold of sufficient capability is to move as many people to sufficient levels of capability or as close to the TSC as possible.

To apply this approach in practice, the first fundamental step in defining the TSC using the AF measures is the "identification method" (see Section 4.5.4.1). This requires determining the threshold level on each dimension below at which there is considered to be a shortfall in sufficient capability. For the capability measure employed here, the ICECAP-O (see Table 12) has four levels of capability for each of its five attributes, conceptually ranging through full capability (level 4) , a lot of capability (level 3), a little capability (level 2) and no capability (level 1) (Coast et al. 2008a). A vast number of different thresholds could be implemented in theory, given that there is no need for the level in which a person is considered to be in capability poverty to be consistent across attributes.

Here, two possibilities are considered for ease of interpretation:

- Option 1: assuming that if a person has at least 'a lot' of capability (i.e. level 3) on each attribute they have sufficient capability ***("33333")***.
- Option 2: assuming that if a person has at least 'a little' capability (i.e. level 2) on each attribute they have sufficient capability ***("22222")***.

Table 12 Thresholds of Sufficient Capability (TSC) on the ICECAP-O

ICECAP-O dimension	ICECAP-O Values	TSC "33333"	TSC "22222"
Attachment (Love and Friendship)			
(4) I can have all of the love and friendship that I want	0.2535	0.2679	0.2412
(3) I can have a lot of the love and friendship that I want	0.2325	0.2679	0.2412
(2) I can have a little of the love and friendship that I want	0.1340	0.1545	0.2412
(1) I cannot have any of the love and friendship that I want	-0.0128	-0.0147	-0.0230
Security (Thinking about the future)			
(4) I can think about the future without any concern	0.1788	0.1234	0.1189
(3) I can think about the future with only a little concern	0.1071	0.1234	0.1189
(2) I can only think about the future with some concern	0.0661	0.0761	0.1189
(1) I can only think about the future with a lot of concern	0.0321	0.0370	0.0578
Role (Doing things that make me feel valued)			
(4) I am able to do all of the things that make me feel valued	0.1923	0.2066	0.2332
(3) I am able to do many of the things that make me feel valued	0.1793	0.2066	0.2332
(2) I am able to do a few of the things that make me feel valued	0.1296	0.1494	0.2332
(1) I am unable to do any of the things that make me feel valued	0.0151	0.0174	0.0272
Enjoyment (Enjoyment and Pleasure)			
(4) I can have all of the enjoyment and pleasure that I want	0.1660	0.1893	0.2132
(3) I can have a lot of the enjoyment and pleasure that I want	0.1643	0.1893	0.2132
(2) I can have a little of the enjoyment and pleasure that I want	0.1185	0.1365	0.2132
(1) I cannot have any of the enjoyment and pleasure that I want	0.0168	0.0193	0.0302
Control (Independence)			
(4) I am able to be completely independent	0.2094	0.2129	0.1936
(3) I am able to be independent in many things	0.1848	0.2129	0.1936
(2) I am able to be independent in a few things	0.1076	0.1240	0.1936
(1) I am unable to be at all independent	-0.0512	-0.0590	-0.0922

TSC "33333" & "22222" - Threshold levels of sufficient capability

The original values of ICECAP-O are then re-scaled so that 1 is equal to the threshold of sufficient capability, so that the new objective only gives priority to those below sufficient capability. The formula for calculating values for each threshold dimension is presented below:

$$V_{tx} = \frac{v_{tx}}{\Sigma_n(v_{t1} \ldots v_{tn})} \tag{7.2}$$

Here V_{tx} = new value on threshold dimension level x, v_{tx}= original value of the threshold dimension level x, $\Sigma_n(v_{t1} \ldots v_{tn})$ = sum of threshold value levels across all dimensions before transformation (i.e., original ICECAP-O values).

(*e.g. from Table 12, level 3 on the attachment attribute for threshold Option 1 "33333", $v_{tx} \approx 0.2325$; $\Sigma[v_{t1}+v_{t2}+\ldots v_{tn}]$ when TSC "33333" ≈ 0.868; thus $V_{tx} \approx 0.2679$*)

Therefore, any scores higher than the sufficient capability threshold for each option will have the equivalent value of the sufficient capability threshold value. Any shortfalls in capability below this threshold are then allocated a shortfall value according to both:

- ***the extent of that shortfall*** (whether at the level of 'no capability' or 'a little capability' for option 1; not applicable for option 2 as only one level of 'no capability' below threshold) and;
- ***the rescaled ICECAP-O population values***. The ICECAP-O general population based value set is additive and on a linear scale, such that the numerical value is meaningful and the values across all attributes can be summed to give an overall index between 0 - representing no capability and 1 - representing full capability. For options 1 and 2, the index score of 1 will now represent TSC in each scenario depending on each threshold level respectively, i.e. a value of 1 represents sufficient capability.

7.2.5 Sufficient Capability Score

Once the threshold has been determined, values for levels below the threshold need to be calculated to reflect societal values of these states of capability well-being. This is required to apply the most complex of the AF measures of multidimensional poverty methods (the adjusted-FGT or M2, see Figure 11). For values below the threshold for a dimension, the following method is used to calculate the capability value compared with the threshold level:

$$V_x = \frac{v_x}{\Sigma v_{t1}+v_{t2}+\ldots v_{tn}} \text{ ; unless } v_x > v_{tx,} \text{ then } V_x \equiv V_{tx} \text{ for TSC} = 1 \qquad (7.3)$$

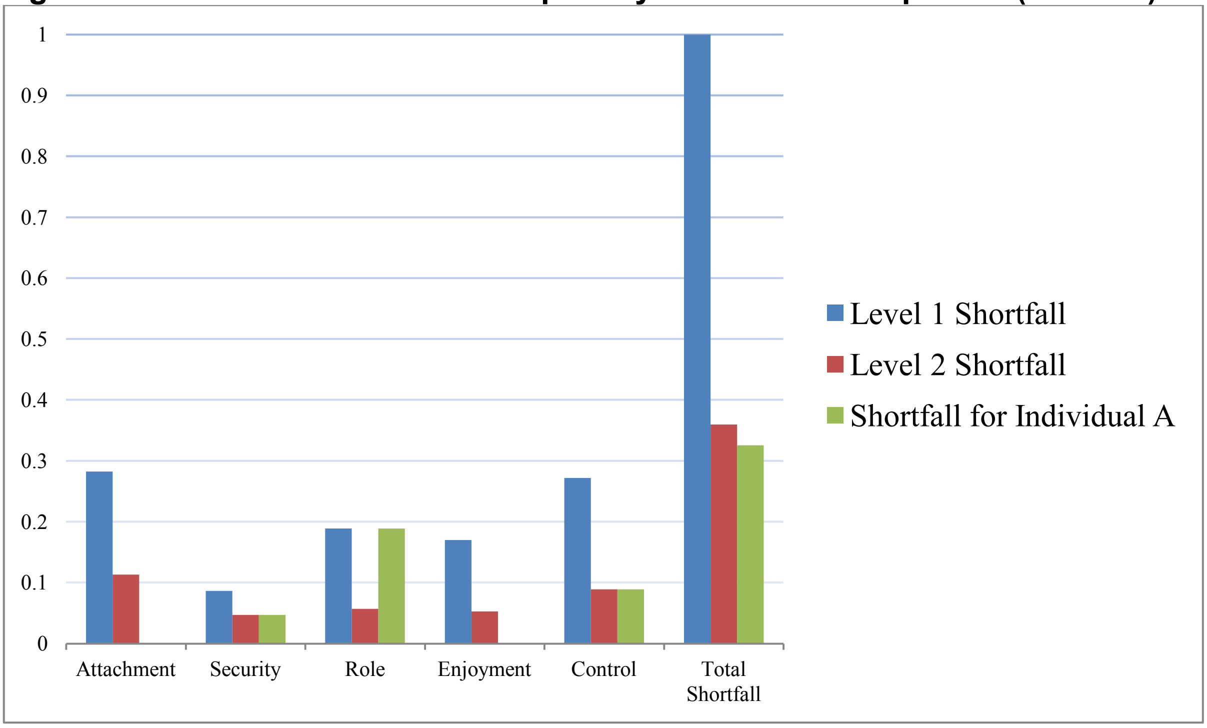
This calculation generates a new 0-1 value scale for ICECAP-O, with 0 still reflecting the “no capability” response levels on all ICECAP-O attributes. However, 1 no longer reflects “full capability” across all attributes, instead this reflects the threshold of sufficient capability (TSC). To calculate an overall reflection of sufficient capability, individual responses are summed across this new scale to calculate an individual’s Sufficient Capability Score (SCS). To calculate SCS for an individual, the values attached for each threshold option proposed in the previous section are presented in Table 12. For example, using threshold option 1 “33333” as the level where sufficient capability is reached, an ICECAP-O profile of “43233” has the same SCS score as an ICECAP-O profile of “44244”. In Table 13 and Figure 12, an individual example of how the sufficient capability score (SCS) is calculated for a given threshold (Option 1 – “33333”) for an individual.

Table 13 Shortfall in Sufficient Capability on ICECAP-O for individual A; (threshold Option 1 "33333")

1. Love and Friendship			Tick one box only in each section
I can have all of the love and friendship that I want	✓	4	
I can have a lot of the love and friendship that I want		3	
I can have a little of the love and friendship that I want		2	
I cannot have any of the love and friendship that I want		1	
2. Thinking about the future			
I can think about the future without any concern		4	
I can think about the future with only a little concern		3	
I can only think about the future with some concern	✓	2	
I can only think about the future with a lot of concern		1	
3. Doing things that make you feel valued			
I am able to do all of the things that make me feel valued		4	
I am able to do many of the things that make me feel valued		3	
I am able to do a few of the things that make me feel valued		2	
I am unable to do any of the things that make me feel valued	✓	1	
4. Enjoyment and pleasure			
I can have all of the enjoyment and pleasure that I want		4	
I can have a lot of the enjoyment and pleasure that I want	✓	3	
I can have a little of the enjoyment and pleasure that I want		2	
I cannot have any of the enjoyment and pleasure that I want		1	
5. Independence			
I am able to be completely independent		4	
I am able to be independent in many things		3	
I am able to be independent in a few things	✓	2	
I am unable to be at all independent		1	

Individual A ICECAP-O profile (42132); Highlight Green = sufficient capability for given attribute and given threshold. Highlight Red = shortfall in sufficient capability for given attribute for given threshold.

Figure 12 Shortfalls in Sufficient Capability for Threshold Option 1 ("33333")



Sufficient Capability Score (SCS) = Sufficient Capability (1) – Total Shortfall (0-1). Shortfall for individual A (42132) for threshold option 1 ("33333") = 0.325; SCS for individual A = 0.675. Capability instrument, ICECAP-O .

7.2.6 Sufficient Capability Over Time

Whilst improvement in individual wellbeing is important, another key calculation in health evaluations is to consider both wellbeing (however defined) and changes in wellbeing over time. This is something which has not been tackled in detail within the capability literature (as emphasised in Chapter 4) and has been identified as an issue for practical evaluations (Alkire et al., 2008). This aspect of the sufficient capability approach can, however, draw on the current methods applied to generate health economics outcomes, like QALYs.

At this stage it is important to note the different anchors used on HRQoL measures such as EQ-5D and the ICECAP-O capability instrument. In general terms, extra-welfarist HRQoL measures are anchored on a 0-1 being dead to full health scale, where it is possible to have states worse than dead depending on the valuation method used (this was discussed in greater depth in Chapter 2). Similarly, the ICECAP-O is anchored on a 0-1 scale, although the anchors for ICECAP-O are no capability to full capability (Coast et al., 2008a). The ICECAP-O is anchored differently to HRQoL measures used to produce QALYs. For that reason, to incorporate the ICECAP-O within a QALY would be inconsistent with the current anchoring system used to generate a health QALY. The zero value on the ICECAP-O index can be interpreted as:

"A number of states may produce such a zero value: assessment of capabilities as being non-existent in relation to all attributes; unconsciousness; and death" (Coast et al., 2008a, p.878)

SCS is a flexible measure which can be applied to maximise capability levels for a given threshold, or inversely minimise shortfalls from the sufficient capability threshold. Therefore, similar approaches to those used for the QALY as a health maximising outcome or the Disability Adjusted Life Years (DALYs) as a disease burden minimising outcome (Murray & Lopez, 1996), could be applied in practice. Three examples of outcomes maximising to sufficient capability and one example of a minimising capability deprivation outcome are explored further next.

7.2.6.1 Poverty Free Years (capability)

The first outcome that is considered here combines the first Alkire-Foster method, the headcount ratio, over time, to calculate a measure of prevalence of poverty within a given population over time. The headcount ratio (H) treats an individual who is below the threshold of sufficient capability on one attribute the same as someone who is capability deprived on all attributes for a given measure. Alkire and Foster (2011a) argue that their other methods of multidimensional poverty are more appropriate to capture the broader influences of deprivation. Thus, the poverty free years (capability) (PFY(c)) outcome will act as a simple method of separating those below a given threshold, with those who have reached sufficient capability. Equation 7.4 is the appropriate calculation for PFY(c)

$$PFY(c) = (1-H)\times T \tag{7.4}$$

where H = headcount ratio and T=time in that state

7.2.6.2 Years of Sufficient Capability (equivalent)

To combine SCS over time for threshold options in Section 7.2.4, Years of Sufficient Capability (YSC) are generated to give a longitudinal measure representing gains in sufficient capability over time. The calculation is represented in Equation 7.5 below:

$$YSC = SCS\times T \tag{7.5}$$

where SCS = Sufficient Capability Score and T = Time

7.2.6.3 Years of Insufficient Capability (equivalent)

To measure shortfalls in sufficient capability over time, Years of Insufficient Capability (YIC) are measured by the combining the inverse of the Sufficient Capability Score (SCS) with time. The calculation of YIC is represented in Equation 7.6

$$YIC = (1-SCS)\times T \tag{7.6}$$

where SCS = the Sufficient Capability Score and T=Time

7.2.6.4 Years of Full Capability (equivalent)

Whilst thresholds of sufficient capability have been suggested to be below the highest possible levels, it may very well be the case that sufficient capability is represented at the top level across all capability dimensions for a given questionnaire. If this is the scenario, then the approach required is comparable to the current health maximisation outcome objective. However, since the values associated with the QALY are generally calculated with a trade-off between quality and quantity of life, it would be misleading to call a similar capability outcome using the ICECAP-O capability valuation dataset a QALY. Therefore, the term Years of Full Capability (YFC) here is used to calculate the maximisation of capability across the population with no time preference. Equation 7.7 below presents the YFC calculation

$$YFC = ICECAP\text{-}O \times T, \text{ where ICECAP-O = original ICECAP-O values and } T = \text{Time} \qquad (7.7)$$

7.3 RESULTS

In this section, the methods of applying the sufficient capability approach are tested on a sample of patients who completed the ICECAP-O at three points. The demographics of the population are presented in Section 7.3.1. In Section 7.3.2, the four Alkire Foster (AF) methods are tested for this dataset. The AF methods were explained in Chapter 4 and an example of how to calculate the methods was previously presented in Figure 11 in this Chapter. The threshold options for calculating a sufficient capability score are then tested in Section 7.3.3. Finally, the four sufficient capability outcomes are tested to see if the different outcomes lead to different changes in capability over time.

7.3.1 Demographics

Table 14 summarises the demographics for the Tayside replacement cohort. At baseline, the average age of this population was 69.27 years. The ICECAP-O scores at baseline for the 106 patients who completed the ICECAP-O was 0.773, less than the ICECAP-O average from the general population valuation dataset (0.815) (Coast et al., 2008a) and from a more recent, larger general population sample (0.832) (Flynn et al., 2011).

Table 14 Baseline Descriptive Statistics for Tayside replacement dataset

	N	n	Mean	SD	Min	Max
Sample size	107					
Missing data	106	1				
Males	106	55				
Employed	106	13				
Living Alone	106	28				
Age (mean)	106	106	69.720	8.854	48.000	89.000
ICECAP-O (base)	106	106	0.773	0.167	0.159	1.000
ICECAP-O (1 year)	106	58	0.862	0.132	0.516	1.000
ICECAP-O (3 year)	106	55	0.832	0.138	0.481	1.000
ICECAP-O complete (base)	42	42	0.789	0.132	0.368	0.998
ICECAP-O complete (1 year)	42	42	0.851	0.134	0.516	1.000
ICECAP-O complete (3 year)	42	42	0.824	0.146	0.481	1.000

N,total population; n,sub population from N; SD, standard deviation; min, minimum; max, maximum

7.3.2 Alkire-Foster measures

In the first analysis, the results for patients who completed the ICECAP-O at all three time-points (n=42) are presented for both threshold options at “a lot of capability 33333” and “a little capability 22222”. The Alkire-Foster methods are calculated at all possible cutoffs (k) for the three time periods for both threshold options. Table 15 presents the AF measures for the “33333” threshold and Table 16 presents the AF measures for the “22222” threshold.

Table 15 AF poverty methods applied to ICECAP-O levels using "33333" as the poverty threshold (n=42)

33333	BASELINE						1-YEAR POST-OP						3-YEAR POST-OP				
CUTOFF(K)	K=1	K=2	K=3	K=4	K=5		K=1	K=2	K=3	K=4	K=5		K=1	K=2	K=3	K=4	K=5
AF						**AF**						**AF**					
H	0.7381	0.5000	0.3333	0.1667	0	**H**	0.3571	0.3095	0.1905	0.1190	0.0238	**H**	0.4762	0.3333	0.2619	0.1905	0
M_0	0.3476	0.3000	0.2333	0.1333	0	**M_0**	0.2000	0.1905	0.1429	0.1000	0.0238	**M_0**	0.2524	0.2238	0.1952	0.1524	0
M_1	0.0931	0.1029	0.0950	0.0647	0	**M_1**	0.0627	0.0659	0.0607	0.0480	0.0143	**M_1**	0.0782	0.0879	0.0870	0.0686	0
M_2	0.0610	0.0698	0.0655	0.0433	0	**M_2**	0.0431	0.0449	0.0417	0.0338	0.0095	**M_2**	0.0519	0.0603	0.0595	0.0445	0

AF, Alkire and Foster multidimensional poverty methods; k, cutoff in number of dimensions for individuals to be poor; H, headcount ratio; M_0, adjusted headcount ratio; M_1, adjusted poverty gap; M_2, adjusted-Foster-Greer-Thorbecke measure. More details on AF methods in Figure 11

Table 16 AF poverty methods applied to ICECAP-O levels using "22222" as the threshold (n=42)

22222	BASELINE						1-YEAR POST-OP						3-YEAR POST-OP				
CUTOFF(K)	K=1	K=2	K=3	K=4	K=5		K=1	K=2	K=3	K=4	K=5		K=1	K=2	K=3	K=4	K=5
AF						**AF**						**AF**					
H	0.1667	0.0476	0.0238	0	0	**H**	0.1190	0	0	0	0	**H**	0.1667	0.0238	0.0238	0	0
M_0	0.0476	0.0238	0.0143	0	0	**M_0**	0.0238	0	0	0	0	**M_0**	0.0429	0.0143	0.0143	0	0
M_1	0.0476	0.0238	0.0143	0	0	**M_1**	0.0238	0	0	0	0	**M_1**	0.0429	0.0143	0.0143	0	0
M_2	0.0100	0.0095	0.0079	0	0	**M_2**	0.0025	0	0	0	0	**M_2**	0.0073	0.0064	0.0064	0	0

AF, Alkire and Foster multidimensional poverty methods; k, cutoff in number of dimensions for individuals to be poor; H, headcount ratio; M_0, adjusted headcount ratio; M_1, adjusted poverty gap; M_2, adjusted-Foster-Greer-Thorbecke measure. More details on AF methods in Figure 11

From both Tables 15 and 16, it is clear that the choice of threshold and cutoff is crucial in measuring the level of poverty in terms of sufficient capability for a given population. Using the simplest AF measures, the Headcount Ratio (H), and the cutoff (k) = 1, comparing Table 15 and 16 shows that while almost three quarters (73.8%) of the population have shortfalls in sufficient capability at the TSC of "33333", less than one fifth (16.7%) of the population have shortfalls at the lower threshold of "22222".

7.3.3 Sufficient Capability for Threshold Option 1 (33333)

Overall, the SCS for the baseline population (n=106) threshold option 1 (33333) is 0.857. This is lower than the average of the sample from the ICECAP-O valuation dataset, which has an average SCS score of 0.894 with the same TSC (see Table 17 for ICECAP-O valuation dataset responses below threshold. Data from Coast et al. 2008a). Table 18 shows the attributes in which shortfalls in sufficient capability occurred pre-intervention, with the "enjoyment" attribute reporting the highest number of shortfall responses below the "33333" threshold (46.23%).

Table 17 ICECAP-O valuation dataset responses below Threshold of Sufficient Capability (n=313)

ICECAP-O	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL
level 2	8.95%	30.03%	23.00%	21.09%	17.89%
level 1	3.83%	12.46%	3.51%	4.15%	2.24%

ICECAP-O original value average, 0.815; SCS (33333), 0.894; SCS (22222), 0.961; level 1, 'no' capability on ICECAP-O; level 2, 'a little' capability on ICECAP-O

Table 18 Responses below Threshold of Sufficient Capability "33333"(baseline)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 2	9	29	38	38	27	26.60%
level 1	2	12	7	11	6	7.17%
Below TSC per attribute	10.38%	38.68%	42.45%	46.23%	31.13%	

n = 106; SCS, Sufficient Capability Score; TSC, Threshold of Sufficient Capability; SCS = 0.857; TSC "33333"

In Table 19, SCS is calculated for patients who completed the ICECAP-O at one year post-operation (n=58). This resulted in an increased SCS score of 0.05 (0.88→0.93) from baseline for these individuals. Patients who completed ICECAP-O at three year post intervention (n=55) also reported an improved SCS from baseline by 0.031 (0.881→0.912), which can be seen in Table 20. In Tables 19 and 20, the improvements in the lower levels of capability come predominantly from the "role" and "enjoyment" ICECAP-O attributes.

Table 19 Responses below Threshold of Sufficient Capability "33333" (baseline and 1 year post operation; n=58)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 2	3	15	21	22	14	25.86%
level 1	1	6	2	5	1	5.17%
Responses below TSC per attribute	6.90%	36.21%	39.66%	46.55%	25.86%	

ICECAP-O responses (1 year)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 2	10	8	11	15	4	16.55%
level 1	0	6	0	0	1	2.41%
Responses below TSC per attribute	17.24%	24.14%	18.97%	25.86%	8.62%	

SCS, Sufficient Capability Score; TSC = Threshold of Sufficient Capability; SCS (baseline "33333") = 0.88; SCS (1 year "33333") = 0.93.

Table 20 Responses below Threshold of Sufficient Capability "33333" (baseline and three year post-operation; n=55)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	Total Responses
level 2	3	17	22	22	13	28.00%
level 1	0	5	1	4	2	4.36%
Responses below TSC per attribute	5.45%	40.00%	41.82%	47.27%	27.27%	

ICECAP-O responses (3 year)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	Total Responses
level 2	4	16	13	11	10	19.64%
level 1	0	5	3	1	1	3.64%
Responses below TSC per attribute	7.27%	38.18%	29.09%	21.82%	20.00%	

SCS, Sufficient Capability Score; TSC, Threshold of Sufficient Capability; SCS (baseline "33333") = 0.881; SCS (3 year "33333") = 0.912.

7.3.4 Sufficient Capability for Threshold Option 2 (22222)

The proportion of responses below SCS threshold option 2 (22222) at baseline, one year and three year post-intervention can be seen in the level 1 ICECAP-O responses on Tables 21,22 and 23 respectively. Overall, the SCS at baseline population (n=106) is 0.940, which is lower than the average of the sample from the valuation dataset with an average SCS score of 0.961 with the same threshold (see Table 18). The "security" and "enjoyment" attributes have the highest proportion of responses below the "22222" threshold (see Table 21). In Table 22, SCS is calculated for patients who completed the ICECAP-O at one year post-operation (n=58). This resulted in an increased SCS score of 0.028 (0.961→0.989). Patients who completed ICECAP-O at three years post intervention (n=55) also reported an improved SCS from baseline of 0.008 (0.967→0.978), which can be seen in Table 23. The "enjoyment" attribute showed the greatest improvement in the follow-up years.

Table 21 Responses below Threshold of Sufficient Capability “22222”(baseline)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 1	2	12	7	11	6	7.17%
Below TSC per attribute	1.89%	11.32%	6.60%	10.38%	2.83%	

n = 106; SCS, Sufficient Capability Score; TSC, Threshold of Sufficient Capability; SCS = 0.940; TSC "22222"

Table 22 Responses below Threshold of Sufficient Capability “22222” (baseline and 1 year post operation; n=58)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 1	1	6	2	5	1	5.17%
Responses below TSC per attribute	1.72%	10.34%	3.45%	8.62%	1.72%	

ICECAP-O responses (1 year)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	TOTAL RESPONSES
level 1	0	6	0	0	1	2.41%
Responses below TSC per attribute	0.00%	10.34%	0.00%	0.00%	1.72%	

SCS, Sufficient Capability Score; TSC = Threshold of Sufficient Capability; SCS (baseline “22222”) = 0.961; SCS (1 year) = 0.989.

Table 23 Responses below Threshold of Sufficient Capability “22222” (baseline and three year post-operation; n=55)

ICECAP-O responses (baseline)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	Total Responses
level 1	0	5	1	4	2	4.36%
Responses below TSC per attribute	0.00%	9.10%	1.82%	7.27%	3.64%	

ICECAP-O responses (3 year)

Below TSC	ATTACHMENT	SECURITY	ROLE	ENJOYMENT	CONTROL	Total Responses
level 1	0	5	3	1	1	3.64%
Responses below TSC per attribute	0.00%	9.10%	5.45%	1.82%	1.82%	

SCS, Sufficient Capability Score; TSC, Threshold of Sufficient Capability; SCS (baseline “22222”) = 0.967; SCS (3 year “22222) = 0.978.

7.3.5. Sufficient Capability Outcomes over time

The four outcomes of sufficient capability [(1) poverty free years (capability), (2) years of sufficient capability, (3) years of insufficient capability, (4) years of full capability are tested over time with this sample population who completed the ICECAP-O at all three time points (n=42). This was used to get an indication of the changes in capability over time for each of the four outcomes proposed here.

7.3.5.1 Poverty Free Years (capability) – (PFY(c))

Poverty Free Years (capability) are calculated both for option 1 (33333) and option 2 (22222). Firstly, option 1 reported a headcount ratio of 0.738 at baseline, which significantly decreased at year 1 (0.358) and year 3 (0.476). Using the area under the curve approach, where the benefits of the intervention are deducted from the benefits of the control group (Drummond et al., 2005), poverty free years (capability) are increased by 0.832 during this time period (i.e. $[3-[0.5(0.738+0.358)+(0.358+0.476)]]-[3-(0.738\times3)] = 0.832$).

As for option 2, the headcount ratio for the arthritis population in capability poverty was considerably less (0.167). At year 1, this reduced to 0.119 but returned to baseline level at year 3. Therefore, the PFY(c) gained from this intervention is less than option 1, with a gain of 0.072 (i.e. $[3-[0.5(0.167+0.119)+(0.119+0.167)]-[3-(0.167\times3)] = 0.072$).

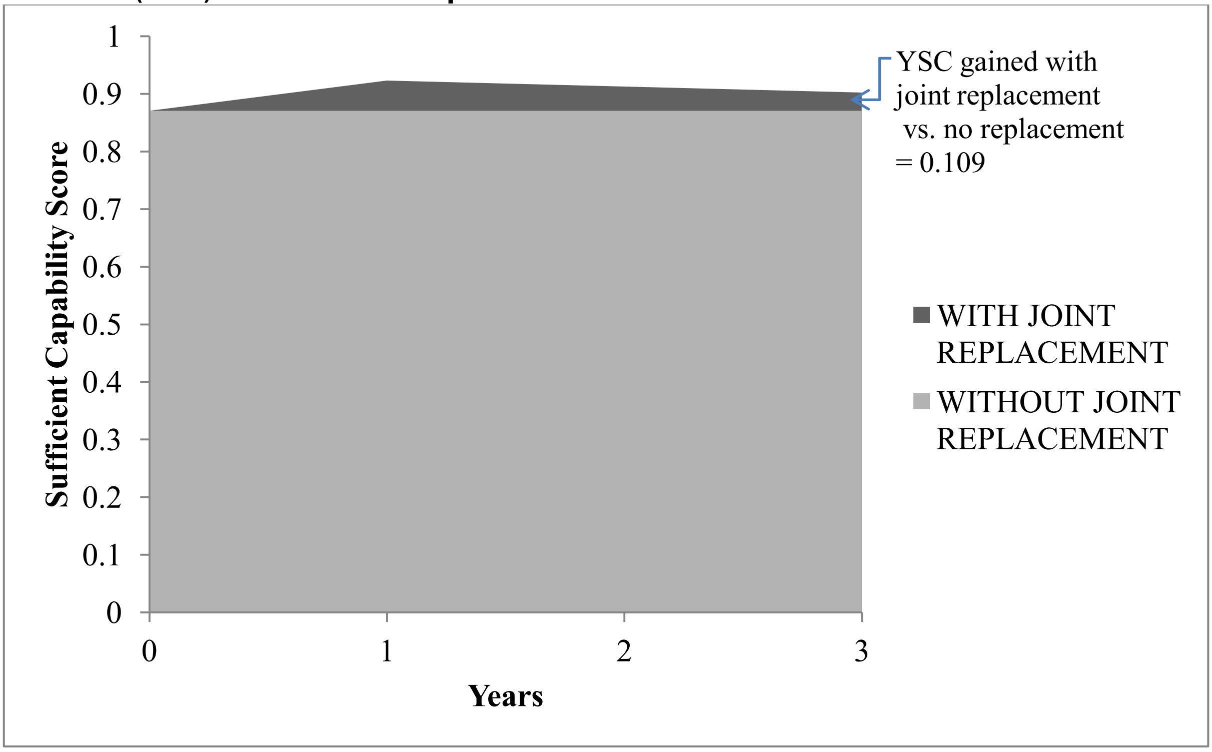
7.3.5.2 Years of Sufficient Capability – (YSC)

This analysis involves calculating the change in SCS over time to generate Years of Sufficient Capability (YSC). YSC is calculated using the area under the curve approach.

For Option 1 (33333) the SCS for the 42 patients at baseline is 0.871, which is assumed to stay constant if the intervention was not provided to the patients. SCS at one year (0.923) and SCS at three year post-intervention (0.902) are used to calculate the intervention group, with each SCS is assumed to be connected linearly. Figure 13 shows this calculation graphically. The darker area on Figure 13 displays the YSC gain from treatment, which is equal to 0.109 YSC gained compared with no intervention provided (i.e. $[0.5(0.871+0.923)+(0.923+0.902)]-[3\times0.871] = 0.109$).

For Option 2 (22222) SCS for the 42 patients at baseline is 0.965, which is assumed to stay constant if the intervention was not provided to the patients. SCS at follow up at one year (0.987) and SCS at three year post-intervention (0.972). Combining these three time-points and comparing results to a constant SCS over the three years results in 0.04 YSC gained compared to no intervention provided (i.e. $[0.5(0.965+0.987)+(0.987+0.972)-[3\times0.965] = 0.04$).

Figure 13 Example of Years of Sufficient Capability (YSC) for Osteoarthritis Patients (n=42) **for threshold option 1 "33333"**



7.3.5.3 Years of Insufficient Capability – (YIC)

The calculation of Years of Insufficient Capability (YIC) is intended to minimise shortfalls in sufficient capability. This is the inverse calculation of YSC. Therefore, for Option 1 (33333) YIC at baseline is 0.129, at the one year follow-up is at 0.077 and at the three year follow-up of 0.098. With the baseline YIC held constant over three years, the YIC saved from the joint replacement is 0.109.

Similarly for Option 2 (22222), YIC for the arthritis patients in this dataset pre-operation is 0.035. YIC at one year (0.013) and three years (0.028) post-operation, results in a YIC saved of 0.040.

7.3.5.4 Years of Full Capability – (YFC)

Years of Full Capability (YFC) are calculated in the same area under the curve approach as for Years of Sufficient Capability (YSC). At baseline, YFC for the arthritis patients are equal to 0.789 (i.e. when the threshold of sufficient capability is equal to 44444 on the ICECAP-O). YFC at year one (0.851) at year three (0.824) follow up result in YFC gained of 0.128 when compared with the baseline score over a three year period (i.e. [0.5(0.789+0.851)+(0.851+0.824)]-[3×0.789] = 0.128).

Table 24 summarises all the different outcomes across both threshold options in this study. While three of the four outcomes produce similar results, the poverty free years (capability) outcome results in a greater improvement. This represented an increase in the population whose capability levels improved to the extent that they met the threshold of sufficient capability following treatment and who were below the threshold at baseline. However, the PFY(c) was calculated using the simplest of the AF methods, the headcount ratio (H). All three other outcomes apply the most complex of the AF methods, the adjusted Foster Greer Thorbecke or M2, which takes account of number of attributes and values of shortfalls below the threshold of sufficient capability.

Table 24 Capability Outcomes for Tayside Replacement Dataset (n=42)

		TSC	Baseline	1-year	3-year	Benefit of intervention (holding baseline constant)
Poverty Free Years (capability)	PFY(c)	33333	0.738	0.358	0.476	0.832
		22222	0.167	0.119	0.167	0.072
Years of Sufficient Capability	YSC	33333	0.871	0.923	0.902	0.109
		22222	0.965	0.987	0.972	0.040
Years of Insufficient Capability	YIC	33333	0.129	0.077	0.098	0.109
		22222	0.035	0.013	0.028	0.040
Years of Full Capability	YFC	44444	0.789	0.851	0.824	0.128

TSC, Threshold of Sufficient Capability on ICECAP-O

7.4. DISCUSSION

This is the first study to formulate a new decision rule based on achieving sufficient capability using a capability questionnaire. This new approach is flexible in how it can be applied in practice, i.e. either maximising in a method similar, although not identical to QALYs, or minimising in a method similar to DALYs, as shown by the development of four new outcomes. However, the objective of "capability efficiency" (Cookson, 2005b) or efficiency in achieving capability needs to be rationalised in association with the theoretical underpinnings of the capability perspective. Maximising total population levels of capability or minimising total levels of capability poverty is not in line with how the capability approach was designed to be implemented in practice, as emphasised in Chapters 3 and 4. This study has therefore adopted methods from the multidimensional poverty literature, closely linked to the capability approach, to develop a practical methodology for generating capability outcomes through what is called here the "sufficient capability approach". The example above highlights the importance of selecting an appropriate threshold of sufficient capability (TSC), as results in terms of the assessment of benefit gained inevitably vary considerably depending on the threshold chosen (as shown in Table 24).

Cookson (2005b) discussed the possibility of aligning the QALY with the capability approach (or what Cookson calls the 'capability QALY'), accounting for both health and non-health functionings that influence capability. Whilst appreciating his meticulous framework, the primary concern with Cookson's approach to incorporating capability (or "capability efficiency") within health economics using the "capability QALY" (Cookson, 2005b).

This Chapter deviates from Cookson's approach by firstly capturing capability directly, although the ICECAP questionnaires were not available at the time of Cookson's proposal. More importantly, his definition of the "efficiency of what?" is not in line with the capability approach, as the sufficient capability approach allows distributional considerations to be handled within the efficiency measure, by improving those with shortfalls in sufficient capability as a priority. Simply re-interpreting the QALY as a 'capability QALY' seems unsatisfactory from a capability perspective, as emphasised in Chapter 4 and the proposal of the sufficient capability approach is offered as an alternative. Four sufficient capability outcomes were developed and calculated for a sample of OA patients to provide an illustration of the approaches.

There are limitations associated with the application of the sufficient capability approach in the example presented in the previous section.

First, the dataset of 107 patients at baseline, and the smaller sample sizes available at follow-up was relatively small. Nevertheless, it was large enough to show significant change in the overall sufficient capability score (SCS) between follow-up periods, in particular for threshold option 1 “33333”, and was useful in illustrating the methods and outcomes of the sufficient capability approach in practice.

Second, and possibly more significant than the small sample size limitation, is the comparison to show how sufficient capability changed and improved over time. There was no control group available whereby an intervention versus no intervention could be compared, which would have given more meaningful results as to whether the sufficient capability outcomes were capturing tangible differences between outcomes and due to the intervention provided to these patients. While there is some evidence which suggests that older age (over 75 years) leads to reduced capability (Flynn et al., 2011), this was based on cross-sectional data. Therefore, a conservative assumption was made to keep capability the same as baseline over the three years as a proxy longitudinal control group.

Third, the selection of the threshold of sufficient capability (TSC) was undoubtedly somewhat arbitrary. The selection of the ‘a little capability or 22222’ threshold was guided partly by capability theorists assertion that all basic capabilities are of equal importance (Nussbaum, 2000; Venkatapuram, 2011); therefore, no capability for any attribute is of chief priority to avoid. The selection of the ‘a lot of capability or 33333’ was guided by previous research with the ICECAP-O questionnaire, which showed that average ICECAP-O UK population values were relatively close to this threshold before re-scaling the valuation dataset (“33333” = 0.868

vs. UK over 65 years old ICECAP-O average values = 0.832 (Flynn et al., 2011)). The "33333" threshold may also provide a more reasonable identification of sufficient capability within a developed nation such as the UK where higher levels of capability would be expected, rather than the lower threshold of "22222". Additionally, the cutoff (k) for the number of capability attributes required to reach TSC was also a subjective choice, made for ease of interpreting the methodology (where, k = 1 meaning that a person was below sufficient capability if they were below the attribute threshold level on any one attribute). Such uncertainty in defining a suitable TSC and cutoff of TSC is open to scrutiny, and requires further exploration and justification for the methods to be taken forward.

Fourth, the ICECAP-O has been used here as the measure of sufficient capability. Whilst this may be appropriate for the arthritis population (mean age approximately 70) in this dataset, to allocate resources across the entire health service, the more recently developed ICECAP-A (Al-Janabi et al., 2012a) is likely to be a more suitable measurement of capability. The design of the ICECAP-A is for the general adult (18+) population, compared to the ICECAP-O, which is designed for older people specifically. Nevertheless, the same methodology can be applied with the ICECAP-A.

Fifth, assumptions are made to combine sufficient capability with time to generate sufficient capability outcomes over time. The assumption of sufficient capability outcomes to be constant over time was made for simplicity, and no discount rate was applied. The second assumption is that the "no capability" score on the 0-1 scale remains at zero as no capability developed in the ICECAP-O valuation (Coast et al., 2008a). Further research is required to develop a better understanding on the capability depreciation (if any) over time, as well as whether there are worse states than the “pits” state of no capability on all the ICECAP-O attributes.

There is no particular need for the sufficient capability methodology to be exclusive to the ICECAP questionnaires. It could be used with other capability instruments, such as the other capability questionnaires described in Chapter 3, and could potentially be applied to generic measures of health like the EQ-5D, with the emphasis being on minimum thresholds for health.

This chapter proposes a method to incorporate the capability approach within health economic evaluations. It draws from capability applications and combines a capability instrument with time for a number of outcomes. The key finding from this chapter is that the choice of threshold of sufficient capability is likely to be crucial to the resource allocation process if this methodology was implemented across a national health service, as shown in Table 24 where the benefits vary depending on the threshold chosen.

CHAPTER 8. METHODOLOGY FOR OBTAINING CAPABILITY OUTCOMES FROM AN ECONOMIC MODEL: A CASE STUDY

8.1 INTRODUCTION

It has already been established in this thesis that capability can be predicted from measures of physical function (see Chapter 6). Additionally, Chapter 7 has outlined a methodology to generate meaningful outcomes which can be of value in allocating scarce resources from a capability perspective. In Chapter 5, the case study for incorporating a capability instrument into an economic model was identified as the Birmingham Rheumatoid Arthritis Model (BRAM). Using the methodology from Chapter 6 and 7, it is now possible to incorporate a capability instrument into an economic model through mapping (Chapter 6) to generate outcomes from the capability approach (Chapter 7). In this chapter, the methods for generating capability outcomes from the BRAM model are explained.

The remainder of the chapter is structured as follows. In Section 8.2, an outline of the main characteristics of the case study is presented. The case study chosen has been implemented on a number of occasions in the past to aid decision-making in the UK (Barton et al., 2004a; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011). A focus on how previous economic outcomes were generated in the selected case study is of primary interest to the generation of capability outcomes in this chapter. Then the methods (Section 8.3) for generating four capability outcomes from the chosen case study are outlined, with a results section (Section 8.4) presenting the required input into the chosen economic model to obtain the outcomes for further analysis in an evaluation. The results of the evaluation are presented in Chapter 9.

8.2 BACKGROUND TO THE BIRMINGHAM RHEUMATOID ARTHRITIS MODEL (BRAM)

Following the justification of the selected case study (Chapter 5), this section details the main features of the case study. The focus, in particular, is on the type of outcomes which have been produced from this case study previously, which act as a comparator to the methodology of generating capability outcomes from the case study in the following section. Information within Section 8.2 is not original work by the author of this thesis and is presented here for illustration of the BRAM model process.

8.2.1 A review of previous BRAM versions

BRAM is an individual sampling model, a type of model that accounts for individual patient pathways at a patient-level basis (Barton et al., 2004a). The model is designed to represent clinical treatment for patients with rheumatoid arthritis, a chronic ailment that is caused by inflamed synovial tissue between joints. Rheumatoid arthritis is a progressive disease that causes functional disability, significant pain and destruction of the joints, which can in turn lead to premature mortality (Kvien, 2004).

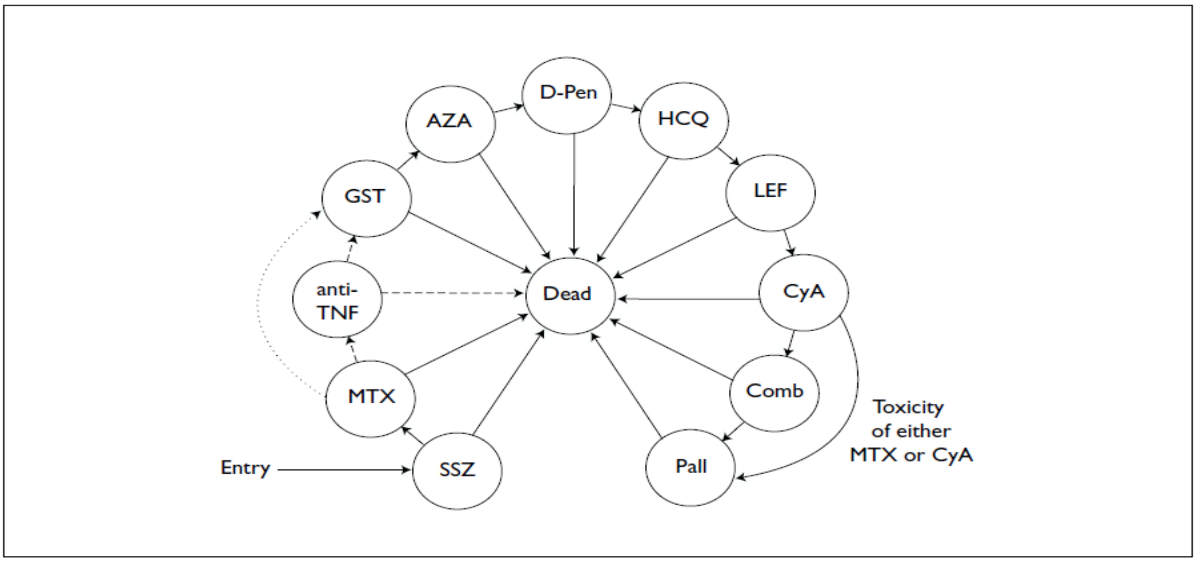
Each individual pathway in an individual sampling model produces a virtual patient history, which aims to give a true reflection of the population treatment strategy under consideration.

It is then from these patient histories that mean population values in terms of costs and benefits can be generated (Barton et al., 2004a).

Over the period between 2000 and 2013, four different BRAM models have been developed for aiding decision-making in terms of the cost-effectiveness of different drug treatment strategies for rheumatoid arthritis within the UK. The first model developed by the West Midlands Health Technology Assessment Collaboration was known as the Birmingham Preliminary Model (or BPM02 from now on, with 02 indicating year of model publication) (Jobanputra et al., 2002). The BPM02 was used to assess the introduction of new biologic treatments, clinically called anti-bodies against tumour necrosis factor (anti-TNFs) drugs which include etanercept (ETN) and infliximab (IFX) to the National Health Service (NHS). The introduction of either of these two drugs, as a third-line therapy or last active treatment, was compared with a sequence of non-biologic disease modifying anti-rheumatic drugs (DMARDs) beginning with sulphasalazine (SSZ) then methotrexate (MTX) then gold (sodium aurothiomalate - GST) which are a likely starting treatment strategy if biologic treatments were not available (Jobanputra et al., 2002). The DMARD sequence employed in BPM02 can be seen in Figure 14 (originally Figure 14 featured in the BPM02 report (Jobanputra et al., 2002)).

In BPM02 and all subsequent BRAM models, the choice of model was an individual sampling model, due to the rigidness associated with Markov modelling in terms of the independence of transition probabilities from timeframe (Markov assumption) and previous health states (homogeneous assumption), as well as the requirement of a fixed cycle length with such a model (Barton et al., 2004b). In an individual sampling model, each patient pathway produces a number of virtual patient histories, which aims to give a true reflection of the population characteristics under consideration (Barton et al., 2004b).

Figure 14 Treatment strategy options and patient pathway represented in BPM02



In Figure 14, a circle represents a different stage in BPM02, with the majority of circles representing a drug treatment, as well as dead and palliation treatment (Pall). Other abbreviations are SSZ, sulphasalazine; MTX, methotrexate; GST, gold; AZA, azathioprine; D-Pen, penicillamine; HCQ, hydroxychloroquine; LEF, leflunomide; CyA cyclosporin; Comb, combination of MTX plus CyA.

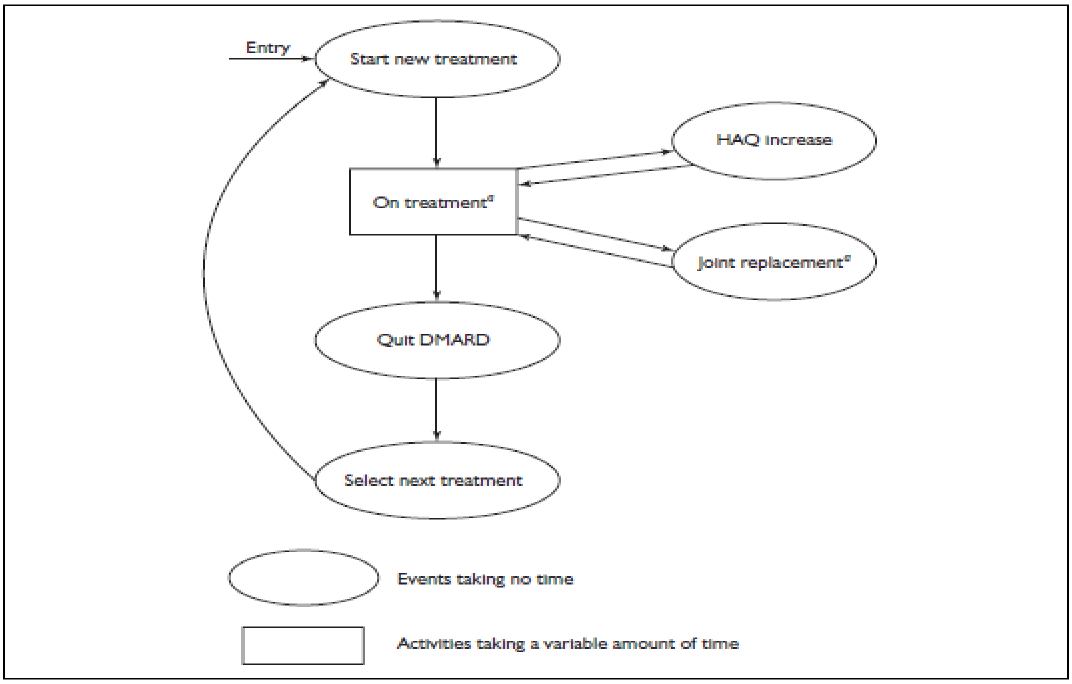
Source: Originally appeared as Figure 9, p. 54 in Jobanputra et al. (2002)

BPM02 classified toxicity or loss of effectiveness as the reasons that individuals quit a treatment within the model. The sole patient characteristic required in BPM02 was a patient's remaining lifetime, which was calculated from relevant life expectancy tables for rheumatoid arthritis patients. The BPM02 model also assumed that improved health related quality of life associated with DMARDs and anti-TNFs was fixed, with the health improvement lost when quitting that treatment. Mortality effects from DMARDs were not included in BPM02. Other limits to changes in individual status over time for the first BRAM version included the inability to account for the effect of disease progression, joint replacement and hospitalisation within BPM02. However, many of the assumptions in BPM02 were made due to the time constraints associated with developing the model, hence the use of the word 'Preliminary' in the title of the first model version (Jobanputra et al., 2002).

More time was required for the further enhancement of the model to better capture changes in patient pathways during drug treatment strategies for rheumatoid arthritis patients. The main recommendations for improving BPM02 were elaborated in a report in 2004 (Barton et al., 2004a). The main change from BPM02 to the subsequent BRAM models is the introduction of the Health Assessment Questionnaire-Disability Index (HAQ-DI), as the primary predictor of changes in health related quality of life as an individual moves through the model. HAQ-DI, a measure of functional disability used commonly in rheumatoid arthritis trials, is calculated on a discrete scale 0-3 at 0.125 increments (Bruce & Fries, 2003). HAQ-DI is also used to predict the likelihood of mortality in subsequent models from BPM02, as higher HAQ-DI scores indicate higher physical function limitations which is negatively linked to life expectancy. Average HAQ-DI scores are also allowed to vary for different treatments and the effect of joint replacement, measured again by HAQ-DI score, is also allowed for in the enhanced model versions. The process of an individual going through the next three versions of BRAM is outlined in Figure 15, which originally appeared in Barton et al. (2004a).

The changes to BPM02 outlined in Barton et al (2004a) were first implemented in a new model in the same year of recommendation (BRAM04 from now on). BRAM04 looked at the introduction of anakinra, a new drug treatment, into a DMARD sequence representing current practice, as opposed to current practice without anakinra (Clark et al., 2004). The same short-term improvement captured in BPM02 when starting a new treatment is now measured using a fixed improvement (decrease) in HAQ-DI score in BRAM04. This allowed patients' HAQ-DI scores to deteriorate whilst on treatment, but HAQ-DI could never go above the worst possible score of 3. This deterioration could now be captured in BRAM04 because of the number of increased initial input characteristics of individuals in the revised model, which not only included HAQ-DI score, but also the age and sex of the individual at the start of treatment.

Figure 15 Individual patient pathway in BRAM04, BRAM06 and BRAM11 models(Barton et al., 2004a)

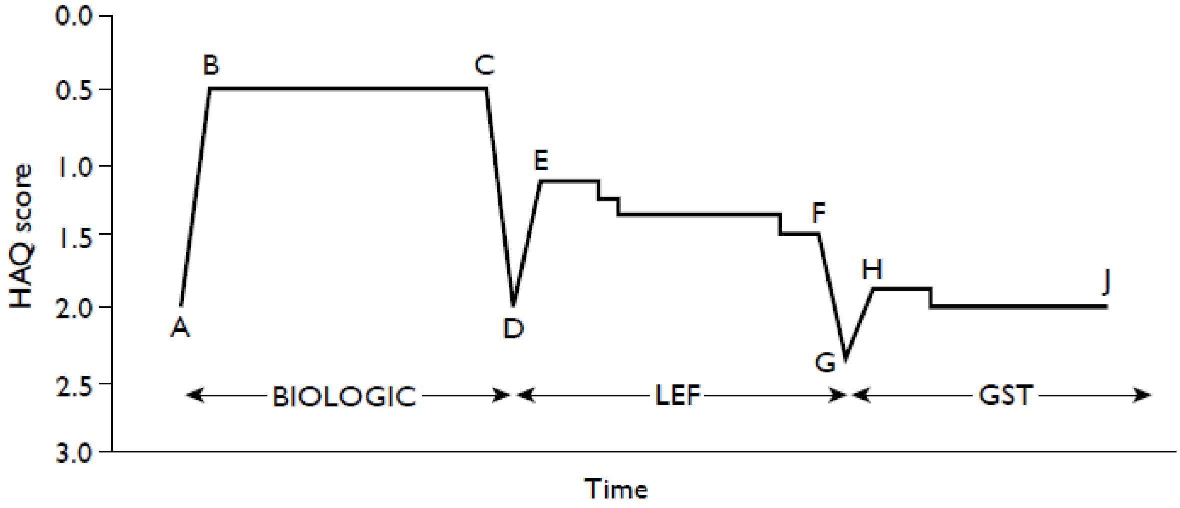


Source: Figure 15 above originally appeared in Barton et al. (2004), Fig. 3, p. 14. DMARD, Disease Modifying anti-Rheumatic Drug: HAQ, Health Assessment Questionnaire.

The next BRAM version, BRAM06, evaluated the cost-effectiveness of etanercept and infliximab once more, as well as a new anti-TNF, adalimumab (Chen et al., 2006). Further changes were instigated in BRAM06, including HAQ-DI improvement being allowed to vary for individuals when starting a new treatment. The model also allowed for greater benefits to be apportioned to those who had higher HAQ-DI scores, as they had greater potential to benefit from treatment. Two early withdrawal steps for patients were introduced into the model, for reasons of toxicity at six weeks and either toxicity or loss of effectiveness at twenty four weeks. BRAM06 was also used for a subsequent NICE report which was concerned with the sequential use of biologic treatments, which included rituximab, a monoclonal antibody targeting B lymphocytes, as an alternative strategy (Barton, 2008).

The final and most recent version of the BRAM is BRAM11 and was developed to evaluate adalimumab, etanercept, infliximab, rituximab and abatacept after the failure of one anti-TNF previously (Malottki et al., 2011). The primary change in this model was the introduction of probabilistic sensitivity analysis (PSA), which reflects the uncertainty entering the model by placing relevant distributions on the input parameters when feasible (Briggs et al., 2006). An example of the change in HAQ-DI for a patient over time from BRAM04, BRAM06 and BRAM11 is presented in Figure 16 (originally appeared in Malottki et al., 2011).

Figure 16 Representation of HAQ-DI change over time for patients in BRAM Models (BRAM04, BRAM06 and BRAM11) (Malottki et al., 2011)



AB shows initial improvement on biologic; this improvement is lost when treatment changes to leflunomide (LEF) represented by CD; DE shows the initial improvement from LEF; higher HAQ-DI scores between EF represents gradual deterioration of the patient over time; similar process for final treatment, until point J where the patient dies from other causes.

Source: Malottki et al. (2011), Fig. 94, p 147

8.2.2 Economic outcomes in previous BRAM versions

The focus of interest in this case study is the costs and benefits used to generate economic outcomes from BRAM models. Costs and QALYs for different strategies are compared using the incremental cost-effectiveness ratio (ICER) approach for all four versions of BRAM, which is the decision-rule favoured by NICE (NICE, 2013).

Costs have been handled in a comparable manner in all versions of the BRAM, accounting for patient start off costs with a new treatment, along with monitoring costs throughout time on each treatment which varied for each drug respectively. Later versions (BRAM06 and BRAM11) accounted for joint replacement and hospitalisation costs either directly in the base case analysis or through sensitivity analysis. However, all evaluations were conducted from the NHS or NHS and personal and social services perspective, as recommended by NICE at the time of the development of the previous BRAM models (NICE, 2008). Therefore, patient indirect costs and carer costs have not been considered in any of the BRAM versions. Patient and carer costs for rheumatoid arthritis patients are recognised as having a considerable impact on those affected by the disease, particularly in terms of lost hours and lost productivity at work (Kvien, 2004).

The economic outcomes from the BRAM models have been consistent throughout all versions, with the BRAM team following NICE guidance in generating QALYs for different treatment strategies. In all four versions of the BRAM, the economic outcome generated is QALYs, although methods for their generation varied for each version of BRAM. In all four models, the HAQ-DI is the primary predictor of quality of life for the respective models in the QALY calculation. This is also true for the majority of other economic models involving rheumatoid arthritis patients when applying a measure of condition-specific or generic health status instrument to predict health utility (Marra et al., 2011).

In the first model version (BPM02), limited time meant that an opinion of the likely relationship between an improvement (i.e. decrease) in HAQ-DI score and EQ-5D score informed the calculation of quality of life, which allowed for the calculation of QALYs. Any HAQ-DI improvement/deterioration was multiplied by 0.2 to generate the change in quality of life. However, such 'quick and dirty' techniques have questionable validity issues (Coast, 1992), due to a subjective opinion being used to generate the relationship between two questionnaires which is not easily comparable with QALYs generated from related studies (Coast, 1992; Fryback et al., 1997). The use of statistical associations such as mapping are now more commonly used and recommended (Brazier et al., 2010), rather than the 'reprocessing'(Coast, 1992) approach applied in BPM02. In BRAM04, BRAM 06 and BRAM11 a statistical approach to identify a relationship was employed (i.e. mapping).

Generating a regression analysis between HAQ-DI and EQ-5D is used in the later three BRAM versions (Barton et al., 2004a). A linear relationship between HAQ-DI and EQ-5D is applied in BRAM04 and BRAM06, whilst BRAM11 employed a quadratic relationship between HAQ-DI and EQ-5D, with Quality of Life predicted by both HAQ-DI and the square function of HAQ-DI. From a frequentist statistical point of view, the use of the quadratic regression would seem to be unwarranted given the earlier regression analysis which suggested that HAQ-DI alone was the best predictor of EQ-5D scores (Barton et al., 2004a). However, the incorporation of PSA into BRAM11 required a Bayesian approach that takes account of the uncertainty between the relationship of predicting EQ-5D from HAQ-DI more systematically. A summary of the algorithms used to generate health related quality of life in each of the models can be seen in Table 25.

The BRAM model has developed from a preliminary model with many constraints into a model which now accounts for different types of uncertainty and economic outcome measures. The question for this research is whether and how it can be further extended to consider capability, rather than QALY, outcomes.

Table 25 BRAM models and Quality of Life Calculation for QALYs

BRAM version	Quality of Life Calculation	Equation
BPM02	HAQ-DI improvement from treatment multiplied by 0.2	HAQ-DI reduction of 1 = 0.2 improvement in quality of life
BRAM04	Linear regression HAQ-DI score used to predict EQ-5D as QoL measure	QoL = 0.862-0.327×HAQ-DI
BRAM06	Same as BRAM04	Same as BRAM04
BRAM11	Quadratic equation with HAQ-DI and HAQ-DI² used to predict EQ-5D	QoL=0.804-0.203×HAQ-DI-0.045×HAQ²

QoL, Quality of Life; HAQ-DI, Health Assessment Questionnaire-Disability Index

8.3 METHODS

In this section, the methods for incorporating a capability instrument, the ICECAP-O, to produce capability outcomes from BRAM are outlined.

8.3.1 BRAM version

The choice of BRAM model to carry out this analysis in reality could be any of the four versions previously described in Section 8.2.2. However, an important aspect of this research is to compare what is done presently in health economics with what could be done with capability outcomes. Therefore, while other comparisons with earlier BRAM versions might provide insightful findings, the choice of BRAM model for this study is BRAM11 (Malottki et al., 2011), given that it relates most closely to current health economic evaluations and models for the rheumatoid arthritis population. Table 26 and Table 27 present the main components of the starting population dataset in BRAM11, specifically age and sex (Table 26) and HAQ-DI as the measure of physical function (Table 27), which make up the main model inputs for the BRAM11 population.

Table 26 Age and sex distribution when starting treatment in BRAM11

	Age(years)							
Sex	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75-84**	**Total**
Male	0.0	0.4	1.9	5.2	6.5	3.8	1.2	19
Female	0.1	1.5	8.2	22.1	27.7	16.3	5.1	81

Source: Malottki et al. (2011), Table 66, p. 149

Table 27 Distribution of HAQ-DI scores among population when starting treatment in BRAM11

HAQ-DI	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	**1**
%	0.0	0.1	0.2	0.5	0.7	1.2	1.7	2.2
HAQ-DI	**1.125**	**1.25**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	**2**
%	2.9	3.6	4.3	5.1	5.8	6.6	7.2	7.7
HAQ-DI	**2.125**	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3**
%	8.1	8.4	8.3	8.0	7.1	5.9	3.7	0.7

Source: Malottki et al. (2011), Table 67, p. 149; %, proportion of population recording a given HAQ-DI score

8.3.2 Dataset

The dataset employed to predict the relationship between a starting measure of physical function (HAQ-DI) and a target measure of capability (ICECAP-O) is the baseline information from the Tayside Joint Replacement dataset. This dataset was also applied in Chapter 6 to investigate the feasibility to map from a measure of physical function (WOMAC) to a measure of capability (ICECAP-O).

8.3.3 Instruments

This section explains the two instruments of interest for this case study, the Health Assessment Questionnaire-Disability Index (HAQ-DI) and the ICECAP-O questionnaire.

8.3.3.1 Starting Measure: Amended version of the Health Assessment Questionnaire – Disability Index (HAQ-DI)

The health assessment questionnaire (HAQ) is primarily used as a measurement of functional disability in patients with rheumatoid arthritis. The HAQ has been frequently used, refined and modified since its inception over thirty years ago (Fries et al., 1980). While the "full HAQ" incorporates five dimensions of outcome (known as the five Ds: "disability", "discomfort", "drug side effects", "dollar costs" and "death"), the most widely used aspect of the HAQ is the disability index, the HAQ-DI (Bruce & Fries, 2003), which is the primary interest in this study, as this was the measure applied previously in BRAM11.

In the traditional HAQ-DI format, there are 20 questions covering 8 areas of physical function (dressing and grooming, rising, eating, walking, hygiene, reach, grip, activities), with 2 to 3 questions for each functioning (see Appendix 6). The highest (worst) score for each area of functioning is used to calculate the HAQ-DI score, with each of the eight categories valued equally. Additionally, polar questions regarding the extra support patients require from individuals, aids and devices are sometimes taken into consideration. When extra assistance is required in a HAQ-DI category, a score of 2 (much difficulty) is recorded for that category, unless 3 (unable to do) has already been reported. The final HAQ-DI score for each person ranges from 0 (no problems functioning) to 3 (not able to function). A simpler measure of the HAQ-DI called the modified HAQ or MHAQ, which only asks one question per functioning (8 in total) has been developed more recently and is now used interchangeably with the original HAQ-DI (Wolfe, 2001).

HAQ-DI scores were processed through responses given in the physical function questions section collected alongside ICECAP-O in the Tayside Joint Replacement cohort. Physical function items from this dataset were taken from the initial pool development of the new Aberdeen measures of impairment, activity limitations and participation restriction (Pollard et al., 2009). The items used in this study were from the development of the activity limitation part of the measure, which came primarily from the WOMAC questionnaire (Bellamy et al., 1988) but also HAQ-DI and SF-36 (Ware, Jr. & Sherbourne, 1992). Previous research has shown a strong relationship between the HAQ-DI and the WOMAC (Bruce & Fries, 2004). Questions were restructured in the WOMAC questionnaire format to ask patients about the degree of their difficulty on a 0-4 scale (no-severe difficulty).

Out of the twenty questions from the eight areas of physical function captured on HAQ-DI, eleven questions were worded in a directly comparable manner, two questions required four physical function questions to capture the HAQ-DI questions fully, while the remaining seven questions had no comparable questions in this dataset. Two of the eight categories on HAQ-DI (eating and grip) had no comparable questions in this dataset. However, since only six out of the eight categories are required to be completed to record a HAQ-DI score (Bruce & Fries, 2003), the study proceeded with fifteen questions across six categories to calculate an amended HAQ-DI score here (see Appendix 7 for amended version of HAQ-DI). Table 28 and 29 represent the 101 individuals who completed both the amended HAQ-DI questionnaire and the ICECAP-O questionnaire to record an overall HAQ-DI score and ICECAP-O responses. Table 28 and 29 represent the prediction dataset characteristics which act as a comparison to the BRAM11 study dataset characteristics as presented in Section 8.3.1.

Table 28 Age and sex population for prediction dataset

	Age(years)								
Sex	**15-24**	**25-34**	**35-44**	**45-54**	**55-64**	**65-74**	**75-84**	**85-94**	**Total**
Male	0	0	0	1	12	25	13	1	54
Female	0	0	0	2	13	13	13	2	47

Table 29 Distribution of amended HAQ-DI scores in prediction dataset

HAQ-DI	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	**1**
%	0.00	0.00	0.99	0.00	1.98	0.99	3.96	3.96
HAQ-DI	**1.125**	**1.25**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	**2**
%	5.94	6.93	4.95	5.94	7.92	10.89	9.90	5.94
HAQ-DI	**2.125**	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3**
%	5.94	5.94	0.99	3.96	3.96	2.97	3.96	1.98

8.3.3.2 Target Measure: ICECAP-O capability questionnaire

The ICECAP-O capability questionnaire has been explained in detail in Chapter 3 of this thesis. In summary, ICECAP-O consists of five attributes of capability well-being: attachment, security, role, enjoyment and control (see Appendix 1 for full questionnaire). These five attributes responses are valued on the ICECAP-O index, where 1 is equal to full capability and zero equal to no capability (Coast et al., 2008a). An important distinction between ICECAP-O and the starting measure (HAQ-DI) is that the best score for ICECAP-O is at the top of the index (one on a 1-0 scale), whilst the best score for HAQ-DI it is at the bottom of the index (zero on a 0-3 scale).

8.3.4 Capability Outcomes

In Chapter 7, the methodology and types of capability outcomes which could be applied in conjunction with the multidimensional poverty methodology were outlined. In this section, the methods required to generate the capability outcomes from the BRAM model are outlined, with a section devoted to each capability outcome.

8.3.4.1 Capability Outcome 1 (CO1): Years of Full Capability (YFC)

Years of Full Capability (YFC) look to sum capability over time in a similar but not identical method as to how QALYs calculate health related quality of life over time. One YFC is equal to full capability for a year while a zero YFC score represents no capability for a given year.

The methods of statistical analysis are explained below. This follows a similar approach to Chapter 6, which looked at predicting a relationship between WOMAC and ICECAP-O.

The first step in the statistical analysis involved the generation of descriptive statistics for all possible dependent and independent variables. Three independent variables (overall HAQ-DI score, age, sex) were explored for predictive significance, as all three variables were present in the prediction and study (i.e. BRAM11) datasets. Overall ICECAP-O score (continuous variable) and ICECAP-O dimensions (categorical variables) were the dependent variables considered. Scatter-graphs were used initially to explore the association between ICECAP-O scores and each of the potential explanatory variables. Box-plots for each of the five attributes of capability well-being were employed to ascertain the relationship between each attribute level and overall HAQ-DI scores.

Since this study is the first mapping attempt from HAQ-DI to capability, there was no *a priori* position on which mapping model was the most appropriate. While a number of potential model specifications are available (Brazier et al., 2010), due to the summary nature of the primary explanatory variable (HAQ-DI), only two regression model specifications were considered for further analysis.

8.3.4.1.1 ICECAP-O as a continuous dependent variable

Ordinary least squares (OLS) regression was the first regression chosen due to the prevalent use of this approach in mapping studies (Brazier et al., 2010), particularly for arthritis mapping studies (Marra et al., 2011). There have been notable limitations using OLS when predicting EQ-5D scores in previous studies, particularly due to ceiling effects (Gray et al., 2006). However, there were no ceiling effects with the ICECAP-O dataset, as only one person recorded the highest capability score achievable. The three explanatory variables consisted of two continuous variables (HAQ-DI score, age) and one discrete variable (sex). Stepwise regression (as explained in Chapter 6) was used to test the significance of the three explanatory variables. Backward elimination stepwise regression was employed here (Draper & Smith, 1998).

8.3.4.1.2 ICECAP-O attributes (5) as categorical dependent variables

There were two regression model options available for use with categorical dependent variables: ordinal logistic (OL) or multinomial logistic (ML) regression (detailed explanation of OL and ML in Chapter 6). If the OL regression is not viable due to the violation of proportionality between categories, ML regression is employed. Three ML methods (expected-utility method; most-likely probability and Monte Carlo simulation, see Chapter 6 for explanations) were considered if the assumptions necessary for OL regression were violated.

8.3.4.1.3 Prediction Measures

While common measures of goodness of fit of regression models such as R^2 play an important role in showing the explanatory power of a model for the dependent variable, the main interest for mapping is in the ability to predict the dependent variable from the explanatory variables in the model. Two common measures have been prominently used in mapping studies (Brazier et al., 2010): mean absolute error (MAE) and root mean squared error (RMSE). All models are tested for these two measures of prediction error, with lower prediction error scores indicating a better model for prediction. All models were also tested for normality and heteroscedasticity. Analysis was carried out using STATA Version 10.1 and Microsoft Excel 2007.

8.3.4.1.4 Prediction Specification

The prediction specifications of YFC in terms of ICECAP-O from BRAM11 are outlined below:

(1) Overall ICECAP-O score = f(HAQ-DI score, age, sex)

The overall HAQ-DI score, age and sex are the independent variables in specification 1. The ICECAP-O, HAQ-DI and age are entered as continuous variables, while sex is inserted as a dummy variable. It is noted that (1) is regression 1 in the results section for YFC

(2) Overall ICECAP-O score = f(HAQ-DI score)

A reduced version of specification 1, with overall HAQ-DI score as the only predictor variable. It is noted that (2) is regression 2 in the results section for YFC

(3) Overall ICECAP-O score = f(HAQ-DI score^2)

An alternative to specification two, which is implemented when the residuals of a regression are skewed and heteroscedastic. The square number of overall HAQ-DI score was chosen after using the "ladder" command in STATA, which produced the lowest p-value and highest chi-square for the HAQ-DI variable. It is noted that (3) is regression 3 in the results section for YFC

(4) ICECAP-O dimensions = f(HAQ-DI score, age, sex)

The five ICECAP-O dimensions are entered as categorical variables. The overall HAQ-DI score and age are continuous independent variables. Sex is entered as a dummy variable.

(5) ICECAP-O dimensions = f(HAQ-DI score)

A reduced version of specification 4. The five ICECAP-O dimensions are inserted as five categorical dependent variables. Overall HAQ-DI score is the sole predictor variable, which is entered as a continuous variable. It is noted that (5) are regressions 4-6 in the YFC results section with 4 (Expected-Utility), 5 (Most Likely Probability) and 6 (Monte Carlo simulation).

Regressions 1-3 prediction errors are estimated through Ordinary Least Squares (OLS), the most commonly used regression method for mapping from one instrument to another in health economics (Brazier et al., 2010). Regression 4-6 can be estimated using OL or ML regression.

8.3.4.2 Capability Outcome 2 (CO2): Poverty Free Years (capability) (PFY(c))

Poverty Free Years (capability), or PFY(c), measures the amount of time an individual is poor or not poor in terms of capability well-being. Whether a person is poor or not depends on the threshold of sufficient capability (TSC), which was discussed in detail in Chapter 7 and is set to a lot of capability on all the ICECAP-O attributes (i.e. 33333) for the remainder of this thesis. To predict the probability of an individual having a poverty free year in terms of sufficient capability for a given HAQ-DI score, a logit regression is employed. The equation for calculating PFY(c) is presented below:

$$PFY(c) = Pr(TSC=1) \qquad (8.1)$$

where Pr(TSC=1) equals the probability of reaching the threshold of sufficient capability (TSC) for a given HAQ-DI score.

Other tests on the prediction dataset for PFY(c) included whether different age or sex characteristics resulted in differing probabilities of poverty in terms of capability for sub-groups.

8.3.4.3 Capability Outcome 3 (CO3): Insufficient Capability Score (ICS)

Insufficient Capability Score (ICS) is determined by the shortfalls of sufficient capability (if any) for an individual at a given time point. Years of Insufficient Capability (YIC) is the method of summing the opposite of sufficient capability score (SCS) over time so that shortfalls in sufficient capability are accounted for. The SCS, ICS and YIC have been explained in detail in Chapter 7.

The calculation of the inverse of the sufficient capability score is explored using a number of regression options. The first regression option tested is simple OLS, calculating ICS for a given HAQ-DI score. An alternative method is also tested here known as the two-part model approach, commonly used in cost-based regression models (Manning et al., 1981; Liu et al., 2010; Winkelmann, 2012) which has also been gaining popularity in mapping studies recently (Chuang & Kind, 2009; Dakin et al., 2010; Dakin et al., 2013b). There are likely to be a high proportion of individuals who reach the threshold of sufficient capability (TSC), so in a similar way that a high proportion of patients record zero costs, the two-part model approach appears appropriate for generating those who record shortfalls in sufficient capability.

The first part of the two part model assesses the probability of sufficient capability for a given HAQ-DI score, using a logit regression. This is equivalent to the probability regression employed for capability outcome PFY(c), which was described in Section 8.3.4.2. The second part of the model assesses the shortfall of sufficient capability for a given HAQ-DI score if the threshold of sufficient capability (TSC) has not been achieved. The values attached to the shortfalls in sufficient capability for ICECAP-O at the TSC of 33333 are presented in the previous chapter (see Chapter 7, Table 12). The algorithm used to calculate ICS for any of the 25 potential HAQ-DI scores, is presented below:

$$\text{ICS} = 1-[\Pr(\text{TSC} = 1) + (1 - \Pr(\text{TSC} = 1))IC\,] \tag{8.2}$$

where *IC* is the predicted ICS, conditional on falling below the threshold of sufficient capability (TSC).

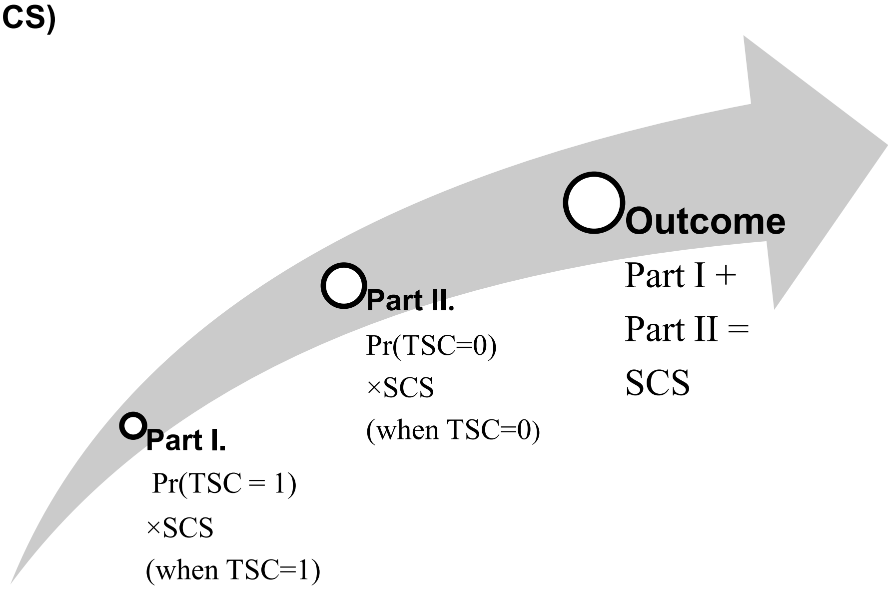
8.3.4.4 Capability Outcome 4 (CO4): Sufficient Capability Score (SCS)

The sufficient capability score (SCS) is the conceptual opposite of the ICS, so that the index of 0-1 for SCS means that a score of 1 is equal to sufficient capability. A score of one for a full year is equal to 1 year of sufficient capability (YSC). Further elaboration of both methods can be found in Chapter 7.

Like ICS, two types of regression models are tested for SCS. Firstly by calculating the relationship between SCS and HAQ-DI, and secondly through a two-part model. In the second approach, SCS is calculated by adding the probability of those who reached sufficient capability for a given HAQ-DI score, to those who are not capability sufficient multiplied by their SCS, which is less than 1. For those who have not reached the threshold of sufficient capability (TSC), their SCS is the opposite of ICS (i.e.1 minus their ICS score). The algorithm for calculating SCS is presented below, and is presented graphically in Figure 17:

$$SCS = Pr(TSC=1)\times 1 + Pr(TSC=0)\times(1-ICS) \tag{8.3}$$

Figure 17 Two part model process to calculate Sufficient Capability Score (SCS)



8.4 RESULTS

8.4.1 Descriptive Statistics

Table 30 presents the demographics of the patient population. Six patients did not fully complete either the ICECAP-O or the amended HAQ-DI, so were excluded from further analysis. The mean ICECAP-O score for the remaining 101 patients was 0.779. The mean amended HAQ-DI score was 1.773. The 101 respondents had a mean age of 69.94 years.

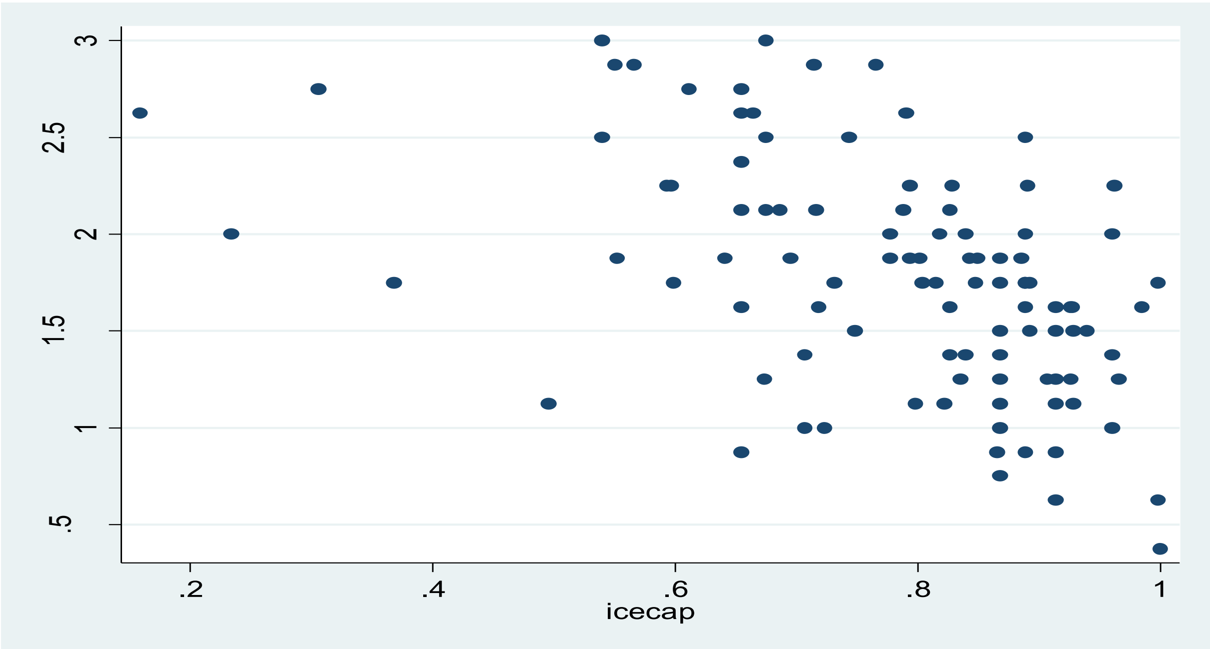
Table 30 Descriptive Statistics for Prediction Dataset

	N	Mean	Standard Deviation	Min	Max
Sample size	107				
Missing data	6(excluded below)				
Males	54				
Employed	13				
Living Alone	27				
Age (mean)	101	69.941	8.812	48.819	89.164
HAQ-DI	101	1.774	0.598	0.375	3
ICECAP-O	101	0.779	0.162	0.159	1
↓ Responses	**Capability Level**	**1 No**	**2 A little**	**3 A lot**	**4 Full**
Attachment		2	9	36	54
Security		10	28	48	15
Role		5	37	46	13
Enjoyment		10	34	46	11
Control		4	27	44	26

N, population numbers; Min, Minimum; Max, Maximum

Figure 18 shows the relationship between ICECAP-O and HAQ-DI and indicates a clear skewness towards higher ICECAP-O scores. This was an early indication that variable transformation would be required. It is also interesting to note that most of the patients reported a level of capability above 0.5 on ICECAP-O, irrespective of HAQ-DI score. Notwithstanding this, those who reported better physical function on HAQ-DI (i.e. less than 1), they also reported high capability (greater than 0.6 in all instances).

Figure 18 Relationship between ICECAP-O and HAQ-DI



8.4.2 Capability Outcome (1): Years of Full Capability

In total, six models predicted ICECAP-O scores or dimensions. Three models (Models 1-3) predicted overall ICECAP-O scores as a continuous variable through OLS regression. Regression 1 considered HAQ-DI score, age and sex as explanatory variables of overall ICECAP-O scores. Stepwise regression eliminated both age and sex as non-significant predictors of capability in this sample. The reduced equation of HAQ-DI as the sole predictor of ICECAP-O is illustrated in Regression 2. Model specification tests showed that Regression 2 has residuals which produced non-normally distributed and heteroscedastic data. Data transformation was required to address these issues, which were corrected for in Regression 3, where the square function of overall HAQ-DI score predicted ICECAP-O.

Regressions 4-6 explored ICECAP-O attributes as dependent variables. The assumption of proportional odds between categorical levels did not hold for ICECAP-O dimensions using the Wald test, so OL regression was not used here. Instead, the three methods of ML regression (Expected Utility Method – Regression model 4, Most-Likely Probability – Regression model 5, Monte Carlo Simulation – Regression model 6) were analysed for their ability to predict ICECAP-O categories with HAQ-DI as the sole explanatory variable.

Table 31 presents the predictive statistics for the six regression models. All regressions recorded predicted ICECAP-O scores equal to or above the observed ICECAP-O scores, which is likely to be due to the negatively skewed ICECAP-O data. Regression 6 (Monte Carlo simulation) predicted the highest observed capability score (i.e. 44444 or ICECAP-O score = 1) at the individual level, while Regression 4 is most distant from this value with its high score at 0.889. Regression 6 also produced the lowest ICECAP-O score achievable at the individual level (i.e. 11111 or 0.000), while the remaining models' minimums were all greater than 0.55. Regressions 1, 2 and 3 have predicted ICECAP-O scores closest to the observed mean.

Table 31 Prediction of ICECAP-O score and dimensions to generate YFC

Regression	n	Observed ICECAP-O	Predicted ICECAP-O	Min	Max	Abs. Diff.	MAE (SE)	RMSE	R^2
1	101	0.779	0.779	0.594	0.963	0.000	0.102(0.093)	0.138	0.262
2	101	0.779	0.779	0.614	0.968	0.000	0.101(0.096)	0.139	0.248
3	101	0.779	0.779	0.572	0.906	0.000	0.099(0.097)	0.138	0.262
4	101	0.779	0.786	0.586	0.889	0.007	0.097(0.099)	0.138	0.266
5	101	0.779	0.829	0.574	0.914	0.030	0.101(0.115)	0.153	0.218
6	101*1000	0.779	0.788	0	1	0.009	0.097(0.099)	0.138	0.262

S.D., standard deviation; RMSE, root mean squared error; MAE, mean absolute error; Abs Diff, absolute difference between predicted and observed ICECAP-O

In this study, the primary interest is in the predictive ability of ICECAP-O from HAQ-DI. All three of the OLS regressions (Regressions 1-3) produce similar predicted average ICECAP-O scores to the observed sample, with the ML regressions (Regressions 4-6) reporting averages higher than the mean (0.007-0.030). Regression 3 (HAQ-DI² predicting overall ICECAP-O score) produced the lowest MAE statistics for the OLS regressions (Regression 1-3) with MAE = 0.099. Two of the multinomial models (Regression 4, Expected Utility; Regression 6, Monte Carlo simulation) produced slightly lower MAE of 0.097 each, but this lower error is negated in the RMSE statistics with Regressions 3, 4 and 6 recording 0.138. Given that the main concern is generating overall ICECAP-O scores for YFC, the simplest method with lowest computational difficulty is Model 3, which is the prediction of ICECAP-O from HAQ-DI². Although Model 4 and Model 6 produce lower MAE, the difference from average observed to predicted ICECAP-O score is greater than the averages from the OLS regressions. Table 32 gives the regression required to predict the ICECAP-O from the amended HAQ-DI used in this study.

Table 32 Predicting the ICECAP-O from the amended HAQ-DI: Mapping Model

Variable	Coefficient	Standard error	t	P>|t|	95% Confidence Interval
HAQ-DI²	-0.0376	0.006	-5.94	0.000	[-0.0502,-0.0250]
Constant	0.9108	0.026	34.83	0.000	[0.8589,0.9627]

The prediction of Years of Full Capability (YFC) from Regression 3 is outlined in Table 33~~0~~.

Table 33 Predicted Years of Full Capability for given HAQ-DI score

HAQ-DI	**0**	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	
YFC	0.9108	0.9102	0.9084	0.9055	0.9014	0.8961	0.8896	0.8820	
HAQ-DI	**1**	**1.125**	**1.25**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	
YFC	0.8732	0.8632	0.8520	0.8397	0.8262	0.8115	0.7957	0.7786	
HAQ-DI	**2**	**2.125**	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3**
YFC	0.7604	0.7410	0.7205	0.6987	0.6758	0.6518	0.6265	0.6001	0.5725

YFC, Years of Full Capability

8.4.3 Capability Outcome (2): Poverty Free Years (capability) (PFY(c))

Whilst the mapping work in Chapter 6 and the mapping for YFC showed that there was little predictive relationship between sex and overall ICECAP-O score, a different finding occurred when predicting PFY(c). Table 34 below shows that males appear more likely to reach sufficient capability than females in this dataset, when the threshold of sufficient capability (TSC) is set to a lot of capability on all attributes (i.e. 33333). The difference between the sexes is statistically different using the chi-square test (p = 0.01927).

Table 34 Capability poverty statistics in Tayside joint replacement cohort

	Female	Male	Total
Capability Poor	35	28	63
Sufficient Capability	12	26	38
Total	47	54	101

The finding presented in Table 34 led to two potential methods for predicting PFY(c). One option, applying the two part model, is to express the probability of having sufficient capability for the whole population, while the second possibility is to separate the probability of sufficient capability for males and females, which may take account of the differences observed in Table 34 more accurately. The probabilities for each scenario are presented in Table 35. All three predictions are carried out by applying logistic regression to predict a binary outcome (poor or not poor in terms of sufficient capability).

Table 35 records an unintuitive result for a HAQ-DI score of 2.25 when broken down by sex. For every other HAQ-DI score, either male or female categories are above or below the probability of TSC for all individuals, irrespective of sex. Also, while females are less likely to have reached TSC at the best HAQ-DI score (0) than males, they are more likely to have reached TSC at the worst HAQ-DI score. Given these ambiguities with the sex statistics, the decision was made to only employ overall probabilities of reaching the threshold of sufficient capability (TSC). The prediction of PFY(c) from HAQ-DI scores can thus be seen in Column 2 of Table 35.

Table 35 Probability that individual has reached threshold of sufficient capability (TSC) given their HAQ-DI score for all, males and females

HAQ-DI	Pr(TSC)-ALL	Pr(TSC)-Male	Pr(TSC)-Female
0	0.9470	0.9828	0.8073
0.125	0.9330	0.9765	0.7791
0.25	0.9155	0.9681	0.7480
0.375	0.8941	0.9567	0.7141
0.5	0.8680	0.9415	0.6776
0.625	0.8367	0.9213	0.6389
0.75	0.7996	0.8951	0.5982
0.875	0.7565	0.8639	0.5562
1	0.7076	0.8190	0.5133
1.125	0.6533	0.7673	0.4702
1.25	0.5948	0.7060	0.4276
1.375	0.5334	0.6362	0.3860
1.5	0.4710	0.5602	0.3461
1.625	0.4095	0.4813	0.3081
1.75	0.3507	0.4033	0.2726
1.875	0.2961	0.3298	0.2398
2	0.2467	0.2639	0.2098
2.125	0.2033	0.2070	0.1826
2.25	0.1658	0.1598	0.1583
2.375	0.1340	0.1217	0.1366
2.5	0.1076	0.0916	0.1176
2.625	0.0858	0.0685	0.1008
2.75	0.0681	0.0508	0.0862
2.875	0.0539	0.0375	0.0736
3	0.0425	0.0276	0.0626

Pr(TSC), probability for reaching threshold of sufficient capability for whole sample; ALL, all population; Male, for men only; Female, for women only

8.4.3 Capability Outcome (3): Years of Insufficient Capability (YIC)

To calculate Years of Insufficient Capability (YIC), the inverse of the sufficient capability score is required. The insufficient capability score (ICS) is composed of the individuals who are classified as poor in the Tayside cohort from Table 34 (i.e. individuals who have not reached sufficient capability). As can be seen from Table 34, 63 of the 101 individuals do not reach the threshold of sufficient capability (TSC). The sixty three people who exhibit shortfalls in sufficient capability are the focus of analysis when generating the insufficient capability score (ICS).

An OLS regression with ICS predicted from HAQ-DI, resulted in ICS scores below zero for some HAQ-DI scores. Therefore, two-part models are a more appropriate way of calculating ICS from a measure of physical function, such as HAQ-DI. Part 1 of the model is the same calculation used to generate PFY(c) in section 8.4.2. However, even when the zeros were excluded from the second part of the two part model, HAQ-DI scores still predicted ICS scores lower than zero. Given the limitations of HAQ-DI on its own, the prediction of ICS based on HAQ-DI^2, similar to the YFC calculation in the previous section, is tested and produced algorithm 8.4 below, which is used to predict ICS in the BRAM model. The formula for calculating ICS from this dataset is presented in equation 8.4.

$$ICS = 0.0964+0.0289(HAQ\text{-}DI^2) \quad (8.4)$$

Algorithm 8.4 resulted in acceptable distribution of ICS scores based on HAQ-DI (0.0964-0.3564) for those who exhibited shortfalls in sufficient capability. This algorithm is conditional on the probability that an individual did not reach TSC for each given HAQ-DI score (i.e. the opposite probability exhibited in column 2, Table 35). The HAQ-DI scores predictions of ICS for this population are outlined in Table 36:

Table 36 Predicted Insufficient Capability Score for given HAQ-DI score

HAQ-DI	**0**	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	
ICS	0.0051	0.0065	0.0083	0.0106	0.0137	0.0176	0.0226	0.0289	
HAQ-DI	**1**	**1.125**	**1.25**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	
ICS	0.0366	0.0461	0.0574	0.0705	0.0854	0.1020	0.1201	0.1394	
HAQ-DI	**2**	**2.125**	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3**
ICS	0.1597	0.1808	0.2025	0.2246	0.2472	0.2701	0.2935	0.3172	0.3413

ICS, Insufficient Capability Score

8.4.4 Capability Outcome (4): Years of Sufficient Capability (YSC)

It has already been explained in the last section, that the two-part model is appropriate for capability outcomes with a high proportion of individuals reaching the optimum point, in this scenario, the threshold of sufficient capability (TSC). The sufficient capability score (SCS) is used to calculate Years of Sufficient Capability (YSC) for the opposite calculation of ICS from Table 36. SCS for a given HAQ-DI can be calculated from Table 37 below. While Table 37 is the direct opposite of Table 36, it is presented here for clarity to show how the SCS is calculated in comparison.

Table 37 Predicted Sufficient Capability Score for given HAQ-DI score

HAQ-DI	**0**	**0.125**	**0.25**	**0.375**	**0.5**	**0.625**	**0.75**	**0.875**	
SCS	0.9949	0.9935	0.9917	0.9894	0.9863	0.9824	0.9774	0.9711	
HAQ-DI	**1**	**1.125**	**1.25**	**1.375**	**1.5**	**1.625**	**1.75**	**1.875**	
SCS	0.9634	0.9539	0.9426	0.9295	0.9146	0.8980	0.8799	0.8606	
HAQ-DI	**2**	**2.125**	**2.25**	**2.375**	**2.5**	**2.625**	**2.75**	**2.875**	**3**
SCS	0.8403	0.8192	0.7975	0.7754	0.7528	0.7299	0.7065	0.6828	0.6587

SCS, Sufficient Capability Score

8.5 DISCUSSION

The primary aim of this chapter was to detail the methods for generating capability outcomes from an existing economic model which has been used to aid health decision-making, the BRAM. The methods for generating four capability outcomes from the relationship between a questionnaire of physical function (the amended HAQ-DI) and a questionnaire of capability well-being (ICECAP-O) has been outlined. The production of each of the four capability outcomes is discussed individually below.

8.5.1 Capability Outcome (1): Years of Full Capability (YFC)

In this study, the ability of econometric regression models to predict capability well-being in terms of ICECAP-O to generate Years of Full Capability (YFC) from a physical function questionnaire (HAQ-DI) for arthritis patients has been investigated. This is the second attempt to use a mapping approach with the capability wellbeing measure. Results from Chapter 6 suggest that the majority of the ICECAP-O attributes can be captured by a measure of physical function. Six models are tested for predictive ability. The regression chosen was the transformed version of HAQ-DI^2 as the predictor of ICECAP-O scores (i.e. Regression 3 in Table 31). Regression 3 produced minimal difference from observed to predicted ICECAP-O average scores, as well as the equally lowest RMSE.

This is the first attempt to generate capability outcomes in terms of YFC, which can be subsequently applied to economic models to test the change in overall capability well-being over time. The results of this Chapter suggest that two methods of regression produce similar error prediction statistics. In a similar manner to health economics mapping work (Brazier et al., 2010), the method with the lowest complexity in application is chosen.

There are a number of limitations associated with this research. The HAQ-DI scores were inferred from physical function questions in the form of the WOMAC (0-4), so a rescaling exercise was required to fit questions to the HAQ-DI 0-3 scale (see Appendix 7). The sample size of 101 individuals is small compared to previous mapping studies using the response mapping approach (Gray et al., 2006; Rivero-Arias et al., 2010), so 1000 Monte Carlo simulations were required in order to reach equivalent sample sizes. However, there are mapping studies with smaller sample sizes (Brazier et al., 2010) and for an exploratory study such as this in a novel area of research, the data are sufficient to provide interesting results.

There was also a limitation with the type of sample, which was comprised of patients awaiting joint replacement and this meant that individuals were not evenly distributed across HAQ-DI scores, unlike in previous mapping studies using the HAQ-DI (Barton et al., 2004a). However, the distribution of HAQ-DI in this prediction dataset is comparable with the patient input into BRAM11 (see Table 27 and 29). Finally, validation of mapping models usually takes place in an external dataset as well as the prediction dataset, but lack of data restricted analysis to internal validation here.

This research looked at all potential mapping algorithms which have been used in previous studies. The use of OLS and ML regression was explored. However, a new Bayesian approach has produced lower predictive errors than the response mapping approach and OLS predictions (Le & Doctor, 2011). Due to the way in which HAQ-DI was calculated here, it could not be fully broken down for dimension-specific analysis. Therefore this new approach to mapping could not be tested here. Future studies with such detailed variables should look at applying this Bayesian approach in statistical analysis.

8.5.2 Capability Outcome (2): Poverty Free Years (capability) (PFY(c))

To predict PFY(c) in this study from HAQ-DI, logistic regressions are applied to test the probability of poverty, given a threshold of sufficient capability (TSC). A calculation of sex and HAQ-DI, and HAQ-DI solely are used to predict the probability of reaching TSC. There was a significant difference in the ability of different sexes to reach the TSC, with men more likely to reach TSC than women. However, the HAQ-DI variable was the only variable used to predict PFY(c) due to the unreliability of splitting an already small sample size in two.

The generation of PFY(c) from a measure of physical function allows for a new type of capability outcome to be generated from economic decision models that have never been used before. The major strength of this research is that it allows the further exploration of capability outcomes such as PFY(c) in decision models which can then be used as a simplistic capability outcome to aid decision-makers.

There are a number of limitations associated with the methodology employed to generate PFY(c). Firstly the relationship between PFY(c) and HAQ-DI is generated by an amended version of HAQ-DI. Therefore, the transferability of this mapping algorithm to other studies is questionable. However, given the time constraints and the data available for this study, this was the best option available for this research.

This study shows how the capability outcome PFY(c) can be generated from a measure of physical function using a two-part model. For policy makers interested in exploring broader benefits of interventions, PFY(c) offers the most straightforward methodology drawn from multidimensional poverty methodology, explored in Chapter 7, to generate capability outcomes. However, it should be noted that such a simplistic method can also be a misleading tool in evaluating improvements in social welfare. The problems around the PFY(c) methodology have already been discussed in detail in Chapter 7.

While this study found that differences exist between the sexes for reaching the threshold of sufficient capability (TSC), the dataset applied here was not large enough to include such a factor in the generation of PFY(c). Previous research on representative samples of the elderly population in the UK have shown little difference in capability well-being overall for sex (Coast et al., 2008a; Flynn et al., 2011). However, when attention switches towards a TSC, there are indications that sex may play an important role in defining whether an individual has sufficient capability, when the TSC is set to a lot of capability on all the ICECAP-O attributes. Comparing TSC achieved with the valuation dataset used for ICECAP-O (Coast et al., 2008a), it does not appear to be limited to a specific condition, as can be seen in Table 38. Future research should focus on investigating the role of sex in achieving the TSC further.

Table 38 Sufficient Capability among ICECAP-O valuation sample of UK elderly population (Coast et al., 2008a)

	Females	Males	Total
Capability Poor	111	76	187
Sufficient Capability	66	60	126
Total	177	136	313

8.5.3 Capability Outcome (3): Years of Insufficient Capability (YIC)

To predict ICS from HAQ-DI scores to generate Years of Insufficient Capability (YIC), two econometric regression models were considered – OLS and a two-part model. When using HAQ-DI alone to predict ICS led to answers outside the 0-1 index of the measure. Conversely, introducing a two part model by itself was not adequate to produce ICS scores within the 0-1 index. Similarly to the generation of YFC as discussed in Section 8.5.1, HAQ-DI required transformation so that only values of ICS between 0 and 1 are recorded.

This is the first attempt to incorporate the most complex multidimensional poverty methodology into an outcome which can be used by decision-makers for resource allocation. This research will permit the exploration of this capability outcome in particular, to help evaluate between interventions, with the aim of minimising shortfalls of sufficient capability.

As with all of the capability outcomes examined here, there are a number of limitations associated with the ICS calculation. Overall HAQ-DI scores are from an amended version, so that outcomes can be generated from BRAM11. It is not recommended to use the mapping algorithm applied here if actual HAQ-DI and ICECAP-O data can be generated in the same dataset. However, this was not feasible with the data available at the time of this research.

The ICS capability aggregation method allows for the assessment of population shortfalls in sufficient capability over time with Years of Insufficient Capability (YIC). Chapter 6 showed that such outcomes which account for broader definitions of well-being can be generated from measures of physical function. However, it is still recommended that primary collection of ICECAP and other capability questionnaires are likely to give a more accurate indication of capability well-being within different population groups.

An unexpected finding from this analysis was the requirement for retransformation of HAQ-DI in the two-part model to generate ICS scores for those individuals whose ICS was greater than zero. Future research should focus on whether this is due to the severity (or lack thereof) of the shortfalls in sufficient capability for this arthritis dataset. If more severe losses of sufficient capability are encountered, the need for re-transformation may not be necessary

8.5.4 Capability Outcome (4): Years of Sufficient Capability (YSC)

Finally, in this chapter it is shown how to generate the sufficient capability score (SCS) through mapping. SCS is the opposite of ICS, such that the aim is to maximise levels of sufficient capability. Reversing the ICS into SCS allows for a closer comparison with traditional economic outcomes from health economic evaluations such as the QALY, as well as implementing the most complex methods of multidimensional poverty measures into a capability outcome that can be used to measure changes in sufficient capability over time (i.e., Years of Sufficient Capability (YSC)).

8.6 CONCLUDING THOUGHTS

This chapter has documented methods of generating capability outcomes from a measure of physical function. This will allow the generation of capability outcomes which can be used in an evaluation framework. It also enables the exploration of the feasibility of considering each of these outcomes in an economic evaluation (or decision-making context). This is explored in greater depth in the next chapter.

CHAPTER 9. GENERATING CAPABILITY OUTCOMES FROM THE BIRMINGHAM RHEUMATOID ARTHRITIS MODEL: RESULTS

9.1 INTRODUCTION

In Chapter 8, the methodology for generating capability outcomes from the Birmingham Rheumatoid Arthritis Model 2011 version (BRAM11) (Malottki et al., 2011) is outlined. This chapter presents the results of the case study, investigating the feasibility of generating capability outcomes from an existing evaluation framework.

As explained in the previous chapter, the BRAM model was chosen after a case study selection process. While four versions of the BRAM have been produced over the last decade (Jobanputra et al., 2002; Clark et al., 2004; Chen et al., 2006; Malottki et al., 2011), the most recent BRAM version (i.e. BRAM11) has been chosen to compare capability outcomes to previous cost-effectiveness results. The results from BRAM11 used for aiding cost-effectiveness for rheumatoid arthritis patients by NICE are summarised in Section 9.2.

To distinguish the work undertaken previously in BRAM11 to that produced using capability outcomes here, BRAM12 will refer to the modifications made to BRAM11 in order to generate new capability outcomes. The 12 in BRAM12 stands for 2012, when the majority of this part of research was carried out. While the methodology for generating capability outcomes is accounted for in the previous chapter, particular changes from BRAM11 to generate capability outcomes in BRAM12 are detailed in Section 9.3. Section 9.4 presents the results of the four capability outcomes from BRAM12. The four capability outcomes are: Years of Full Capability (YFC), Poverty Free Years (capability) – PFY(c), Years of Insufficient Capability (YIC), and Years of Sufficient Capability (YSC). Each of the capability outcome results are discussed in turn in section 9.5 before comparing the results in a case study discussion in section 9.6.

9.2 MODEL SPECIFICATION

The work in BRAM11 was carried out independently by researchers prior to this. It is presented as work that was previously completed but its inclusion in the thesis is required to give a fuller understanding of the case study specification presented in section 9.3.

9.2.1 BRAM11 – Modelling Benefits

The most recent Birmingham Rheumatoid Arthritis Model (BRAM11) evaluated the cost-effectiveness of adalimumab (ADA), etanercept (ETN), infliximab (IFX), rituximab (RTX) and abatacept (ABT) following the failure of a first tumour necrosis factor (TNF) inhibitor for rheumatoid arthritis patients. Differences in cost-effectiveness in the individual sampling model are compared by measuring the differences in costs and QALYs for the interventions (Malottki et al., 2011).

Quality of life (QoL) scores in BRAM11 are calculated by obtaining an Ordinary Least Squares (OLS) quadratic equation between the Health Assessment Questionnaire – Disability Index (HAQ-DI) and the health utility instrument EQ-5D. The relationship between the two questionnaires was predicted from a dataset where both questionnaires were collected simultaneously (Hurst et al., 1997). The algorithm applied to predict Quality of Life (QoL) in BRAM11 is presented in algorithm 9.1 below:

$$QoL = 0.804 - 0.203 \text{ x HAQ-DI} - 0.045 \text{ x HAQ-DI}^2 \quad (9.1)$$

From the above algorithm, QoL represents the dependent variable, which in this scenario is the EQ-5D global index score, where individual preferences are scaled to a 0-1 death to full health index (Dolan, 1997). The parameters represent the ability of HAQ-DI and HAQ-DI^2 to predict EQ-5D. The constants are mean numbers around this relationship, which are allowed to vary around the 95% confidence interval for each parameter respectively in BRAM11. BRAM11 was the first BRAM version to use a quadratic equation. This was appropriate as Probabilistic Sensitivity Analysis (PSA) was carried out in BRAM11 for the first time, and so the outcomes required a Bayesian approach to handling uncertainty. Previous BRAM models have relied on a linear relationship between EQ-5D and HAQ-DI only (Barton, 2011).

The BRAM models are known as individual sampling models. For each individual (virtual) drug treatment pathway, there is an underlying pattern that the patient is assumed to follow based on clinical guidance. Each individual starts and spends time on a treatment, before quitting treatment for reasons of loss of effectiveness or toxicity. When a treatment is stopped, the following treatment begins and this cycle continues until no remaining treatments are available. The final treatment is palliation. HAQ-DI is accounted for at different time points in the model. A fixed improvement on starting a treatment is assumed, which is lost when that treatment ceases. HAQ-DI declines over time to reflect the likelihood of frailty with age. This is modelled as an increase in HAQ-DI of 0.125 every 2.7 years for all treatments except for palliation, where HAQ-DI increases occur every 2 years (Barton, 2011).

The starting population for BRAM11 are those who have already failed their first TNF inhibitor. Six treatment sequences are compared in BRAM11. Five interventions - ADA, ETN, IFX, RTX and ABT – as alternative treatment commencing a sequence following the failure of the first TNF inhibitor, and a sixth sequence starting with a conventional disease modifying anti-rheumatic drug (DMARD), leflunomide (LEF). In the first five sequences, LEF is the second therapy after the first drug has failed for each respective sequence.

For each new treatment, a random number is drawn for each individual, which measures the improvement in terms of HAQ-DI, using an appropriate distribution. Another random number is drawn simultaneously to measure the length of time on treatment for each drug (Malottki et al., 2011). Qol is combined with time on treatment to generate Quality Adjusted Life Years (QALYs). QALYs are discounted at a rate of 3.5% per annum as recommended by NICE (NICE, 2008).

9.2.2 BRAM11 – Modelling Costs

Costs for each treatment include the cost of the drugs themselves, as well as additional costs relating to the administration and other costs involved with the consumption of each treatment. Costs were calculated for the year 2008. Costs are discounted at 3.5% per annum as recommended by NICE (NICE, 2008).

9.2.3 BRAM11 Results

BRAM11 is run using a fixed random number seed for at least 10,000 virtual patients. The Incremental Cost Effectiveness Ratio (ICER) is the method used to compare the costs and effects of the different sequences. The threshold of willingness to pay for an additional QALY is £20,000-£30,000, where interventions are expected to be below for approval within the UK, as recommended by NICE (NICE, 2013).

Quasi-confidence intervals are employed for ICERs, representing sampling uncertainty for each model run and not parameter uncertainty. Fixed stopping rules for Quasi-confidence intervals are used to ascertain acceptable levels of uncertainty around the ICERs for individuals sampled for each sequence (as practiced in BRAM04 and BRAM06). The first rule entailed that when one treatment dominated another, quasi-confidence intervals for cost and QALYs were required to avoid zero. Decision stopping rule two was employed in scenarios where no strict dominance occurred and a quasi-confidence interval [lower (L), upper (U)] ratio for the ICER had to fall below the following ratios before sampling was deemed adequate to account for sampling variation:

U < 5,000 or L > 200,000: U/L < 2.5

U < 10,000 or L > 100,000: U/L < 2.0

U < 20,000 or L > 50,000: U/L < 1.5

U < 30,000 or L > 30,000: U/L < 1.2

L < 30,000 or U > 30,000: U/L < 1.1

For example, a cost per QALY gain of £25,000 would require confidence intervals (£22,750,£27,250) or narrower to meet the 1.2 lower to upper ratio for ICERs less than £30,000 (i.e. £27250/£22,750 = 1.2).

Table 39 presents the baseline cost and QALYs for the six treatment sequences from BRAM11 if a deterministic analysis, rather than a PSA had been run. RTX has the lowest ICER compared with standard treatment (LEF) and dominates three of the other four sequences (ETN, IFX and ADA). The cost per additional QALY of ABT over RTX is over £100,000, considerably higher than the current guidance.

Table 39 BRAM11 deterministic results

Options	abbreviation	full name	Mean Costs (£)	Mean QALYs	ICER vs. LEF (£)	Dominated	ICER ABT vs. RTX
Option 1	ETN	etanercept	75,100	2.80	38,900	Yes- RTX	
Option 2	IFX	infliximab	73,000	2.80	36,100	Yes- RTX	
Option 3	ADA	adalimumab	74,800	2.89	34,300	Yes- RTX	
Option 4	RTX	rituximab	69,400	3.10	21,100	No	
Option 5	ABT	abatacept	93,000	3.28	38,400	No	>100,000
Option 6	LEF	leflunomide	49,000	2.13			

ETN, etanercept; IFX, infliximab; ADA, adalimumab; RTX, rituximab; ABT, abatacept; LEF, leflunomide; QALYs, Quality Adjusted Life Years; ICER, Incremental Cost-Effectiveness Ratio

9.3 BRAM12

In BRAM12 the majority of the features from BRAM11 are preserved. Costs and discount rates remained unchanged. The only change for BRAM12 is the measurement of benefit for each model run. The methods for generating the benefit in BRAM12 (i.e. capability outcomes) were explained in detail in the previous chapter.

The method of combining costs and effects is also the same in BRAM12 as BRAM11 as they are compared between different drug treatment sequences using ICERs. Since this is the first attempt to combine capability with costs over time, there is no accepted willingness to pay per unit of capability improvement *a priori* as there is for QALYs. Therefore the model run stopping rules for number of patients to run through BRAM12 need to be very precise in terms of the confidence interval ratio between upper and lower levels of ICERs. All capability outcomes are required to reach the upper to lower ratio of 1.2 before enough patients are deemed to have run through the model for additional confidence, given that there is no established decision rule for capability outcomes (e.g. WTP per QALY gain threshold of £20,000-£30,000 in UK). Costs and capability outcomes are discounted at a rate of 3.5% per annum.

The results for four capability outcomes are presented in the next section. They are Years of Full Capability (YFC); Poverty Free Years (capability) – PFY(c); Years of Insufficient Capability (YIC); Years of Sufficient Capability (YSC). Methods for generating each outcome were detailed in the previous chapter. This chapter is not reported as an economic evaluation, as the motivation of this research was the feasibility of generating capability outcomes from a model. An economic evaluation is therefore necessary and how this evaluation meets standard economic evaluation guidance (Husereau et al., 2013) is reported in Appendix 8.

The final sub-section of the results in this chapter present the differences in outcome between BRAM11 and BRAM12, which looks at the range between most effective and least effective intervention, as well as the ordering of sequences in terms of most to least effective. From the BRAM11 deterministic results in Table 39, the difference is 1.15 QALYs gained by ABT compared to LEF. Additionally the lowest cost per additional unit of outcome gained compared with standard treatment is also analysed. In BRAM11 the cost per QALY gained of RTX over LEF was £21,100.

9.4 RESULTS

9.4.1 Capability Outcome (1): Years of Full Capability (YFC)

YFC is the measurement of capability over time based on the original ICECAP-O valuation index of 0-1 no capability to full capability (Coast et al., 2008a). Table 40 below presents the findings from the first run of BRAM12. The stopping rule criteria of the 1.2 ratio from upper to lower confidence intervals resulted in 100,000 patients through this model run and 250,000 patients for the ICER for rituximab (RTX) compared with abatacept (ABT).

The treatment with highest YFC gained is abatacept (ABT) with 9.91 YFC compared with leflunomide which produced 9.35 YFC. The difference between lowest and highest effective gains is 0.56 YFC. RTX compared with LEF produced the lowest cost per additional YFC of approximately £40,200.

Table 40 Years of Full Capability (YFC) & ICER (100,000 patients)

Option	Cost (£)	QSE	YFC	QSE
1. ETN	75293	126	9.70	0.0139
2. IFX	73266	118	9.70	0.0139
3. ADA	74825	125	9.73	0.0139
4. RTX	69472	116	9.86	0.0143
5. ABT	92932	161	9.91	0.0141
6. LEF	49096	78	9.35	0.0134

Option	ICERvsLEF (£)	Quasi-CI [L-U]		ICERvsRTX (£)	Quasi-CI [L-U]	
1. ETN	76700	70000	82400			
2. IFX	69500	64300	75600			
3. ADA	68100	63300	73600			
4. RTX	40200	38000	42700			
5. ABT	78500	74600	82700	471000*	343000	752000

ETN, etanercept; IFX, infliximab; ADA, adalimumab; RTX, rituximab; ABT, abatacept; LEF, leflunomide; YFC, Years of Full Capability; ICER, Incremental Cost-Effectiveness Ratio; QSE. Quasi Standard Error.
*250,000 patient simulations were required to get the confidence level ratio below 2.5

9.4.2 Capability Outcome (2): Poverty Free Years (capability) (PFY(c))

PFY(c) is the measurement of the probability of freedom from sufficient capability poverty over time. Whether an individual is classified as free from capability poverty or capability poor is dependent on the threshold of sufficient capability (TSC) – the capability poverty line, which is set to a lot of capability on the ICECAP-O attributes (i.e., 33333).

Table 41 presents the PFY(c) for the second running of BRAM12. 40,000 patients were required to reach the stopping rule confidence interval ratio of 1.2 for ICERs. From Table 41, it can be seen that ABT produces the highest number of PFY(c) at 4.75 years. Leflunomide has the lowest number of poverty free years at 3.72. The distance between most and least effective is 1.03 years. RTX compared with LEF produced the lowest cost per additional poverty free year of approximately £24,400.

Table 41 Poverty Free Years (capability) - PFY(c) & ICER (40,000 patients)

Option	Cost (£)	QSE	PFY(c)	QSE
1. ETN	75271	199	4.29	0.0167
2. IFX	73126	186	4.30	0.0167
3. ADA	75074	197	4.42	0.0168
4. RTX	69336	183	4.55	0.0184
5. ABT	92892	255	4.75	0.0179
6. LEF	49106	124	3.72	0.0153

Option	ICERvsLEF (£)	Quasi-CI [L-U]		ICERvsRTX (£)	Quasi-CI [L-U]	
1. ETN	45900	43200	49000			
2. IFX	41500	39100	44300			
3. ADA	37200	35400	39300			
4. RTX	24400	23300	25700			
5. ABT	42400	40900	44000	115000	95800	143000

ETN, etanercept; IFX, infliximab; ADA, adalimumab; RTX, rituximab; ABT, abatacept; LEF, leflunomide; PFY(c), Poverty Free Years (capability); ICER, Incremental Cost-Effectiveness Ratio; QSE, Quasi Standard Error

9.4.3 Capability Outcome (3): Years of Insufficient Capability (YIC)

YIC is the combination of the insufficient capability score (ICS) as calculated in the previous chapter with time. Unlike the three other capability outcomes, the aim is to minimise YIC rather than maximise.

Table 42 presents the results of the third run of BRAM12. 100,000 virtual patients were required to reach the 1.2 confidence interval stopping rule for the ICER, except for one ICER calculation which required 250,000 to get to an acceptable ratio level. The strategy with the best years of insufficient capability is abatacept (ABT) with 2.10 YIC. The worst (highest)

YIC is leflunomide (LEF) with 2.49. The difference between the highest and lowest YIC is 0.38. The strategy with the lowest cost per year reduction in YIC compared with leflunomide (LEF) is rituximab (RTX) at approximately £62,600.

Table 42 Years in Insufficient Capability (YIC) & ICER (100,000 patients)

Option	Cost (£)	QSE	YIC	QSE
1. ETN	75293	126	2.26	0.0054
2. IFX	73266	118	2.26	0.0054
3. ADA	74825	125	2.23	0.0054
4. RTX	69472	116	2.16	0.0053
5. ABT	92932	161	2.10	0.0052
6. LEF	49096	78	2.49	0.0057

Option	ICERvsLEF (£)	Quasi-CI [L-U]		ICERvsRTX (£)	Quasi-CI [L-U]	
1. ETN	114700	108500	121500			
2. IFX	105800	99800	112300			
3. ADA	100100	95200	104500			
4. RTX	62600	59900	65500			
5. ABT	114500	110600	118700	402800*	358100	459100

ETN, etanercept; IFX, infliximab; ADA, adalimumab; RTX, rituximab; ABT, abatacept; LEF, leflunomide; YIC, Years of Insufficient Capability; ICER, Incremental Cost-Effectiveness Ratio; QSE, Quasi Standard Error.
*250,000 patient simulations were required to get the confidence level ratio below 1.5

9.4.4 Capability Outcome (4): Years of Sufficient Capability (YSC)

YSC is the combination of the Sufficient Capability Score (SCS) as calculated in the previous chapter with time. It is the opposite of YIC, such that the aim is to maximise capability levels below sufficient capability, rather than minimise shortfalls in sufficient capability. This was explained in detail in the Chapter 7. YSC allows a direct comparison of the sufficient capability methodology with the previous maximising capability outcomes to be drawn.

Table 43 presents the final BRAM12 run. 100,000 patients were required to meet the 1.2 confidence interval stopping rule for ICERs. The ICER between RTX and ABT required

250,000 virtual patient runs. From Table 43, the highest YSC is for ABT at 11.01. The lowest YSC is for LEF at 10.431. The difference between lowest and highest is 0.58 YSC. The lowest cost per additional year of sufficient capability is for RTX compared with LEF, which costs approximately £38,800 for an additional year of sufficient capability.

Table 43 Years of Sufficient Capability (YSC) & ICER (100,000 patients)

Option	Cost (£)	QSE	YSC	QSE
1. ETN	75293	126	10.79	0.0165
2. IFX	73266	118	10.79	0.0154
3. ADA	74825	125	10.82	0.0154
4. RTX	69472	116	10.96	0.0159
5. ABT	92932	161	11.01	0.0157
6. LEF	49096	78	10.43	0.0150

Option	ICERvsLEF (£)	Quasi-CI [L-U]		ICERvsRTX (£)	Quasi-CI [L-U]	
1. ETN	73100	67300	80200			
2. IFX	67100	61800	73500			
3. ADA	65700	60800	71400			
4. RTX	38800	36600	41400			
5. ABT	75600	71700	80000	449000*	322000	742000

ETN, etanercept; IFX, infliximab; ADA, adalimumab; RTX, rituximab; ABT, abatacept; LEF, leflunomide; YSC, Years of Sufficient Capability; ICER, Incremental Cost-Effectiveness Ratio; QSE, Quasi Standard Error
*250,000 patient simulations were required to get the confidence level ratio below 2.5

9.4.5 Comparison of Model results

Table 44 presents a summary of the differences in BRAM model runs for each outcome respectively. In comparing the capability outcomes from BRAM12 with the QALY outcomes from BRAM11, there appears to be very few differences between the orderings of drug treatment strategies. The sensitivity of change between highest and lowest outcome in terms of effectiveness in capability outcomes ranges from 1.03 for PFY(c) to 0.38 for YIC. The ICER for unit change in capability outcomes ranges between £24,400 and £62,600.

Table 44 Summary results from BRAM12 Case Study

BRAM version	Outcome	Strategy ordering	Difference most-least effective	Range	ICER RTX vs. LEF	WTP (£) for unit change?
BRAM11	QALYS	5,4,3,2,1,6	1.15 QALYs	[3.10-2.13]	£21100 per QALY gain	<30,000
BRAM12	YFC	5,4,3,2,1,6	0.56 YFC	[9.91-9.35]	£42700 per YFC gain	???
BRAM12	PFY(c)	5,4,3,2,1,6	1.03 PFY(c)	[4.75-3.72]	£24400 per PFY(c)	???
BRAM12	YIC	5,4,3,2,1,6	0.38 YIC	[2.10-2.49]	£62600 per YIC avoided	???
BRAM12	YSC	5,4,3,2,1,6	0.58 YSC	[11.01-10.43]	£38800 per YSC gained	???

QALYs, Quality Adjusted Life Years; YFC, Years of Full Capability; PFY(c) Poverty Free Years (capability); YIC, Years of Insufficient Capability; YSC, Years of Sufficient Capability; RTX, rituximab; LEF, leflunomide; ICER, Incremental Cost Effectiveness Ration; WTP, willingness to pay of decision-makers for a positive unit change in outcome; ???, value unknown.

9.5 DISCUSSION OF EACH MODEL RUN IN BRAM12

9.5.1 Capability Outcome (1): Years of Full Capability (YFC)

In the first model run of BRAM12, Years of Full Capability (YFC) are produced to measure the overall benefit of treatment, using the ICECAP-O valuation dataset to anchor capability scores on a 0-1 no capability-full capability index. The treatment strategy commencing with ABT produced the highest YFC (9.91), whilst LEF produced the lowest YFC (9.35). Applying the incremental cost-effectiveness ratio (ICER) for costs and YFC for different treatment strategies, the most cost-effective treatment strategy compared with standard treatment (LEF) is RTX at £40,200.

This is the first attempt of modelling a capability outcome from an economic model used previously to aid decision-making in the UK. YFC allows for the full incorporation of the ICECAP-O index to be implemented in a capability outcome, where the objective is to maximise total levels of YFC within the population. It shows that YFC can be generated for an economic model by mapping from physical function (HAQ-DI) to capability (ICECAP-O), which has been described in detail in Chapter 8. This allows for the generation of capability outcomes in a wider variety of clinical areas in health where capability data such as ICECAP-O is not widely available or collected over a long time period.

Nevertheless, there are some limitations to this research. As with the generation of QALYs in BRAM11, the generation of YFC relies on the change in physical function (HAQ-DI) within BRAM12. Therefore, the primary driver of change in YFC is the change in physical function. If capability outcomes such as YFC are aiming to capture broader benefits than health alone, such a reliance on a measure of physical function could prove misleading in terms of the true capability benefits for the population under consideration.

Given that the interest in this research is in the change of benefits from the outcome measure used in BRAM11 (QALYs) with those used in BRAM12, no other changes to BRAM12 were made. A full evaluation using a decision model applying the capability approach would also aim to capture a broader range of costs associated with the societal perspective, rather than just the NHS and personal and social service costs as currently recommended (NICE, 2013). Additionally, the same method of discounting QALYs at 3.5% per annum is also in place for capability outcomes here. The issue of whether discounting outcomes in health economics is applicable for capability outcomes is not explored here, but it is likely to raise similar questions as those raised with discounting of QALYs (Gravelle & Smith, 2001; Brouwer et al., 2005; Gravelle et al., 2007; Claxton et al., 2011).

No other study has tried to generate capability outcomes from an economic model. One study has compared ICECAP-O with health utility measures such as EQ-5D and SF-6D, in order to generate QALYs using the capability instrument (Rowen et al., 2012). However, there is no particular rationale for implementing the ICECAP-O in a QALY framework. The QALY is the benchmark outcome in health economic evaluations and there appears to be a certain amount of support to input any new measure within health into the current dead-full health QALY index criterion (Rowen et al., 2012). For this to be implemented in practice, a close crossover of scales is required between measures (Brazier et al., 2004a). So far, there is conflicting evidence as to whether the ICECAP-O and the EQ-5D-3L are complements (Davis et al., 2013) or substitutes (Makai et al., 2013) for each other. So far, this research has been limited to small sample sizes and particular patient groups. Further research is required to verify these results and whether or not the instruments could, or should, be used interchangeably.

For policymakers interested in maximising levels of capability across a population, YFC offers a method for implementing such an objective. In order for capability questionnaires to be comparable with YFC from the ICECAP-O, such as Adult Social Care Outcome Toolkit (ASCOT) (Netten et al., 2012) and OxCAP (Lorgelly et al., 2008; Anand et al., 2009) questionnaires, they would need to be re-scaled in a similar approach to the ICECAP-O valuation index (Coast et al., 2008a). This would allow for meaningful comparisons to be made across studies, on a no capability to full capability index.

9.5.2 Capability Outcome (2) Poverty Free Years (capability) (PFY(c))

The second run of BRAM12 produced Poverty Free Years (capability) (PFY(c)), an outcome predicting the likelihood of a population free from capability poverty throughout their time during treatment. For BRAM12, PFY(c) is calculated over the remaining lifetime of rheumatoid arthritis patients. The treatment commencing with ABT produced the highest PFY(c) at 4.7547 poverty free years, while LEF yielded the lowest PFY(c) of 3.7215. The most cost effective treatment compared with LEF is RTX at £24,400 per additional PFY(c).

This is the first attempt to incorporate poverty methodology within a health economics framework. Poverty, in this scenario, refers to capability poverty – a level of capability which is insufficient to have a decent level of individual well-being. By implementing the headcount ratio methodology (Sen, 1976; Foster et al., 1984) within an outcome, the overall likelihood of falling below a decent level of capability well-being within a population can be judged.

Additionally, poverty and capability have rarely been measured over time (Alkire et al., 2008) to the same degree as health economic outcomes are extrapolated over time in decision models. Therefore the calculation of PFY(c) represents another first as it is a combination of a measure of freedom from capability poverty over the lifetime of patients within the model.

The limitations concerning predicting capability from physical function are relevant here as for all capability outcomes in BRAM12. Furthermore, there are weaknesses apparent for the PFY(c) outcome specifically. Measuring capability poverty using a binary outcome of poverty or no poverty suffers from the notable drawbacks of using a headcount ratio, which were explored in detail in Chapters 4 and 7. In summary, for an outcome like PFY(c), the depth or severity of poverty within the population cannot be ascertained. Thus, it can be misleading when many people just below the poverty line improve enough to cross the poverty threshold, while those worse off neglibly improve their wellbeing. This has been shown previously when relying on a single indicator of poverty (income) which can lead to misleading poverty of population sub-groups in the United States. The African-American population sub-group was evaluated as poorest when using income only, but when other factors were accounted for such as educational attainment, health status and health insurance, the Hispanic group was the worst off in that jurisdiction (Alkire & Foster, 2011a). A headcount ratio such as PFY(c) can therefore overestimate the improvements within a population and should be handled with caution.

The measurement of poverty presented here in terms of PFY(c) has not been attempted before in a health economic evaluation framework and is recognised as an area which has been neglected within the discipline (Mooney, 2009). In this scenario, poverty is put forward in terms of shortfalls in sufficient capability, with the aim of minimising the levels of capability poverty, falling short of the threshold of sufficient capability required for a decent standard of living. Therefore, poverty in this scenario is just as applicable to developed nations as it is to human development in developing countries.

The issue of addressing equity within health economic outcomes is a key aim for many health economists looking for an alternative way of assessing improvements in individual well-being (Coast, 2004; Cookson et al., 2009; Culyer & Bombard, 2012). Numerous attempts have tried to account for equity in terms of fair-innings (Williams, 1997), proportional shortfall (Stolk et al., 2004) and severity of illness (Nord et al., 1999), with varying degrees of success. More recently, equity weights have been attached to QALY outcomes by the social value of a QALY team (Baker et al., 2010). However, for all of these attempts, there appears an inability to consistently apply such methods in generating economic outcomes or in defining what equity issues are of primary importance in evaluations, although there has been some success in the Netherlands, primarily the proportional shortfall method (van de Wetering et al., 2013).

The capability approach allows researchers to apply a consistent theoretical basis to evaluations and focus on helping those with the lowest levels of capability as a priority (Sen, 1985; Sen, 2009). Having such a theoretical basis can allow for a clear and coherent framework for evaluation, where generating meaningful outcomes is just one of a number of key changes required to evaluations from the current health economic paradigm. PFY(c) offers one of a number of potential capability outcomes which can be used in this mode.

The production of PFY(c) allows for a measure of prevalence of capability poverty within health economic evaluations and how this changes over time with different interventions. Whilst not tested here, an outcome of prevalence of poverty in terms of health could be applied in a similar manner. However, there is a theoretical basis for poverty measurement within the capability context. Such use of a poverty measure of health would have to be rationalised as to why it would be useful to decision-makers. For those policymakers interested in overall levels of capability poverty, PFY(c) provides a practical outcome which can be generated in economic evaluation formats such as BRAM12 here.

9.5.3 Capability Outcome (3): Years of Insufficient Capability (YIC)

YIC is the first attempt to incorporate the multidimensional poverty methodology, specifically the adjusted-Foster Greer Thorbecke (FGT) or M2 poverty measure (Alkire & Foster, 2011a) into an outcome which can be tracked over time. The calculation of insufficient capability score (ICS) is combined with time to give the prevalence of poverty within a population, as well as accounting for severity of poverty to get YIC, with the aim of minimising levels of YIC. In the third run of BRAM12, the treatment strategy commencing with ABT produced the lowest YIC at 2.1049, while LEF produced the highest YIC at 2.4876. The cost of reducing YIC by a year is lowest for RTX compared with LEF at £62,600 per year of insufficient capability averted.

YIC represents a direct application of the adjusted-FGT measure (Alkire & Foster, 2011a) over time and accounts for the severity of capability poverty within a population. The measure of YIC overcomes the particular weaknesses outlined for PFY(c) in terms of addressing the sensitivity of capability poverty, giving additional weight to lower capability levels.

The calculation of YIC in BRAM12 is limited by the fact that it is predicted from a measure of physical function (HAQ-DI), a limitation for all capability outcomes analysed here. Furthermore, YIC aims to minimise shortfalls in sufficient capability and is therefore not directly comparable with the other capability outcomes in BRAM12 or QALYs from BRAM11, which look to maximise their specific objectives respectively. YIC appears more suitable to draw comparisons with DALY calculations, which look to reduce morbidity and mortality for populations. However, such data were not available for comparison here. Also, given that the objective is to reach zero levels on YIC, a complication with the current BRAM12 model is that zero levels are comparable to state of being dead, which represents QALY zero levels in the original BRAM11. This is not ideal for YIC as its optimum is comparable with a state which would not be the objective of decision-makers (i.e. being dead). While the YIC does not produce any surprising results in relation to this complication here, it is an issue that would need to be addressed in further modelling using this unique capability outcome considered here.

This is the first attempt to apply multidimensional poverty methodology in an outcome that accounts for changes in capability over time. Previously, the multidimensional poverty methodology has been applied in global studies, within a multidimensional poverty index (MPI) now routinely collected by the United Nations (UN) to compare levels of poverty across a number of dimensions internationally (Klugman, 2010; Klugman, 2011). However, there have been relatively few attempts to quantify levels of capability over time, as well as combining such outcomes with population values attached to different capability states (Chiappero-Martinetti & Roche, 2009). This is where work carried out in health economics can aid the development of the methodology further, by combining the methodology outlined by Alkire and Foster (Alkire & Foster, 2011a; Alkire & Foster, 2011b) with common health economic techniques applied to assess changes in states of well-being over time for individuals.

The YIC offers a method for policymakers to assess the shortfalls of sufficient capability within a population, with the aim of minimising shortfalls. It accounts for severity of shortfalls in sufficient capability, something which is unaccounted for in the PFY(c) calculation. It could also potentially offer an alternative mechanism to DALYs (Murray & Lopez, 1996; Fox-Rushby & Hanson, 2001). DALYs are used frequently within assessing health in developing countries. However, assumptions in relation to age weights inherent in DALY calculation have caused considerable controversy (Anand & Hanson, 1997; Arnesen & Nord, 1999). Therefore the YIC outcome could provide a viable alternative to those who have objections to the underlying basis of the DALY.

9.5.4 Capability Outcome (4): Years of Sufficient Capability (YSC)

YSC allows for the combination of the Sufficient Capability Score (SCS) with time, so that the outcome can assess sufficient levels of capability over time. It offers a reversal of the YIC calculation, in that the aim is to maximise capability levels to sufficient capability, rather than minimise shortfalls of sufficient capability levels. This allows for a direct comparison with the other two capability outcomes from BRAM12 as well as the original QALY results from BRAM11. In the final run of BRAM12, the strategy commencing with ABT produced the highest YSC of 11.0105, while LEF generated the lowest YSC of 10.431 years of sufficient capability. The most cost-effective treatment in comparison to LEF is rituximab (RTX) which costs £38,800 per additional YSC.

This study shows how to implement the most complex multidimensional poverty methodology using an outcome measure that specifies the relevant attributes and thus is more akin to outcomes found in economic evaluations. However, unlike the usual use of common health economic outcomes, like QALYs where the aim is to maximise, the YSC objective is to reach a sufficient level rather than the highest point achievable as the optimum. It offers a method for comparing a capability outcome incorporating the most complex multidimensional poverty methodology, which can be compared with QALYs and the other maximising capability outcomes generated here.

The YSC outcome generated here suffers from the same limitations as the previous three capability outcomes in BRAM12, relying on the relationship between capability and physical function to predict YSC. However, this is not an issue of conceptual validity; rather it is one of data availability. Along with the assumptions made in the original BRAM11, this limitation is not unique to this research nor to the use of capability measures, but is a broader issue in health economics when relevant measures have not been collected alongside a trial. However, the relationship between physical function and capability is not well established, with the only primary research to date that has explored this relationship being limited to the exploration in Chapter 6 of this thesis.

There is no clear method as to how equity should be incorporated into the current health economic paradigm, with numerous equity considerations of potential importance to decision-makers (Dolan et al., 2005b; Culyer & Bombard, 2012). Methods range from adjusting QALYs with equity weights (Baker et al., 2010) to having a list of all equity considerations to be reviewed in order of importance on a case-by-case basis (Culyer & Bombard, 2012). There is no consensus to as to which method is most appropriate in the UK. A capability framework which addresses concerns about shortfalls in sufficient capability in an outcome over time offers an alternative to such attempts to incorporate equity implicitly into evaluations. YSC and the other capability outcomes considered in this chapter address equity considerations in terms of shortfalls of sufficient capability explicitly within an evaluation between different interventions.

Some developed nations have come to consensus as to how equity should be incorporated into economic evaluations such as the Netherlands who have adopted the method of proportional shortfall (van de Wetering et al., 2013) and Norway which adopted a scale of severity of illness to include alongside traditional QALY outcomes in evaluations (Nord, 2012). However, no such consensus is applicable to all nations, with diverse values on equity likely in different jurisdictions. Addressing equity explicitly in the primary economic outcome has the advantage of knowing that certain values are built in to the outcome and therefore not troubled by how to combine measures of equity and efficiency simultaneously. By shifting the objective of the capability outcomes such as YSC here to maximising levels of sufficient capability, the optimum criterion shifts focus further to those individuals who are deemed to be of primary concern in terms of their shortfalls of sufficient capability.

YSC allows decision-makers to maximise the achievement of capability gains in those individuals deemed not to have a level of sufficient capability within the population under evaluation. The aim is to increase levels up to the threshold of sufficient capability, with changes above this threshold excluded from analysis. The YSC offers the blending of multidimensional poverty methodology with an outcome that can track changes in time, a common feature in health economic evaluations. However, the rationale for the TSC needs to be based on solid reasoning by the researcher, decision maker or society overall *a priori*.

Further research would be required for a TSC which reflected a societal valuation of sufficient capability. Finally, the maximisation objective for YSC may be a more familiar and therefore convenient tool over the minimisation objective also considered with YIC in helping to aid the decision-making process.

9.6 DISCUSSION COMPARING BRAM12 MODEL RESULTS

In this case study, BRAM12 generated four capability outcomes, the first such undertaking for both economic modelling and measuring capability over time. This chapter gives a menu of capability outcomes to decision-makers looking to apply broader benefit measures in evaluations between competing interventions. It shows that it is feasible to generate capability outcomes from a decision analytic model, relying on a relationship between physical function and capability to generate meaningful outcomes.

The results in BRAM12 suggest that there is no variation in the ordering of strategy outcome benefits from the QALY BRAM11 results, which is more than likely due to the HAQ-DI having a prominent role in providing the basis of change in well-being for all model outputs. The difference between the highest and lowest benefit strategy varies for the four capability outcomes, as can be seen in Table 44. PFY(c) produce the highest difference between most and least effective. YIC proved the smallest difference to change from lowest to highest. Similar sensitivity is observed between YFC and YSC. However, considering all outcomes have different index ranges compared to QALY calculations from BRAM11, such differences should be handled with care in terms of judging the most appropriate capability outcome.

This is the first attempt to explore the feasibility of developing and using potentially appropriate capability outcome for aiding decision-making in health. It compares four capability outcomes, outcomes which have been theoretically and methodologically explained in detail in the previous two chapters. The methods used in BRAM12 offers guidance and practical advice on how to generate capability outcomes, which can produce meaningful comparisons between interventions.

There are limitations associated with relying on this information solely to judge the most appropriate capability outcome to recommend for common application. The primary limitation here is the inability to compare outcomes here with another intervention to assess the change in outcomes in a resource allocation decision, when only one intervention can be funded. However, the work that has been carried out here will allow others to compare the outcomes generated here with their research as a comparator, which will help to assess the validity of the capability outcomes further.

There is no accepted willingness to pay for an additional unit change in capability outcomes which decision-makers have become accustomed to with QALYs. However, considering the stage of development of the capability outcomes, it is not surprising that willingness to pay thresholds for capability outcomes are not available. Again, once the capability outcomes are tested for different interventions, a greater understanding of the likely willingness to pay for additional capability outcome gains can be established.

That all capability outcomes here relied on the HAQ-DI measure of physical function is a clear drawback. However, this weakness is not only limited to this study but also traditional economic outcomes generated from models such as BRAM11, where health utility is predicted from other routine data collected in clinical trials.

Health interventions for older people (Grewal et al., 2006), public health (Lorgelly et al., 2010a), complex interventions (Payne et al., 2013) and social care (Netten et al., 2012) have already been identified as areas where adopting a broad capability approach theory to evaluation could prove worthwhile. Until now, there has been no formulation of how an evaluation framework adopting capability theory would take place. Rather it has been suggested in open terms as an alternative to current practice (Anand & Dolan, 2005; Coast et al., 2008c; Ruger, 2010b; Smith et al., 2012), with little formulation as to how an evaluation framework based on capability could be implemented. This chapter shows that capability outcomes can be generated from an economic model and implemented to evaluate changes in capability between different interventions.

While this research is the closest attempt to implementing a full capability evaluation, a number of significant questions remain. This chapter has presented a menu of capability outcomes that could be useful in a decision-making context. If the decision of most appropriate capability outcome is judged in a traditional health economics way of the most sensitive to change of all potential measures (Grieve et al., 2009; Turner et al., 2013), then the PFY(c) outcome would be the most appropriate tool (see Table 44 for sensitivity between different interventions). However, this method of deciding which health utility instrument to use is not transferable to deciding which capability outcome is most appropriate. If this were the case recent evidence of the sensitivity to change is greater on the EQ-5D-3L than the ICECAP-O (Couzner et al., 2013b), but both questionnaires are on different scales so are not directly comparable. From a theoretical perspective, the PFY(c) has a number of limitations that can give misleading results. Using both sensitivity and theoretical rationale, the YSC is likely to be most appropriate, which allows for a wider application of capability instruments to generate capability outcomes more so than some of the other outcomes explored in this chapter. Further research is necessary to assess the ability to compare capability questionnaires which value shortfalls in sufficient capability differently.

While this chapter has presented a number of capability outcomes for decision-makers to apply, it cannot be determined from this research how much a decision-maker is willing to pay for a unit change per capability outcome. In Table 44, the comparison of the most cost-effective treatment for QALYs from BRAM11, along with the willingness-to-pay of NICE to recommend a treatment as cost-effective over another strategy is presented. While it is possible to compare the ICER for each intervention as an estimation for a likely lower threshold of willingness to pay for a unit of capability improvement (given the ICER in BRAM11 is £21,200/QALY – lower than the £30,000 cost per additional QALY recommended currently), such inference of a likely threshold for capability change for decision-makers from this single study is tenuous. Further studies would be required to get a better idea of the willingness to pay for capability outcomes.

Finally, this chapter, in particular, has focused on generating capability outcomes from an economic model. In order to apply a full capability based evaluation, a broader perspective of costs as well as benefits should be adopted. Inclusion of additional costs such as lost productivity and costs to family may have a significant impact on findings. Future research could explore a full capability based evaluation, measuring broader benefits and costs, as well as addressing how discounting cost and benefits should be handled from a capability perspective.

CHAPTER 10. THESIS DISCUSSION AND CONCLUSION

10.1 INTRODUCTION

In this thesis, a case has been argued for the adoption of an alternative theoretical basis for evaluating health interventions in economic evaluations, namely the capability approach. Many of the early criticisms of capability theory were concerned with the practical difficulty of capturing, within applied analysis, the richness of data on people's choices to live lives that were of value to them (Sugden, 1993). Developing outcomes for assessing individual well-being has been a challenging process for researchers interested in the adoption of the capability theory, but still, a number of contributions have been made within the health economics field to measure capabilities directly (Grewal et al., 2006; Coast et al., 2008a; Al-Janabi et al., 2012a; Lorgelly et al., 2008; Anand et al., 2009; Kinghorn, 2010). Until now, less attention has been placed on how such measures could be used within an evaluation framework.

The research presented here has been primarily concerned with how a capability evaluation space and objective for capability instruments could be practically applied within a health economic evaluation and still have close ties with the theoretical basis of the capability approach. The ICECAP-O questionnaire was the choice of measuring capability, due to the questionnaire's development through qualitative interviews to reflect the most important capabilities for the population under examination (Grewal et al., 2006) and also the fact that population values have been developed for different capability levels on the questionnaire's attributes (Coast et al., 2008a).

In the remainder of this chapter, the key findings from this thesis are outlined in Section 10.2. In Section 10.3 the strengths of the research undertaken are outlined, with a focus on the primary advances made within the thesis. In Section 10.4, a counter-balance to the strengths of the research is presented by discussing the limitations involved with the work undertaken.

In Section 10.5, attention is turned towards the research implications of the findings presented throughout the thesis. Emphasis is placed on the implications for health economists and decision-makers who are interested in a broader assessment of well-being within an evaluation framework compared with the current prevailing applications that focus on health alone. Section 10.6 presents further research questions which have been raised through the findings from this research. Guidance is offered on future research topics which would be of most benefit for moving the operationalisation of the capability approach, as presented here, towards a similar level of the current economic applications within healthcare. In Section 10.7, a conclusion is drawn from this discussion and the thesis overall.

10.2 KEY FINDINGS

There are a number of distinct findings from the work conducted in this thesis to the health economics and capability disciplines. They are (i) the development of a methodology, i.e. sufficient capability, that allows the assessment of capability levels through a practical framework within an appropriate evaluative space for the practical application of the capability approach; (ii) the development of capability outcomes, for which capability questionnaires can be used in practice to aid decision-making (iii) a better understanding of the relationship between capability and health related functioning as captured by two questionnaires.

10.2.1 The illustration and development of the sufficient capability methodology

This thesis represents the first attempt to operationalise the capability approach within a practical evaluation framework that can be used by decision-makers. The current "reference case" for health economic evaluations by the National Institute for Health and Care Excellence (NICE) within the UK consists of the implementation of the cost-utility analysis framework, with Quality Adjusted Life Years (QALYs) as the primary outcome of health benefit from treatment (NICE, 2013). The reference case that is currently employed by NICE to allocate resources across a healthcare system was challenged throughout the thesis and an alternative theory relying on Sen's capability approach was argued.

In Chapter 2, it was shown that alternative outcomes to the QALY are also focused on producing health outcomes, rather than a broader measure of well-being. While willingness to pay offers an alternative to health outcomes as a measure of utility, it falls short of a multidimensional evaluation space, which the extra-welfarism framework was developed to incorporate (Hurley, 1998; Brouwer et al., 2008). Therefore, a truly extra-welfarist approach measures more than one aspect of individual's well-being, rather than health or utility only. Therefore, for decision-makers who want to account for the effects of health interventions beyond the health of the patient alone, current empirical applications offer no method of accounting for such multidimensional benefits.

In Chapter 3, it was found that the justification of developing the extra-welfarist approach came from Amartya Sen's critique of welfare economic theory. Sen developed an alternative theory to welfare economics based on individuals capabilities, that is, the ability of individuals to do and ways of being in their life that are valuable to them (Sen, 1992; Sen, 1993; Sen, 2009). The capability approach was explained as an alternative to welfarism and the differences between extra-welfarism employed in health economics and the capability approach were discussed. A full application of the capability approach within health economics, as Sen intended, would require broader aspects of well-being to be captured than are currently within health related quality of life instruments used to produce QALY outcomes (Verkerk et al., 2001; Coast et al., 2008c; Smith et al., 2012).

Two conceptualisations of the capability approach within the health field have cautioned against the use of a monist outcome like QALYs (Ruger, 2010a; Venkatapuram, 2011). This is in spite of both conceptualisations focusing on the two areas that the QALY measure typically captures well, i.e. reduced morbidity and increased life expectancy. Some health economists have attempted to align capability theory with the QALY outcome (Cookson, 2005b; Bleichrodt & Quiggin, 2013). It was argued in this thesis that such QALY formulations are a dilution of the capability theory to fit within a traditional health economics approach, rather than shaping an outcome based on capability theory.

Other attempts have been made to capture capability directly within health economics. The OxCAP (Oxford Capability Questionnaires) group compiled questions from household and panel survey data to elicit capabilities from questions which could be construed as reflecting

Nussbaum's ten central human capabilities (Anand et al., 2009; Lorgelly et al., 2008; Simon et al., 2013). The Adult Social Care Outcomes Toolkit (ASCOT) outcome was developed to capture social care related quality of life to generate a social care QALY comparable with a health QALY (Netten et al., 2012). The ICECAP questionnaires directly attempt to measure capabilities, as opposed to the OxCAP indicators of capability (Anand et al., 2009). The ICECAP measure emphasise capabilities which are important to the UK population, developed through qualitative research (Grewal et al., 2006; Al-Janabi et al., 2012a). Given that the ICECAP questionnaires have population valuations attached to each attribute level on the respective questionnaire, it was decided that for the incorporation of the capability approach within health economic evaluations, the ICECAP measures are the most developed at this stage.

In light of the findings from Chapter 2 of the criticisms that have been levelled at the QALY, namely the narrow evaluation space and the health maximand decision rule, Chapter 4 attempted to identify alternative decision rules within the capability literature. From the studies that were identified as relevant within the comprehensive pearl growing review, the potential scope of capability assessed was vast with large numbers of dimensions of capability and health playing a major role in how many studies assessed capability well-being. Of particular interest was for the studies identified which focused on health, health itself was rarely the primary objective for measuring people's capabilities. Rather, the focus was on where poor health would have a negative impact on capabilities and therefore the person's overall well-being (Nikiema et al., 2012; Mabsout, 2011; Ferrer & Carrasco, 2010).

In terms of a decision rule for capability, while there were some examples of maximisation (Netten et al., 2012; Renouard, 2011; Tikly & Barrett, 2011), they were in the minority. Much of the capability research was focused on poverty analysis, in particular, the multidimensional influences on poverty. Much of this multidimensional poverty research focused on the avoidance of multidimensional states of poverty, rather than reaching an optimum capability state available. This was primarily led by the seminal work of Alkire and Foster on the formalisation of multidimensional poverty within the capability approach as a measurement of "unfreedom" (Alkire & Foster, 2011a).

Following on from the capability empirical literature review in Chapter 4, Chapter 7 attempted to develop a decision rule based on the concept of sufficient capability. The methods for sufficient capability drew primarily on the multidimensional poverty methods of Alkire and Foster. Chapter 7 developed a methodology as to how instruments which capture capability could be used within health economics by applying the sufficient capability approach. Unlike previous attempts to align the QALY with the capability approach, this work attempted to align the theoretical underpinnings of capability theory with a direct measure of capability well-being, the ICECAP-O questionnaire (Grewal et al., 2006; Coast et al., 2008a).

In Chapter 8, the methodology for choosing a case study to operationalise the sufficient capability methodology within an economic model was detailed. The chosen case study was the Birmingham Rheumatoid Arthritis Model (BRAM). Methods were implemented to incorporate the ICECAP-O capability measure within the BRAM model to produce sufficient capability outcomes. Capability outcomes were estimated through mapping between a proxy of the Health Assessment Questionnaire Disability Index (HAQ-DI) onto the ICECAP-O.

10.2.2 Development of capability outcomes

Before this thesis, little research had been carried out on how applications of the capability approach are used in practice. In Chapter 4, a review of capability applications suggests that the health maximisation objective of health economics does not necessarily transpire when applying the capability approach in practice. Therefore, previous attempts of re-interpreting the QALY as an outcome of "capability efficiency" (Cookson, 2005b) or "capability as menus" (Bleichrodt & Quiggin, 2013), without a consideration of whether the maximisation principle is applicable in a capability based evaluation.

In Chapter 7, an alternative proposal to the maximisation principle of health economic outcomes was presented. By employing the Alkire-Foster methods of multidimensional poverty, the most frequent method of aggregating capability attributes to form an indicator as discovered in Chapter 4, four potential capability outcomes were developed and illustrated in Chapter 7, based on a principle called "sufficient capability".

Finally, in Chapter 9, the results of the BRAM model case study were presented. A number of potential outcomes of sufficient capability were estimated. These outcomes were compared to the orderings produced with previous cost per QALY and incremental cost effectiveness ratio (ICER) results with BRAM. While the orderings of treatment strategies in BRAM did not vary with the original QALY orderings, when ICERs were introduced, the cost per unit of capability changed significantly depending on the threshold of sufficient capability and outcome implemented. Therefore, the choice of capability outcome and sufficient capability threshold is of crucial importance for the application of this approach in practice.

10.2.3 Understanding the relationship between health and capability wellbeing

The influence of health upon capability is essential for the adoption of the capability approach within health economics and related health disciplines. This thesis has shown the ability to predict capability levels on a questionnaire of capability well-being from a condition-specific health status questionnaire for OA patients.

Chapter 6 set out an agenda for exploring the relationship between capability and measures of health through the process of mapping between two instruments. Mapping was used to explore the relationship between a measure of condition-specific health status, the WOMAC Osteoarthritis Index, with a measure of capability, the ICECAP-O for a group of OA patients who required a hip or knee replacement. The aim in Chapter 6 was to show whether capability levels could be inferred from changes in health status. Given that much of the data available within clinical trials are often measures of health related well-being, it would be a big barrier for a generic measure of capability to overcome if it could not account for changes in health states.

In Chapter 6, it was shown that the categories of pain, stiffness and physical function on the WOMAC, were able to predict the majority of levels on the ICECAP-O attributes (security, role, enjoyment and control). The strongest relationship was between the physical function category and the control attribute on the ICECAP-O (R^2=0.2143). The attachment attribute levels on the ICECAP-O could, unsurprisingly, not be predicted from the WOMAC. While the WOMAC was not able to capture all changes on the ICECAP-O, all categories on the WOMAC showed some relationship with ICECAP-O attributes, adding strength to the argument that the capability questionnaire could be used in health economics as a broader outcome measure than health alone.

10.3 STRENGTHS OF RESEARCH

The main strengths of this thesis is that it is the first to (i) align the capability conceptualisations to health with those developing capability questionnaires; (ii) conduct a systematic literature search strategy for review of capability applications; (iii) explore the feasibility of mapping from condition-specific health status to capability questionnaires; (iv) develop a capability objective for economic evaluations; and (v) incorporate capability outcomes within an economic model.

10.3.1 Aligning capability conceptual theory with questionnaire development

In Chapter 3, a literature review of research which could be classified within the fields of the capability approach and health economics was presented. While this is not the first attempt to look at the types of capability outcomes available for use within health economics (Lorgelly et al., 2010a), it is the first attempt to compare the merits of different approaches to developing capability questionnaires within health economics. It is also the first attempt to align the two main conceptualisations of the capability approach in health with health economics (Ruger, 2010a; Venkatapuram, 2011). While both conceptualisations appear to resist health economic outcomes currently implemented, Chapter 3 attempts to align differences noted by the conceptual philosophers of the capability approach with the practical developers of capability questionnaires for healthcare assessment. This alignment was further guided by the development of an alternative objective of sufficient capability in Chapter 7, to the currently used objective of QALY maximisation.

10.3.2 Implementing an explicit literature search strategy for capability reviews

Chapter 4 completed an empirical literature review of the most up to date methods of measuring capabilities and using objectives and decision rules used for such outcomes. A multidisciplinary search was conducted so that a broad consensus of how the capability approach is currently being applied could be assessed. No other review of the application of the capability approach has targeted which attributes are captured by instruments, how such instruments are aggregated, as well as how such instruments could be used to aid decision-making within their respective fields.

Previous reviews of empirical capability literature were not explicit in how studies were chosen for their review or what kind of search strategy took place (Kuklys & Robeyns, 2005; Robeyns, 2006; Chiappero-Martinetti & Roche, 2009). The search strategy employed in Chapter 4 was the Comprehensive Pearl Growing strategy. Keyword search terms were implemented, as well as a two stage categorisation process which has been used in previous health economic systematic reviews (Roberts et al., 2002). The comprehensive pearl growing strategy is most appropriate when a keyword search would return tens of thousands of studies, which happens when words have multiple meanings and uses within the academic literature (e.g. costs, equity and capability). Given that the aim of the search was to assess the application across as many disciplines as possible, the pearl growing methodology was particularly suitable to the review as reported here (Dolan et al., 2005b).

10.3.3 The feasibility of mapping from health status to capability

This was the first study to map from a condition-specific measure of health to a measure of capability. This has led to a greater understanding of the relationship between the WOMAC categories of pain, stiffness and physical function with the capability well-being attributes on the ICECAP-O questionnaire. Previous research has suggested that the ICECAP-O may only be a complement to health status questionnaires such as the EQ-5D-3L (Davis et al., 2013). However, the ability of the WOMAC categories to predict ICECAP-O attribute levels would suggest that the ICECAP-O is capturing something completely separate to changes in health status somewhat of a contradiction to the findings of Davis and colleagues (2013).

10.3.4 The development of a capability objective for economic evaluations

Previous attempts to operationalise the capability approach within health economics have relied purely on economic methodology to conceptualise the way a capability evaluation could be carried out using the QALY outcome (Cookson 2005; Bleichrodt & Quiggin 2013). In Chapter 7, both methods from health economics and the findings from Chapter 4 on empirical research within the capability approach were used to develop a new objective. This objective is referred to as "sufficient capability", which incorporates the practicability of economic outcomes, as well as guidance from the capability approach as to what the objective of a capability evaluation should entail. All outcomes tested within the objective in Chapter 7 were also illustrated with patient data, unlike prior attempts to align the QALY measure with the capability approach, which relied purely on mathematical formulations and was not illustrated in practice using data (Cookson, 2005b; Bleichrodt & Quiggin, 2013).

10.3.5 The incorporation of capability outcomes within an economic model

The sufficient capability outcomes developed in Chapter 7 were also tested within a case study. The Birmingham Rheumatoid Arthritis Model (BRAM) model was used previously to aid decision-making in the UK. A number of potential capability outcomes that were developed in Chapter 7 were subsequently tested in the BRAM case study model. This was the first study to apply capability outcomes within an economic evaluation which analysed the different drug treatment strategies for rheumatoid arthritis patients. The modelling of capability outcomes over time within the BRAM represents the first measure of capability longitudinally, over a patient's life cycle as captured through the BRAM. This case study, detailed through Chapters 8 and 9, offers a substantial contribution to the operationalisation of the capability approach compared to previous attempts in the literature.

10.4 LIMITATIONS OF RESEARCH

Whilst this was the first time that research in this area has undertaken the focus of translating capability questionnaires (specifically the ICECAP-O) into model outputs, it must be noted there are limitations with the research conducted in this thesis. The areas in which limitations occur are (i) the framing of the capability application review; (ii) generalising the relationship between health status and capability; (iii) defining sufficient capability; (iv) methods for incorporating the ICECAP-O into an economic model; and (v) the BRAM model case study itself.

10.4.1 Capability empirical review search strategy

A limited timeframe was selected for study inclusion within the capability empirical review. This was done to focus on the current approaches of applying the capability approach in assessing well-being. Additionally, the search strategy was chosen to ease the burden of sifting through an insurmountable number of unrelated studies. Therefore, a search based over a longer time period and wider search strategy scope may have captured additional studies involved in this area. However, the search strategy choice and timeframe were judged to be suitable as the low likelihood of obtaining many additional studies, which would have had little benefit to the chapter objective of getting a consensus on the evaluative space and decision rule that applied the capability approach across disciplines.

10.4.2 The relationship between health status and capability

A relatively small sample size of 107 patients requiring joint replacement was used for the mapping study, although smaller mapping sample sizes have been used in practice (Brazier et al., 2010). Therefore, any inferences from the results of Chapter 6 must account for this limitation. The validity of the mapping was limited to internal validation of the dataset for osteoarthritis patients, with follow-up responses used to check the predictions from baseline. While this validation approach has precedence from a previous mapping study involving

WOMAC (Barton et al., 2008), the clear recommendation is that mapping functions should be externally validated when an external dataset is available (Dakin et al., 2013b).

One study with the ICECAP-O indicated that the capability measure may be a complement rather than a substitute to the EQ-5D for patients of a falls prevention clinic (Davis et al. 2012). A contrasting study has suggested that the ICECAP-O is able to capture health and wellbeing more generally for older hospital patients (Makai et al., 2013). No direct comparison with a measure used to generate QALYs was carried out in this thesis. This would have given a stronger justification for the use of a measure of capability well-being instead of a health related quality of life instrument in economic evaluations.

Other mapping studies have attempted to include additional data beyond the two measures in the mapping process, e.g. clinical or socio-demographic data. The prediction of capability may have been improved slightly for the patients from WOMAC to ICECAP-O if a measure of psychological well-being had been included. However, this broader approach also has the limitation which makes such mapping algorithms less useful when only WOMAC data has been collected, so this was deemed an acceptable sacrifice for the research undertaken in chapter 6.

10.4.3 Defining Sufficient Capability

The testing of the sufficient capability methodology relied on one patient group, so a comparison between different population groups could not be carried out. Therefore, the decision rule could not be tested between two population groups to show cases where different interventions would be prioritised over others.

The thresholds of sufficient capability tested were set at levels between the highest and lowest capability levels to show how the sufficient capability methodology would work in practice. Developing "sufficiency" in terms of capability is likely to vary between different populations. However, no determination of what sufficient capability should or should not be was made within the thesis.

Compared with previous methods of applying the capability approach, while a number of alternatives were tested, no specific outcome is recommended beyond the further exploration of the sufficient capability approach. This is in comparison with the clear objective of the maximisation of a specific outcome (QALYs) by other researchers (Cookson, 2005b; Bleichrodt & Quiggin, 2013).

10.4.4 Methods for incorporating ICECAP-O into an economic model case study

To incorporate the ICECAP-O into the BRAM, a mapping function between the HAQ-DI and the ICECAP-O was required. Previously a mapping between the HAQ-DI and the EQ-5D was used to calculate QALY gains from the different drug treatment strategies for rheumatoid arthritis patients within BRAM (Malottki et al., 2011).

While the majority of questions from the HAQ-DI were collected in the Tayside dataset, the questionnaire was not completed in its entirety within the mapping sample. The primary objective of the dataset was to develop a new Aberdeen measure of impairment, activity limitation and participation restriction using item response theory across a number of related questionnaires, resulting in HAQ-DI questions being excluded from collection (Pollard et al., 2009).

To calculate the necessary six category scores to complete a HAQ-DI overall score, some questions from the WOMAC and SF-36 were used as proxies. While this is far from the ideal scenario, these were the best data available to explore a capability measure in an economic model. A further limitation was that the mapping functions were validated within the baseline dataset. This was as a consequence of the data used to calculate the adjusted HAQ-DI was not collected at follow-up. Only the WOMAC questionnaire and the ICECAP-O were completed at 1 year and 3 year follow up.

10.4.5 The BRAM case study

Due to time and budget constraints of the Ph.D. studentship, an original case study which was developed from inception was outside the remit of this research project. This was primarily due to the lack of capability data collected over a period of time long enough to develop a case study and the fact that the capability measures are emerging. Therefore, a body of data does not already exist. Case studies with direct data collected over a reasonable time period are likely to have more certainty with their findings. However, this was not possible here. For the purposes of testing and generating capability outcomes in a model previously used to aid decision-making in health, the methods adopted within the BRAM case study were deemed acceptable for the purposes of this thesis. The potential for a number of case studies was explored. However, due to the extent of the research required to carry out one case study and the lack of relevant ICECAP data, the BRAM12 model was the only case study carried out in this thesis. Ideally, an area where the capability approach is believed to be beneficial, such as public health interventions (Lorgelly et al., 2010a; Saith, 2011), would have been chosen for a case study and compared with the findings from BRAM12.

Within the selected case study, only the outcomes from the BRAM model case study were changed. This was so that the focus was on changes from the outcome side due to the incorporation of the ICECAP-O into the model and not as a result of other changes which could have been made. Therefore, aspects within the model on the costs side, which focus on costs to the NHS and personal and social services only, may be broadened within a full capability evaluation to societal costs. This was beyond the objectives of this thesis, as its primary focus is on the benefit aspect of economic evaluation.

The method of comparing costs with capability outcomes was the incremental cost-effectiveness ratio (ICER). This method was used because there was no alternative metric for combining the changes of costs and effects from the studies explored from the capability literature in Chapter 4. The ICER did have the advantage of being a direct comparison with the previous QALY output from the original BRAM results. While alternatives to ICERs may be an option for a full capability evaluation, it was not investigated in any great detail here.

10.5 IMPLICATIONS OF FINDINGS FOR HEALTH ECONOMICS AND DECISION-MAKERS IN HEALTHCARE

There are a number of important implications from this research. They are (i) a new method for allocating resources using the capability approach as the theoretical basis; and (ii) the development of sufficient capability as a method of incorporating distributional concerns within the outcome of interest.

10.5.1. A new method for allocating resources across healthcare

This thesis offers an alternative framework for aiding decision-making in health, which attempts to incorporate more than health improvement into the outcome measure. For decision-makers interested in allocating resources across a health service, this research will be of particular interest to those who believe economic outcomes should capture a broader impact on individual well-being, rather than health only.

Currently, health economic evaluations are primarily employed within the health service for measuring technical efficiency (i.e. choosing one of a number of potential interventions for a given population group). This would seem an appropriate use of the current QALY, given the different methods of calculating health status within QALYs (e.g. EQ-5D, SF-6D, HUI3), making comparisons between QALYs across a health service nonsensical. While this may be an adequate approach for an ever expanding health service with infinite resources, the recent years of curbed spending within the NHS would suggest that such an approach may not be helpful in allocating resources efficiently, which includes the withdrawl of non-effective interventions. Efficiently, as understood in this thesis within the capability approach, is the achievement of a sufficient level of capability for the maximum number of the population, given the finite resources available for a given period. When people have sufficient capability, resources are aimed at those with the lowest levels below the threshold of sufficient capability, with the aim of improving levels to a society's given threshold point of sufficient capability. This methodology is focused on allocating resources in healthcare, but if sufficient capability was a main criterion across public services, this could allow for comparisons of impact of capability wellbeing between resources from different public sectors attempting to achieve similar goals (e.g. education versus sports funding versus health interventions for reducing obesity).

10.5.2 Sufficient Capability and Extra-Welfarism

The sufficient capability approach also gives an alternative method for applying the capability approach within health economics outside of the QALY measurement. Two notable conceptualisations of the capability approach within health rejected the use of the QALY (Ruger, 2010a; Venkatapuram, 2011), so the approach offered here may be more in line with the objectives of capability theorists too. Health economists who are interested in measuring broader benefits than health alone may also favour this approach.

A number of ways of handling distributional concerns with QALYs in the current extra-welfarist framework have been made. The Dutch approach is the inclusion of such equity weights within the QALY, namely the fair innings approach combined with health maximisation to calculate proportional shortfall (Stolk et al., 2004). Other studies have done similar research, albeit with different objectives from the Dutch, which has been referred to a the social value of a QALY project (Baker et al., 2010). The Norwegian approach outlines equity considerations alongside the QALY outcome, keeping concerns from equity separate from the QALY score (Nord, 2012). A similar approach has attempted to develop an equity checklist which a decision-maker should account for as well as the QALY calculation (Culyer & Bombard, 2012). While there does not appear to be a general consensus about the method of incorporating equity information given to decision-makers in QALY outcomes, research like the studies just referred to would indicate that the QALY will continue to play an important role in aiding decision-making in health, in the short-term at least. However, it is argued in this thesis that the problems with the health maximising objective used for QALYs is down to a combination of the limited evaluative space on health alone and an inappropriate objective of QALY maximisation from a capability perspective.

The sufficient capability approach may also allow a mechanism for the comparison across well-being from different areas of public policy. While questionnaires such as ICECAP were specifically developed for use within the health and social care setting (Grewal et al., 2006), the questionnaires give no direct consideration of health within the questionnaire. While some may see this as a disadvantage within the health setting, the flexibility of the non-specific measures of capability well-being may offer a truer guide as to what areas of public funding are contributing most to a general population's overall well-being. The lack of a direct measure of health does not appear to be a problem as changes in functional health are captured within the ICECAP-O measure, as explored in Chapter 6.

Given that the capability approach is a direct critique of welfare economics, it is not so clear whether welfare economists will be sympathetic to the approach outlined here. However, this thesis has attempted to outline where the true “extra” in “extra-welfarism” can be found (Birch & Donaldson, 2003), within a more direct representation of Sen’s capability theory in an evaluation framework.

10.6 FUTURE RESEARCH RECOMMENDATIONS

While this thesis has addressed a number of issues with the applications of a capability approach within economic evaluation, a number of lingering questions remain following this research.

10.6.1 The relationship between capability and health measures

As noted in Section 10.4.3, research with the ICECAP-O questionnaire and the EQ-5D-3L have reported conflicting arguments about the role of the ICECAP-O as a complement (Davis et al., 2013) or a substitute (Makai et al., 2013) to the EQ-5D measure of health related quality of life. Research undertaken in Chapter 6 would lead to agreement with the latter argument of a substitution effect between health and capability questionnaires, considering that aspects of capability can be predicted from categories of pain, stiffness and physical function on the WOMAC Osteoarthritis Index (Mitchell et al., 2013). Given the limited sample sizes of both the study in this thesis and the previous studies mentioned above, future research should explore whether or not the ICECAP questionnaires can accurately assess changes in a health related measure and non-health (e.g. social care) of quality of life over time.

One method of exploring this relationship further would be the “probabilistic mapping” approach (Le & Doctor, 2011). A relatively large dataset would be required to test the substitution ability of a measure of capability with a measure of health related utility. This framework will also be useful to validate the WOMAC to ICECAP-O predictions for joint replacement patients when a larger dataset is available. The probabilistic framework produced lower error statistics than OLS or the response mapping approach from SF-12 to EQ-5D (Le & Doctor, 2011).

It is also important to note that the ICECAP-O and other capability questionnaires applied in health research have so far looked at measuring capability or measuring functioning. As was shown in the review of capability applications in Chapter 4, no attempt has been made to directly incorporate the third tenet of the capability approach within questionnaires, i.e. individual agency (see Section 3.2.4). While research has been undertaken to measure agency goals from pre-existing datasets (Alkire, 2005; Ibrahim & Alkire, 2007; Burchardt, 2009), little research has been undertaken on the combination and trade-offs of agency enhancement with capability and functioning to assess the overall well-being of an individual. How this concept of agency relates to health related quality of life could also provide an interesting avenue for future research.

10.6.2 Defining Sufficient Capability for a society

The sufficient capability approach offers a new rationale for the implementation of capability instruments within economic evaluations. An illustrative example of the sufficient capability outcomes was presented in Chapter 7, with a direct application through mapping in BRAM in

Chapter 9. However, a number of questions about the application of the newly developed methodology for health economic evaluations remain.

The sufficient capability approach applied here used the ICECAP-O (for 65 years), as this was the most developed questionnaire of the ICECAP family at the time of this research. The ICECAP-O questionnaire values also relied on the population age group the questionnaire is aimed at, i.e., the over 65s (Coast et al., 2008a). Attempts to apply the approach across a health service would necessitate a questionnaire more applicable to the general population, such as the ICECAP-A (for adults aged 18 and over) (Al-Janabi et al., 2012a). However, there is concern about how older people would answer the broader questionnaire, as the over 65 year olds struggled to interpret the achievement and progress category on the ICECAP-A (Al-Janabi et al., 2013). A mapping between the ICECAP-O and the ICECAP-A may allow for older people to complete the questionnaire developed specifically for them, yet be able to compute meaningful comparison with the rest of the general adult population's important capabilities.

While the case study showed how capability outcomes could be implemented within economic models, it was beyond the scope of the case study to ascertain the usefulness of a charge in capability outcomes for decision-makers. While it did take some time for willingness to pay thresholds of a QALY to emerge (McCabe et al., 2008), attempts are currently being made to systematically develop thresholds that decision-makers use to justify policy decisions on the basis of cost-effectiveness (Claxton et al., 2013). However, a similar undertaking to that of Claxton and colleagues (2013) was beyond the scope of this thesis, which focused on developing a capability methodology for evaluation, as well as how capability instruments could be applied within economic models with limited data through mapping. Future research could qualitatively explore decision-makers views on the meaningfulness in changes of outcome measures such as Years of Sufficient Capability (YSC) for patients in the health service.

Another area for further research which would take the findings from this thesis forward would be an example of outcome results between interventions. While the cost per additional QALY is the favoured method by NICE to compare outcomes, alternative methods for prioritising resources such as league tables may also shed light on how a sufficient capability approach may lead to different priority orderings across a health service. Ideally, this would present a scenario where capability outcomes show different results from QALY outcomes, to show that the capability approach offers an alternative that does not result in the same orderings as the QALY outcome. This is a method which has been used by advocates of the contingent valuation methodology, which have attempted to show differences in orderings between willingness to pay surveys and QALY outcomes (Sach et al., 2007).

10.7 CONCLUSION

The research presented in this thesis has attempted to implement the capability approach within an evaluation framework, something which was deemed impossible twenty years ago (Sugden, 1993). A sufficient capability methodology was developed from multidimensional poverty methodology to generate capability outcomes from the ICECAP-O capability questionnaire. This thesis sheds light on the relationship between health and capability questionnaires, showing that some questions of a broader measure of well-being than health can be estimated from condition-specific health questionnaires. While this thesis has done much to implement the capability approach within an evaluation framework, further research is required to show if the sufficient capability methodology offers a practical alternative to the current methodology applied in health economic evaluations currently.

From this research it can be said that condition-specific health status categories of pain, stiffness and physical function on the WOMAC questionnaire are able to predict the majority of capability attributes captured on the ICECAP-O for arthritis patients requiring joint replacement. The existence of this relationship allowed the prediction of relevant outcomes for a capability evaluation to be predicted from a BRAM model. However, further research is required on defining sufficient capability for a given society before this approach to allocating resources could be implemented across a health service. More research is also required to show whether a shift to the sufficient capability approach would lead to any change in priority given to certain interventions over others, compared with the current health maximisation approach used with health status instruments like the EQ-5D to generate QALY outcomes.

Appendix 1: The ICECAP-O Questionnaire(Coast et al., 2008a)

ABOUT YOUR QUALITY OF LIFE

By placing a tick (ü) in ONE box in EACH group below, please indicate which statement best describes your quality of life at the moment.

1. Love and Friendship

I can have all of the love and friendship that I want	☐	4
I can have a lot of the love and friendship that I want	☐	3
I can have a little of the love and friendship that I want	☐	2
I cannot have any of the love and friendship that I want	☐	1

2. Thinking about the future

I can think about the future without any concern	☐	4
I can think about the future with only a little concern	☐	3
I can only think about the future with some concern	☐	2
I can only think about the future with a lot of concern	☐	1

3. Doing things that make you feel valued

I am able to do all of the things that make me feel valued	☐	4
I am able to do many of the things that make me feel valued	☐	3
I am able to do a few of the things that make me feel valued	☐	2
I am unable to do any of the things that make me feel valued	☐	1

4. Enjoyment and pleasure

I can have all of the enjoyment and pleasure that I want	☐	4
I can have a lot of the enjoyment and pleasure that I want	☐	3
I can have a little of the enjoyment and pleasure that I want	☐	2
I cannot have any of the enjoyment and pleasure that I want	☐	1

5. Independence

I am able to be completely independent	☐	4
I am able to be independent in many things	☐	3
I am able to be independent in a few things	☐	2
I am unable to be at all independent	☐	1

Appendix 2: Data Extraction Sheet for Capability Application Review

Criteria	**Justification**
Name of author(s), title of study, year of publication	Summary information necessary for descriptive statistics
Are details available on the type of the attributes within the capability related measure?	Understanding the components of capability related measures across discipline
Type of application of the capability approach? For example poverty and well-being assessment in advanced economies.	The Robeyns' (2006) groups of studies where the capability approach has been applied should help to analyse similar studies together.
Was the capability related measure developed for a specific context? If so, which context?	It has been argued by those who have applied of the capability approach that measures can be developed to address a specific policy question
Country study conducted	Can the study findings be applied in a UK setting?
Was the study country/area specific or cross-national/disciplinary? Which country and what area of focus?	It is important to ascertain the potential for interdisciplinary research, as areas which are applied within a number of fields/countries, may be more adaptable to a health analysis setting
Are comparisons made between different population groups?	An important role in allocating resources is the commensurate nature of population comparisons
Objective of study?	Health maximisation, poverty reduction etc...
Are decision criteria/rules discussed? What methods were used?	If a measure has been promoted within a study, do the authors suggest how decision-makers should interpret such results for aiding decision-making?

Appendix 3: Studies included in review of capability applications

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Grewal et al., 2006)	2006	1	A1	x.	1	C	Attachment Security Role Enjoyment Control	ICECAP-O	Index of capability versus preference-based measure of health related utility
(Walker, 2006)	2006	1	A1	viii.	4	C	Autonomy Knowledge Social Relations Respect and recognition Aspiration Voice Bodily Integrity and bodily health Emotional integrity and emotions	Education capabilities	gender equity in capability & equality of capabilities in education
(Anand & Santos, 2007)	2007	1	A2	vi.	1	F	Sexual Assault Domestic Violence Violence Assault (3 attributes measured across eight indicators)	OCAP capability indicators	N/A
(Distaso, 2007)	2007	1	A1	i.	2	F	Consumption Income Distribution Life Expectancy (boys) Life Expectancy (Girls) Health Education Employment Pollution Aesthetic and cultural values	Multidimensional sustainability wellbeing index	Index of sustainability (-1-+1)
(Clark & Qizilbash, 2008)	2008	1	B1	iii.	4	F	Clean water Health Access to health care Housing Jobs Education Freedom Nutrition Safety Self-worth and respect Survival Religion	"core" poverty	Identifying the core poor as a priority
(Coast et al., 2008a)	2008	1	A1	x.	1	C	(same as Grewal et al. 2006)	ICECAP-O	Attributes are valued on a 1(full capability) to 0 (no capability) scale. Preference expressed for CCA

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Krishnakumar & Ballon, 2008)	2008	1	A1	i.	6	F	Knowledge Living conditions (2 basic capabilities measured across six indicators)	Capability Indices (CI)	Structural equation modelling approach to measuring the influences on capabilities
(Murphy & Gardoni, 2008)	2008	1	B3	iv.	8	C	N/A	N/A	Capability assessed by two separate thresholds. Minimum levels of capability attainment acceptable after a hazard. It can be tolerable for some individuals to fall below acceptable threshold, so long as lower capability attainment is temporary, reversible, and probability of falling below tolerable threshold is sufficiently small
(Wagle, 2008)	2008	1	A1	iv.	5	F	Education Condition of Health Treated with respect Occupational prestige Employment industry	The multidimensional poverty model	Across measures of economic well-being, social inclusion and capability: poverty is rated as follows: abject poor, very poor, poor and non-poor
(Walker, 2008)	2008	1	A1	viii.	1	C	Knowledge Social Relations Critical Thinking Imagination and empathy Recognition and respect Active and experiential learning Autonomy Confidence Active Citizenship Deliberative dialogues Having economic opportunities	Functional capabilities	Equality in capabilities for all students
(Anand et al., 2009)	2009	1	A2	iv.	1	F	Nussbaum's 10 central human capabilities (see Table X) (measured over 64 indicators)	OCAP Capability indicators	N/A
(Burchardt, 2009)	2009	1	B3	iv.	1	A	N/A	N/A	A definition called "capability as autonomy", to include the conditions in which goals, aspirations and preferences are formed
(Di Tommaso et al., 2009)	2009	1	A2	vi.	2	F	Bodily health Bodily integrity (measured across three indicators)	Sub-section of Nussbaum's central human capabilities	N/A
(Hobson & Fahlén, 2009)	2009	1	A1	vii.	2	F	Working times Flexibility Employment situation Perceived economic well-being	Measures of Work Family Balance in the European Social Survey	Shifting focus of WFB from activation and increased productivity to measures of agency freedom and the enrichment of the quality of life

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Kerstenetzky & Santos, 2009)	2009	1	A1	iii.	6	F	Be well-sheltered Be healthy Do gratifying work Enjoy good schooling level Have protected children Be free from hunger and undernourishment Dress adequately Enjoy access to public services Not suffer discrimination Live without fear Participate in community life Participate in the associative life of the city Be happy and proud oneself	Index of Freedom (IF)	Poverty as insufficiency of basic capabilities
(Rosano et al., 2009)	2009	1	A1	v.	2	F	Income Personal assets Property	equivalent economic situation indicator (ISEE) -	Equivalence scales to give the true extent of poverty versus traditional poverty lines
(Wagle, 2009)	2009	1	A1	iv.	5	F	Educational attainment Degree Health condition Occupation prestige	Capability Indicators	Arbitrary thresholds of poverty. Poor in one dimension = capability deprivation
(Young, 2009)	2009	1	A1	viii.	8	C	Functional life skills learning Cognitive life skills learning Interpersonal like skills learning Personal life skills learning as agency freedom Cross-cutting basic capabilities	Basic capability from learning	A threshold of basic learning
(Addabbo et al., 2010)	2010	1	A1	vi.	2	F	Healthy life Safety Knowledge Emotions Integration and affiliation Expression Participation	Gender Budgets	gender budgets could become a tool for assessing the gender division of labour, the distribution of resources and the share of individual and public responsibilities
(Barrientos, 2010)	2010	2	B1	iii.	6	F	Housing Employment Health Income Household dynamics Education Registration	Basic threshold of social protection	Chie Solidario aims to equalise capability by ensuring the poorest can achieve a minimum set of basic functionings
(Clark & Hulme, 2010)	2010	1	B3	iv.	8	C	N/A	N/A	Making a case for the incorporation of time into the measurement of poverty, often neglected.

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Ferrer & Carrasco, 2010)	2010	1	A1	x.	8	C	18 Questions related to capability for changes in diet or physical activity	Assessing Patients' Capability in Diet and Physical Activity	Clinical success depends on events outside the control of the health care system: patient's ability to manage their health behaviours should be of focus
(Floro & Pichetpongsa, 2010)	2010	1	B1	vi.	3	F	Personal income Work intensity Education attainment	Individual well-being index (WBI)	Poverty line measures of income do not show the effect and work effort of these home-based workers. Improvement through WBI (0-1) more N.B.
(Gardoni & Murphy, 2010)	2010	2	A1	i.	3,5	F	Longevity Physical and mental health Affiliation and mobility Command over resources (measured across 8 indicators)	Disaster Impact Index (DII)	0-1 (no consequences-maximum consequences)
(Jordan et al., 2010)	2010	1	A1	iv.	7	F	Employment Income Wealth Income Passivity Health Safety Housing Basic Infrastructure Education Social Capital Governance	Cape York Institute for Policy and Leadership 'capability indicators'	Statistical equality between indigenous and non-indigenous and cultural differences cannot be assumed away
(Kleine, 2010)	2010	1	A1	ix.	6	C	Principal: Choice Secondary: Easier communications Increased knowledge Greener environment Increased income Increased mobility More time More voice More autonomy	Choice Framework	Empowering consumers and producers in decision-making through the Choice Framework
(Murphy & Gardoni, 2010)	2010	1	A3	iv.	8	C	N/A	N/A	A just society maximises the variability within groups and minimizes the variability among groups
(Roelen et al., 2010)	2010	1	B1	iii.	3	F	Education poverty Health poverty Shelter poverty Water and sanitation poverty Child work Leisure poverty Social inclusion and protection poverty	Child poverty index for Vietnam:	Poverty cutoff in two dimensions, normalised child poverty gap versus headcount

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Alkire & Foster, 2011a)	2011	1	A1	i.	3, 5	F	United States: Health status Health insurance Income Education Indonesia: Expenditure Health measured as Body Mass Index Years of schooling	Multidimensional Poverty Index (MPI)	Reducing unfreedoms
(Anand et al., 2011)	2011	1	A1	i.	6	C	Health Freedom of Political Expression Freedom of Political Participation Freedom of Religion Freedom of Thought Emotional Capabilities Security Environment and Social Relations Discrimination Outside of Work Discrimination at Work	Nehring-Puppe type index	count of dimensions where above a threshold level
(Arndt & Volkert, 2011)	2011	1	B2	iv.	2	F	Social opportunities Economic Facilities Political Freedoms (measured across 13 indicators)	German Poverty and Wealth Reports	N/A
(Bérenger & Verdier-Chouchane, 2011)	2011	1	A1	vi.	8	F	Health Education Participation (across six indicators)	The Relative Women Disadvantage Index combined with Womens Quality of Life Index -	a high disparity that disadvantages women) to 1 (no disparity)
(Beyazit, 2010)	2011	1	B3	iv.	8	C	N/A	N/A	Moving away from sole focus of monetary gain
(Binder & Broekel, 2011)	2011	1	A1	iv.	1	F	Being happy Being educated Being healthy Being nourished Being well-sheltered Having satisfying social relations	basic functionings':	Conversion efficiency: converting resources into achieved functioning
(Burchardt & Vizard, 2011)	2011	1	A1	iv.	8	F	Life Physical security Health Education and learning Standard of living Productive and valued activities Participation, influence, voice Individual, family and social life Identity, expression, self-respect	The Equality and Human Rights Commission	Human Rights in terms of the achievement of substantive freedoms and opportunities below a minimum threshold

							Legal security		
Authors	**Year**	**Wave**	**Search cat.**	**Group**	**Region**	**Focus**	**Attributes**	**Aggregation**	**Objective/Decision Rule**
(Grunfeld et al., 2011)	2011	2	A1	ix.	3	C	Health Agriculture output Agriculture income Education Road improvements	Capability, Empowerment, and Sustainability virtuous spiral model (CESVS).	The CES virtuous spiral model assumes a minimum set of capabilities to make effective use of ICTS
(Hobson et al., 2011)	2011	2	A1	vii.	2	A	Employment and working time Parental leave Work environment and work culture	Agency and capabilities to achieve a work-life balance	Agency freedom through work-life balance (WLB) through flexibility, rights to reduce hours and parental leave
(Kivunike et al., 2011)	2011	1	B2	ix.	4	F	Social opportunities Economic facilities Political freedoms	Instrumental Freedoms	n/a
(Kleine, 2011)	2011	1	A1	ix.	1,6	C	(same as Klein 2010)	Choice Framework	Empowering consumers and producers in decision-making through the Choice Framework
(Mabsout, 2011)	2011	1	A1	x.	4	C	Education Earnings share Control over earnings Decision-making	health functioning model	Shortfalls in health functioning, in relation to the decision-making role of women
(Maddox & Esposito, 2011)	2011	2	B3	viii.	8	F	N/A	N/A	Minimum threshold of functioning (sufficiency) versus a dichotomy split literate/illiterate
(Matsuyama & Mori, 2011)	2011	1	A3	iv.	8	C	N/A	N/A	Distribution of goods equalizing well-being freedom (achievement), the total amount of goods being fixed.
(Nguefack-Tsague et al., 2011)	2011	2	A2	i.	8	F	Life expectancy Education GDP	HDI	N/A
(Reitinger et al., 2011)	2011	1	A2	vii.	8	C	Life Knowledge and aesthetic experience Work and play Friendship Self-integration Self-expression Transcendence Fairness	Area of Protection (AoP)	N/A
(Renouard, 2011)	2011	1	B1	vii.	8	C	To be integrated into networks To commit oneself to a project within a group To have specific attachments to others To try and value others' objectives considering them as ends	Relational Capability	The maximisation of relational capability for Corporate Social Responsibility

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Smith & Barrett, 2011)	2011	2	B2	viii.	4	F	Freedom from hunger during a school day Ability to meet financial costs of primary education Freedom from need to be self-reliant, economically, emotionally or practically Opportunities to pursue primary education whilst living close to family Opportunity to use language of instruction outside of school or opportunity to be schooled in language used in pupil's community Opportunity to study or read at home	Basic Capabilities in education quality	N/A
(Tikly & Barrett, 2011)	2011	1	B2	viii.	4	C	Inclusion Relevance Democratic	Measure of good quality education	N/A
(Al-Janabi et al., 2012a)	2012	1	A1	x.	1	C	Stability Attachment Autonomy Achievement Enjoyment	ICECAP-A	Health status an influence on capability well-being, not over-riding priority
(Ansari et al., 2012)	2012	1	A3	iii.	8	C	N/A	N/A	Bonding and Bridging social capital to enhance capabilities
(Arndt et al., 2012)	2012	2	B1	iii.	3,4	F	Severe water deprivation Severe sanitation facilities deprivation Severe shelter deprivation Severe education deprivation Severe information deprivation	Welfare indicators	First order dominance
(Callander et al., 2012b)	2012	1	A1	iv.	7	F	Income Health Education	Freedom Poverty Measure	Not in Poverty, At risk of Freedom Poverty, Financial Poverty, Freedom Poverty, Extreme Poverty.
(Callander et al., 2012c)	2012	1	B1	iv.	7	F	(same as Callander et al. 2012b)	Freedom Poverty Measure	(Not in Poverty, At risk of Freedom Poverty, Financial Poverty, Freedom Poverty, Extreme Poverty.)
(Callander et al., 2012a)	2012	2	A1	iv.	7	F	(same as Callander et al. 2012b)	Freedom Poverty Measure	Focus on multiple deprivations rather than income poverty solely
(Hatakka & Lagsten, 2011)	2012	2	A2	ix.	2	F	Educational Personal Professional (measured across 8 indicators)	Capability set from internet use	N/A
(Kelly, 2012)	2012	1	A3	viii.	1	C	N/A	N/A	Capability focus rather than effectiveness

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
(Kleine et al., 2012)	2012	1	B2	ix.	1,3,6	A	(same as Kleine 2010)	Choice Framework	N/A
(Lewis, 2012b)	2012	1	B1	x.	5	C	N/A	Three questionnaires suggested: 1. Residential Environment Assessment Tool (REAT) 2. Design for Health (DFH) 3. System for Observing Play and Leisure Activity (SOPLAY)	Equality of Capabilities
(Lewis, 2012a)	2012	1	A1	x.	5	C	N/A	System for Observing Play & Leisure Activity in Youth (SOPLAY) and System for Observing Play & Recreation in Communities (SOPRAC)	Built Environment Capability
(Maguire et al., 2012)	2012	1	B2	viii.	5	C	Practical reasoning Educational resilience Knowledge and imagination Learning disposition Emotional integrity Social relations Respect, dignity and recognition	Arts Education Pathway Model	None as such (1 to 5 strongly disagree: 5 strongly agree on questions)
(McAllister et al., 2012)	2012	2	B3	x.	1	A	N/A	N/A	Trade-offs between empowerment and health status
(McLean & Walker, 2012)	2012	1	A2	viii.	4	F	Informed vision Affiliation Resilience Social and collective struggle Emotional awareness Integrity Assurance and confidence Knowledge and practical skills	Public-Good Professional Education Index'	Education of professionals orientated to poverty reduction and the public good
(Netten et al., 2012)	2012	2	B1	x.	1	F	Personal cleanliness and comfort Accommodation cleanliness and comfort Food and drink Safety Social participation and involvement Occupation Control over daily life	ASCOT	SC-QALY ('ideal' social care state-death)

Authors	Year	Wave	Search cat.	Group	Region	Focus	Attributes	Aggregation	Objective/Decision Rule
							Dignity		
(Nikiema et al., 2012)	2012	1	A1	x.	4	C	Knowing where to go to seek care Getting permission to go Getting money for treatment Distance to health facility Having to take transportation Not wanting to go alone Concern that there may not be a female health provider	The index of perceived ability to overcome healthcare seeking	Having no problems on the women's perceived ability to access health care
(Norcia et al., 2012)	2012	1	B2	iii.	2	F	A place for shelter and protection	Functioning well-being	N/A
(Notten & Roelen, 2012)	2012	2	B1	i.	1,2	F	Housing Neighbourhood Basic services Financial resources (measured across 13 indicators)	Deprivation indicators	Absolute adjusted headcount with a cumulative threshold of one deprivation.
(Nussbaumer et al., 2012)	2012	2	B1	i.	4	F	Cooking Lighting Services provided by means of household appliances Entertainment/education Communication (measured across six indicators)	Multidimensional Energy Poverty Index (MEPI)	Energy poverty minimisation
(Parks, 2012)	2012	1	B1	iii.	6	F	Health Education Employment Income Housing and living environment Family and community life Transport and communication Participation (measured across 40 indicators)	Human achievement index (HAI)	sufficiency economy
(Peris et al., 2012)	2012	1	A2	ii.	6	A	Who are we? What problems does our community have? Which of these problems are about our natural resources? What is the most important problem we want to act on? What work do we have to do to improve our problem? What do we hope the future will be like after the project?	El Almanario - A Small Grants Programme (SGP) tool	Empowering communities

							Costs of the projects to do the jobs we have planned and the months in which each job will be done.		
Authors	**Year**	**Wave**	**Search cat.**	**Group**	**Region**	**Focus**	**Attributes**	**Aggregation**	**Objective/Decision Rule**
(Perrons, 2012)	2012	1	A1	iv.	1	F	Health Education Income	Regional Development Index (RDI)	Comparable to the HDI
(Van Ootegem & Verhofstadt, 2012)	2012	1	A2	iv.	2	C	Happy life Achievement of dreams and goals Healthy life Education Information and culture Social life Environment Personal integrity	Life Domains	N/A
(Walker, 2012)	2012	2	A3	viii.	8	C	N/A	N/A	Capability approach advanced opposed to human capital approach in reducing education inequality
(Wust & Volkert, 2012)	2012	2	A1	iv.	2	F	Education/Leisure Health Social Participation Income Poverty	Multidimensional childhood capability deprivation	Accounting for more than financial poverty

Author: First Author; Year: Year of Publication; Group: Thematic Grouping of studies within similar subject area; Focus: Aspect of the Capability Approach study focused on;

Thematic Groups: i. General assessment of Human Development, ii. Assessing small scale development projects, iii. Identifying the poor in developing countries, iv. Poverty and well-being assessment in advanced economies, v. Deprivation of disabled people, vi. Assessing gender inequalities, vii. Debating policies, viii. Education, ix. Technology, x. Health

Focus: C, Capabilities; F, Functioning; A, Agency

Region: 1. UK; 2, Other Europe; 3. Asia; 4. Africa; 5. North America; 6. South and Latin America; 7. Oceania; 8. Non-specific

Appendix 4: Studies excluded at the second stage of categorisation

Author	Year	Wave	Search Classification	Group	Primary Focus	Region	Primary Focus of Study	Reason for exclusion
(James, 2006)	2006	1	A4	ix.	F	8	This paper shows how measuring the benefits of the internet in developing nations can be captured through Sen's functionings approach.	This study was an ethnographic study and no measure or decision-rule was presented within this study.
(Hopper, 2007)	2007	1	A4	x.	C	5	This paper looks at the capability framework to understand recovery from schizophrenia.	This study offers no measure or decision rule so is excluded from the review.
(Alkire, 2008)	2008	1	B4	iv.	F	8	This book chapter look at methods as to how dimensions of multidimensional poverty measures should be chosen.	The focus of this review is on already developed measures and methods of aggregation. It is assumed that choosing attributes has already taken place, so this research is excluded from this study.
(Betti & Verma, 2008)	2008	1	B4	iv.	F	2	This study looks at methods of measuring monetary and non-monetary benefits of poverty using fuzzy set theory.	While the paper talks about methods for measuring well-being, it is not directly linked to applying the capability approach and is not considered any further for this review.
(Harreveld & Singh, 2008)	2008	1	B4	viii.	C	7	This study aims to demonstrate that the capability approach provides a useful framework for interpreting the brokering of learning	No measurement of capability is offered in this paper. Whilst the capability approach is recommended, it is not clear from this study what decision rule could be in place from this paper
(Hulme & McKay, 2008)	2008	1	B4	iii.	F	8	This book chapter looks at methods of identifying and measuring chronic poverty measures	This chapter reviews previous applications, but does not offer a unique measure, nor decision rules which such a measure should be based on. Therefore, this research is excluded from this review
(Krishnakumar & Nagar, 2008)	2008	1	B4	iv.	F	8	This study looks at the latent variable models used to weight and aggregate unobservable variables	Whilst related to measures proposing capability indices, this paper does not provide a measure or decision rule, but a review of the statistical techniques used
(Lessmann, 2009)	2009	1	B4	iv.	C	8	This paper compares the theory of the capability approach with the conditions of life approach	No measurement of capability or decision rule is suggested. A critique of the CA is offered where time needs to be adopted closer within the CA

Author	Year	Wave	Search Classification	Group	Primary Focus	Region	Primary Focus of Study	Reason for exclusion
(Cornelius & Wallace, 2010)	2010	2	A4	iv.	C	1	This paper offers an attempt of conceptualising the capability approach for assessing cross-sector urban regeneration projects, with the impact on those of greatest disadvantage as a priority	No measure of capability or decision rule is discussed in this paper, as it offers an early formulation of how the capability approach could be applied for assessing cross-sector regeneration projects
(Lorgelly et al., 2010a)	2010	1	A4	x.	C	8	This paper provides a review of public health outcomes and the approach to capability in health economics	No unique measure or decision rule is put forward within this paper
(Binder & Coad, 2011)	2011	1	A4	iv.	F	1	This study investigates the circulatory problem of Sen's capability approach: individual resources, conversion factors and valuable functionings endogeniety. This is tackled econometrically using reduced-form vector autoregressions. They find that income and being happy can be considered both as a functioning and resource for other functionings. Being well-sheltered and having satisfying social relations are independent of other influences here	This paper looks at the relationship between different functionings, rather than a particular outcome measure or decision rule
(Gries & Naudé, 2011)	2011	1	B4	vii.	C	8	This paper developed an entrepreneurship model based on the capabilities approach	While this study looked at the role of entrepreneurship in developing capability, it did not develop an index as such, nor how such an index could be used to aid decision-making
(Mrcela & Sadar, 2011)	2011	2	A4	vii.	C	2	This study looks at Slovenian parents work-life balance effect on their capabilities. The qualitative study is primarily interested in the gender inequality of roles for the parents	No measure of capabilities to focus on or decision rule is given within this study, so it is excluded from the review
(Vaughan, 2011)	2011	1	A4	ix.	C	7	This study looks at the role of ICT4D within indigenous communities in Australia	The paper does not develop attributes to use or discuss how such a measure could be use to aid decision-making, so it was excluded
(Wang, 2011)	2011	1	A4	viii.	C	3	This paper set out to measure social exclusion through the capability approach and Sen's definitions of deprivation (constitutive deprivation, instrumental deprivation, active	Paper has no measure of capability embedded within study or no decision rule beyond inequality measurement

							deprivation, passive deprivation).	
Author	**Year**	**Wave**	**Search Classification**	**Group**	**Primary Focus**	**Region**	**Primary Focus of Study**	**Reason for exclusion**
(Decancq & Lugo, 2012)	2012	1	A4	i.	F	8	Eight types of weighting structure were explored: frequency; statistical; most-favourable; equal or arbitrary; expert opinion; price based; self-stated; hedonic	While Decancq provides an illuminating review of methods to weight indices, this is not the primary focus of this review, so has been excluded
(Fertig, 2012)	2012	2	A4	viii.	C	4	This paper explores the role of increased capability and agency in converting children's functioning achievement	The paper does not offer a measure of capability or a decision rule as to what should be the objective of such an evaluation.
(Foster et al., 2012)	2012	2	A4	iv.	F	8	This paper examines the robustness of rankings from multidimensional indices such as the Human Development Index (HDI), Index of Economic Freedom and the Environmental Performance Index. While previous research showed redundancy when high correlation existed between indicators, this paper suggests that higher correlations are associated with more robust rankings	The paper is not directly linked to the capability approach, but looks at the method of aggregation of the HDI as well as a number of indices. Therefore, the paper is excluded from the review.
(Trani et al., 2012)	2012	3	B4	viii.	C	3	This paper argues against viewing education as a basic commodity, which is not appropriate within Conflict-Affected Fragile States (CAFS) such as Afghanistan and that an assessment in line with the capability approach would encourage children with disabilities to participate more in education	This paper offers no methods for measuring capabilities or a decision rule of priority, so is excluded from the remainder of the review
(Walby, 2012)	2012	1	B4	iv.	F	1	This presents a critique of Sen's capability approach. It believes that the focus of equality could be overturned by using Sen's choice theory	This study is a critique of the capability approach, which does not inform outcome measurement or decision rules, so has been excluded

Author: First Author; Year: Year of Publication; Group: Thematic grouping of studies within similar subject area; Focus: Aspect of the Capability Approach study focused on;
Thematic Groups: i. General assessment of Human Development, ii. Assessing small scale development projects, iii. Identifying the poor in developing countries, iv. Poverty and well-being assessment in advanced economies, v. Deprivation of disabled people, vi. Assessing gender inequalities, vii. Debating policies, viii. Education, ix. Technology, x. Health
Focus: C, Capabilities; F, Functioning; A, Agency
Region: 1. UK; 2, Other Europe; 3. Asia; 4. Africa; 5. North America; 6. South and Latin America; 7. Oceania; 8. Non-specific

Appendix 5: Western McMasters and Ontario (WOMAC) Osteoarthritis Index

During the past 48 hours...					
	None (0)	Mild (1)	Moderate (2)	Severe (3)	Extreme (4)
PAIN: How much pain have you had...					
1. When walking on a flat surface?					
2. when going up or down stairs?					
3. at night while in bed?					
4. while sitting or lying down?					
5. while standing?					
STIFFNESS : How severe has your stiffness been...					
6. after you first wake up in the morning?					
7. after sitting or lying down or while resting later in the day?					
PHYSICAL FUNCTION: How much difficulty have you had...					
8. when going down the stairs?					
9. when going up the stairs?					
10. when getting up from a sitting position?					
11. while standing?					
12. when bending to the floor?					
13. when walking on a flat surface					
14. getting in or out of a car, or getting on or off a bus?					
15. while going shopping?					
16. when putting on your socks or panty hose or stockings?					
17. when getting out of bed?					
18. when taking off your socks or panty hose or stockings?					
19. while lying in bed?					
20. when getting in or out of the bathtub?					
21. while sitting?					
22. when getting on or off the toilet?					
23. while doing heavy household chores?					
24. while doing light household chores?					

Note: not actual WOMAC Questionnaire layout. Above is truncation of actual questionnaire structure

(for original WOMAC questionnaire, see Bellamy (2004))

Appendix 6: The HAQ-DI (adjusted from Barton et. al., 2004a)

We are interested in learning how your illness affects your ability to function in daily life. Please feel free to add any comments at the end of this form				
PLEASE TICK THE ONE RESPONSE WHICH BEST DESCRIBES YOUR USUAL ABILITIES OVER THE PAST WEEK				
	Without ANY difficulty Score = 0	With SOME difficulty Score = 1	With MUCH difficulty Score = 2	Unable to do Score = 3
1. DRESSING & GROOMING - Are you able to:				
Dress yourself including tying shoelaces and doing buttons?				
Shampoo your hair?				
2. RISING - Are you able to:				
Stand up from an armless straight chair?				
Get in and out of bed?				
3. EATING - Are you able to:				
Cut your meat?				
Lift a cup or glass to your mouth?				
Open a new carton of milk (or soap powder)?				
4. WALKING - Are you able to:				
Walk outdoors on flat ground?				
Climb up five steps?				
5. HYGIENE - Are you able to:				
Wash and dry your entire body?				
Take a bath?				
Get on and off the toilet?				
6.REACH - Are you able to:				
Reach and get a 5 lb object (e.g. a bag of potatoes) from above your head?				
Bend down to pick up clothing from the floor?				
7. GRIP - Are you able to:				
Open car doors?				
Open jars which have been previously opened?				
Turn taps on and off?				
8. ACTIVITIES - Are you able to:				
Run errands and shop?				
Get in and out of a car?				
Do chores such as vacuuming, housework or light gardening?				

How the HAQ-DI is calculated:

Add the highest score for each of the eight categories together, before dividing by eight to get a score between 0-3.
22 polar questions relating to aids, devices & help required are also asked (not presented here).
For each category where such help is required, a minimum score of 2 is recorded for that category.

Appendix 7: Amended HAQ-DI (worked example)

Degree of Difficulty	*Questionnaire*	None	Mild	Moderate	Severe	Extreme	Score
HAQ Score		0	1	2	3	4	0-4
1. Dressing & Grooming:							
Dressing yourself	HAQ-DI			**X**			
Putting on/off shoes	WOMAC			**X**			2
Washing your hair	HAQ-DI			**X**			
2. Rising:							
Rising from sitting	WOMAC				**X**		3
Rising from bed	WOMAC			**X**			
Getting into bed	HAQ-DI			**X**			
3. Eating:							
n/a							-
4. Walking:							
Short distances	SF-36				**X**		3
Climbing stairs	SF-36			**X**			
5. Hygiene							
Washing and drying	HAQ-DI			**X**			
Getting in/out of bath	WOMAC				**X**		3
Getting on/off toilet	WOMAC			**X**			
6. Reach							
Bending to floor	WOMAC				**X**		3
7. Grip							
n/a							-
8. Activities							
Going shopping	WOMAC			**X**			
Getting in/out of car	WOMAC			**X**			2
Light domestic duties	WOMAC			**X**			
Total Score (0-24)							**16**
Average (Total/6)							**2.67**
HAQ_DI (Average*0.75, as HAQ-DI on a 0-3 scale)							**2**

n/a , not available

Appendix 8: Consolidated Health Economic Evaluation Reporting Statement for BRAM12

Items to include when reporting economic evaluations of health interventions Section/item	Item No	Recommendation	Reported on page No/ line No
Title and abstract			
Title	1	Identify the study as an economic evaluation or use more specific terms such as "cost-effectiveness analysis", and describe the interventions compared.	N/A
Abstract	2	Provide a structured summary of objectives, perspective, setting, methods (including study design and inputs), results (including base case and uncertainty analyses), and conclusions.	N/A
Introduction			
Background and objectives	3	Provide an explicit statement of the broader context for the study. Present the study question and its relevance for health policy or practice decisions.	p.222-223
Methods			
Target population and subgroups	4	Describe characteristics of the base case population and subgroups analysed, including why they were chosen.	p. 222
Setting and location	5	State relevant aspects of the system(s) in which the decision(s) need(s) to be made.	p. 223
Study perspective	6	Describe the perspective of the study and relate this to the costs being evaluated.	p. 228
Comparators	7	Describe the interventions or strategies being compared and state why they were chosen.	p. 227
Time horizon	8	State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate.	p. 227
Discount rate	9	Report the choice of discount rate(s) used for costs and outcomes and say why appropriate.	p. 264

Choice of health outcomes	10	Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	p. 264
Measurement of effectiveness	11a	*Single study-based estimates:* Describe fully the design features of the single effectiveness study and why the single study was a sufficient source of clinical effectiveness data.	N/A
11b		*Synthesis-based estimates:* Describe fully the methods used for identification of included studies and synthesis of clinical effectiveness data.	N/A
Measurement and valuation of preference based outcomes	12	If applicable, describe the population and methods used to elicit preferences for outcomes.	p. 235-242
Estimating resources and costs	13a	*Single study-based economic evaluation:* Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	N/A
13b		*Model-based economic evaluation:* Describe approaches and data sources used to estimate resource use associated with model health states. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs.	p. 262 Unchanged from BRAM11 sources
Currency, price date, and conversion	14	Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate.	p. 262
Choice of model	15	Describe and give reasons for the specific type of decision-analytical model used. Providing a figure to show model structure is strongly recommended.	p. 223
Assumptions	16	Describe all structural or other assumptions underpinning the decision-analytical model.	p. 35

Analytical methods	17	Describe all analytical methods supporting the evaluation. This could include methods for dealing with skewed, missing, or censored data; extrapolation methods; methods for pooling data; approaches to validate or make adjustments (such as half cycle corrections) to a model; and methods for handling population heterogeneity and uncertainty.	p. 232-234
Results			
Study parameters	18	Report the values, ranges, references, and, if used, probability distributions for all parameters. Report reasons or sources for distributions used to represent uncertainty where appropriate. Providing a table to show the input values is strongly recommended.	p. 234; 246; 250.
Incremental costs and outcomes	19	For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios.	p. 265-269
Characterising uncertainty	20a	Single study-based economic evaluation: Describe the effects of sampling uncertainty for the estimated incremental cost and incremental effectiveness parameters, together with the impact of methodological assumptions (such as discount rate, study perspective).	N/A
20b		*Model-based economic evaluation:* Describe the effects on the results of uncertainty for all input parameters, and uncertainty related to the structure of the model and assumptions.	p. 270-271; 273; 276; 278.
Characterising heterogeneity	21	If applicable, report differences in costs, outcomes, or cost-effectiveness that can be explained by variations between subgroups of patients with different baseline characteristics or other observed variability in effects that are not reducible by more information.	N/A

Discussion			
Study findings, limitations, generalisability, and current knowledge	22	Summarise key study findings and describe how they support the conclusions reached. Discuss limitations and the generalisability of the findings and how the findings fit with current knowledge.	p. 270-280
Other			
Source of funding	23	Describe how the study was funded and the role of the funder in the identification, design, conduct, and reporting of the analysis. Describe other non-monetary sources of support.	N/A
Conflicts of interest	24	Describe any potential for conflict of interest of study contributors in accordance with journal policy. In the absence of a journal policy, we recommend authors comply with International Committee of Medical Journal Editors recommendations.	N/A

N/A, not applicable. Reporting form from Husereau et al. (2013)

References

Addabbo, T., Lanzi, D., & Picchio, A. (2010). Gender Budgets: A Capability Approach. *Journal of Human Development and Capabilities, 11,* 479-501.

Al-Janabi, H., Flynn, T., & Coast, J. (2011). QALYs and Carers. *Pharmacoeconomics, 29,* 1015-1023.

Al-Janabi, H., Flynn, T., & Coast, J. (2012a). Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Quality of Life Research, 21,* 167-176.

Al-Janabi, H., Keeley, T., Mitchell, P., & Coast, J. (2013). Can capabilities be self-reported? A think aloud study. *Social Science & Medicine, 87,* 116-122.

Al-Janabi, H., Peters, T., Brazier, J., Bryan, S., Flynn, T., Clemens, S. et al. (2012b). An investigation of the construct validity of the ICECAP-A capability measure. *Qual Life Res*.

Alkire, S. (2008). Choosing Dimensions: The Capability Approach and Multidimensional Poverty. In N.Kakwani & J. Silber (Eds.), *The Many Dimensions of Poverty* (pp. 89-119). Basingstoke: Palgrave Macmillan.

Alkire, S. (2009). Concepts and Measures of Agency. In K.Basu & R. Kanbur (Eds.), *Arguments for a Better World: Essays in Honor of Amartya Sen* (pp. 455-474). Oxford, UK: Oxford University Press.

Alkire, S., Qizilbash, M., & Comim, F. (2008). Introduction. In F.Comim, M. Qizilbash, & S. Alkire (Eds.), *The Capability Approach: Concepts, Measures and Applications* (pp. 1-25). Cambridge, UK: Cambridge University Press.

Alkire, S. & Deneulin, S. (2009). A Normative Framework for Development. In S.Deneulin & L. Shahani (Eds.), *An Introduction to the Human Development and Capability Approach* (pp. 3-31). London: Earthscan.

Alkire, S. (2002). *Valuing freedoms: Sen's capability approach and poverty reduction*. Oxford: Oxford University Press.

Alkire, S. (2005). Subjective Quantitative Studies of Human Agency. *Social Indicators Research, 74,* 217-260.

Alkire, S. & Foster, J. (2011a). Counting and multidimensional poverty measurement. *Journal of Public Economics, 95,* 476-487.

Alkire, S. & Foster, J. (2011b). Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality, 9,* 289-314.

Anand, P. & Dolan, P. (2005). Equity, capabilities and health. *Social Science & Medicine, 60,* 219-222.

Anand, P., Hunter, G., Carter, I., Dowding, K., Guala, F., & van Hees, M. (2009). The Development of Capability Indicators. *Journal of Human Development and Capabilities, 10,* 125-152.

Anand, P., Hunter, G., & Smith, R. (2005). Capabilities and Well-Being: Evidence Based on the Sen–Nussbaum Approach to Welfare. *Social Indicators Research, 74,* 9-55.

Anand, P., Krishnakumar, J., & Tran, N. B. (2011). Measuring welfare: Latent variable models for happiness and capabilities in the presence of unobservable heterogeneity. *Journal of Public Economics, 95,* 205-215.

Anand, P. & Santos, C. (2007). Violent crime, gender inequalities and well-being: models based on a survey of individual capabilities and crime rates for England and Wales. *Revue d'économie politique, 117,* 135-160.

Anand, P. & van Hees, M. (2006). Capabilities and achievements: An empirical study. *Journal of Socio-Economics, 35,* 268-284.

Anand, S. & Hanson, K. (1997). Disability-adjusted life years: a critical review. *Journal of Health Economics, 16,* 685-702.

Anderson, E. (2010). Justifying the Capabilities Approach to Justice. In H.Brighouse & I. Robeyns (Eds.), *Measuring Justice. Primary Goods and Capabilities* (Cambridge: Cambridge University Press.

Andronis, L., Malottki, K., Moore, D., & Barton, P. (2009). *Providing Public Health information to prevent skin cancer: Modelling strategies for primary prevention of skin cancer* London: National Institute for Health and Clinical Excellence.

Ansari, S., Munir, K., & Gregg, T. (2012). Impact at the "Bottom of the Pyramid": The Role of Social Capital in Capability Development and Community Empowerment. *Journal of Management Studies, 49,* 813-842.

Arends-Kuenning, M. & Amin, S. (2001). Women's Capabilities and the Right to Education in Bangladesh. *International Journal of Politics, Culture, and Society, 15,* 125-142.

Arndt, C., Distante, R., Hussain, M. A., Østerdal, L. P., Huong, P. L., & Ibraimo, M. (2012). Ordinal Welfare Comparisons with Multiple Discrete Indicators: A First Order Dominance Approach and Application to Child Poverty. *World Development, 40,* 2290-2301.

Arndt, C. & Volkert, J. (2011). The Capability Approach: A Framework for Official German Poverty and Wealth Reports. *Journal of Human Development and Capabilities, 12,* 311-337.

Arnesen, T. & Nord, E. (1999). The value of DALY life: problems with ethics and validity of disability adjusted life years. *British Medical Journal, 319,* 1423-1425.

Arrow, K. J. (1963). Uncertainty and the Welfare Economics of Medical Care. *The American Economic Review, 53,* 941-973.

Asali, M., Reddy, S., & Visaria, S. (2005). Intercountry comparisons of poverty based on a capability approach: a pilot study. New York: Columbia University [On-line].

Atkinson, A. B. (2003). Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches. *Journal of Economic Inequality, 1,* 51-65.

Auguste, P., Barton, P., Hyde, C., & Roberts, T. E. (2011). An economic evaluation of positron emission tomography (PET) and positron emission tomography/computed tomography (PET/CT) for the diagnosis of breast cancer recurrance. *Health Technology Assessment, 15,* 1-54.

Backhouse, R. E. & Medema, S. G. (2009). Retrospectives: On the Definition of Economics. *The Journal of Economic Perspectives, 23,* 221-234.

Baker, R., Bateman, I., Donaldson, C., Jones-Lee, M., Lanscar, E., Loomes, G. et al. (2010). Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technology Assessment, 14,* 1-161.

Balestrino, A. (1996). A note on functioning-poverty in affluent societies. *Notizie di Politeia, 12,* 97-106.

Baltussen, R. & Niessen, L. (2006). Priority setting of health interventions: the need for multi-criteria decision analysis. *Cost Effectiveness and Resource Allocation, 4*.

Barrientos, A. (2010). Protecting Capability, Eradicating Extreme Poverty: Chile Solidario and the Future of Social Protection. *Journal of Human Development and Capabilities, 11,* 579-597.

Barton, G., Sach, T., Jenkinson, C., Avery, A., Doherty, M., & Muir, K. (2008). Do estimates of cost-utility based on the EQ-5D differ from those based on the mapping of utility scores? *Health and Quality of Life Outcomes, 6*.

Barton, P. (2008). *Further cost-effectiveness analysis of sequential analysis of sequential TNF inhibitors for rheumatoid arthritis patients.*

Barton, P., Jobanputra, P., Wilson, J., Bryan, S., & Burls, A. (2004a). The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technology Assessment, 8,* 1-91.

Barton, P. (2011). Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. *Rheumatology, 50,* iv32-iv38.

Barton, P. & Andronis, L. (2009). *Prevention of cardiovascular disease at population level: Modelling strategies for primary prevention of cardiovascular disease* London: National Institute for Health and Clinical Excellence.

Barton, P., Andronis, L., Briggs, A., McPherson, K., & Capewell, S. (2011). Effectiveness and cost effectiveness of cardiovascular disease prevention in whole populations: modelling study. *British Medical Journal, 343*.

Barton, P., Bryan, S., & Robinson, S. (2004b). Modelling in the economic evaluation of health care: selecting the appropriate approach. *Journal of Health Services Research & Policy, 9,* 110-118.

Bellamy, N., Buchanan, W., Goldsmith, C. H., Campbell, J., & Stitt, L. (1988). Valdiation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis in the knee or the hip. *J Rheumatol, 30,* 167-178.

Bellamy, N. (2004). *WOMAC Osteoarthritis Index User Guide* (Rep. No. Verson VII). Brisbane, Australia.

Bérenger, V. & Verdier‒Chouchane, A. (2011). From the Relative Women Disadvantage Index to Women's Quality‒of‒Life. *Journal of Human Development and Capabilities, 12,* 203-233.

Betti, G. & Verma, V. (2008). Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective. *Statistical Methods and Applications, 17,* 225-250.

Beyazit, E. (2010). Evaluating Social Justice in Transport: Lessons to be Learned from the Capability Approach. *Transport Reviews, 31,* 117-134.

Binder, M. & Broekel, T. (2011). Applying a Non‒parametric Efficiency Analysis to Measure Conversion Efficiency in Great Britain. *Journal of Human Development and Capabilities, 12,* 257-281.

Binder, M. & Coad, A. (2011). Disentangling the Circularity in Sen's Capability Approach: An Analysis of the Co-Evolution of Functioning Achievement and Resources. *Soc Indic Res, 103,* 327-355.

Birch, S. & Donaldson, C. (1987). Applications of cost-benefit analysis to health care: Departures from welfare economic theory. *Journal of Health Economics, 6,* 211-225.

Birch, S. & Donaldson, C. (2003). Valuing the benefits and costs of health care programmes: where's the 'extra' in extra-welfarism? *Soc Sci Med, 56,* 1121-1133.

Birch, S. & Gafni, A. (2003). Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results. *Health Policy, 64,* 207-219.

Birch, S. & Gafni, A. (2006). Decision rules in economic evaluation. In A.M.Jones (Ed.), *The Elgar Companion to Health Economics* (pp. 492-502). Cheltenham, UK: Edward Elgar.

Bleichrodt, H. & Quiggin, J. (2013). Capabilities as menus: A non-welfarist basis for QALY evaluation. *Journal of Health Economics, 32,* 128-137.

Boadway, R. & Bruce, N. (1984). *Welfare Economics*. Oxford, UK: Basil Blackwell Publisher Limited.

Boardman, A. E. (2006). *Cost-benefit analysis: concepts and practice*. (3rd ed.) Upper Saddle River, New Jersey: Pearson/Prentice Hall.

Boorse, C. (1975). On the Distinction between Disease and Illness. *Philosophy & Public Affairs, 5,* 49-68.

Boorse, C. (1977). Health as a Theoretical Concept. *Philosophy of Science, 44,* 542-573.

Bosch, J. L. & Hunink, M. G. M. (1996). The Relationship between Descriptive and Valuational Quality-of-life Measures in Patients with Intermittent Claudication. *Medical Decision Making, 16,* 217-225.

Brandolini, A. & D'Alessio, G. (1998). Measuring well-being in the functioning space. Rome: Banca d'Italia [On-line].

Brazier, J., Ratcliffe, J., Salomon, J. A., & Tsuchiya, A. (2007). *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford: Oxford Uiversity Press.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *J Health Econ, 21,* 271-292.

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

Brazier, J. & Tsuchiya, A. (2010). Preference-based condition-specific measures of health: what happens to cross programme comparability? *Health Economics, 19,* 125-129.

Brazier, J., Yang, Y., Tsuchiya, A., & Rowen, D. (2010). A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ, 11,* 215-225.

Brazier, J. E., Kolotkin, R. L., Crosby, R. D., & Williams, G. R. (2004b). Estimating a Preference-Based Single Index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) Instrument from the SF-6D. *Value in Health, 7,* 490-498.

Brennan, A., Chick, S. E., & Davies, R. (2006). A taxonomy of model structures for economic evaluation of health technologies. *Health Economics, 15,* 1295-1310.

Brennan, V. K. & Dixon, S. (2013). Incorporating Process Utility into Quality Adjusted Life Years: A Systematic Review of Empirical Studies. *Pharmacoeconomics, 31,* 677-691.

Briggs, A., Claxton, K., & Sculpher, M. (2006). *Decision Modelling for Health Economic Evaluation*. Oxford, UK: Oxford University Press.

Briggs, A., Sculpher, M., & Buxton, M. (1994). Uncertainty in the economic evaluation of health care technologies: The role of sensitivity analysis. *Health Economics, 3,* 95-104.

Briggs, A. H. & O'Brien, B. J. (2001). The death of cost-minimization analysis? *Health Economics, 10,* 179-184.

Brooks, R. (1996). EuroQol: the current state of play. *Health Policy, 37,* 53-72.

Broome, J. (2006). *Weighing Lives*. Oxford: Oxford University Press.

Broome, J. (2009). Why Economics Needs Ethical Theory. In K.Basu & R. Kanbur (Eds.), *Arguments for a better world: essays in honor of Amartya Sen* (pp. 7-14). Oxford, UK: Oxford University Press.

Brouwer, W. B. F., Culyer, A. J., van Exel, N. J., & Rutten, F. F. H. (2008). Welfarism vs. extra-welfarism. *Journal of Health Economics, 27,* 325-338.

Brouwer, W. B. F., Niessen, L. W., Postma, M. J., & Rutten, F. F. H. (2005). Need for differential discounting of costs and health effects in cost effectiveness analyses. *British Medical Journal, 331,* 446-448.

Bruce, B. & Fries, J. (2004). Longitudinal comparison of the Health Assessment Questionnaire (HAQ) and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). *Arthritis Care & Research, 51,* 730-737.

Bruce, B. & Fries, J. F. (2003). The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *The Journal of Rheumatology, 30,* 167-178.

Burchardt, T. (2009). Agency Goals, Adaptation and Capability Sets. *Journal of Human Development and Capabilities, 10,* 3-19.

Burchardt, T. & Vizard, P. (2011). 'Operationalizing' the Capability Approach as a Basis for Equality and Human Rights Monitoring in Twenty–first–century Britain. *Journal of Human Development and Capabilities, 12,* 91-119.

Button, K. J. (2003). *Recent developments in transport economics*. Cheltenham: Edward Elgar.

Buxton, M. J., Drummond, M. F., Van Hout, B. A., Prince, R. L., Sheldon, T. A., Szucs, T. et al. (1997). Modelling in Economic Evaluation: An Unavoidable Fact of Life. *Health Economics, 6,* 217-227.

CADTH (2006). *Guidelines for the Economic Evaluation of Health Technologies: Canada*. (3rd ed.) Ottawa: Canadian Agency for Drugs and Technologies in Health.

Caiels, J., Forder, J., Malley, J., Netten, A., & Windle, K. (2010). *Measuring the outcomes of low-level services: Final report* (Rep. No. Discussion Paper 2699). University of Kent: Canterbury: PSSRU.

Callander, E. J., Schofield, D., & Shrestha, R. (2012a). Capacity for Freedom – Using a New Poverty Measure to Look at Regional Differences in Living Standards within Australia. *Geographical Research, 50,* 411-420.

Callander, E. J., Schofield, D., & Shrestha, R. (2012b). Towards a holistic understanding of poverty: A new multidimensional measure of poverty for Australia. *Health Sociology Review, 21,* 141-155.

Callander, E., Schofield, D., & Shrestha, R. (2012c). Capacity for Freedom – New Way of Measuring Poverty Amongst Australian Children. *Child Indicators Research, 5,* 179-198.

Caro, J. J., Nord, E., Siebert, U., McGuire, A., McGregor, M., Henry, D. et al. (2010). The efficiency frontier approach to economic evaluation of health-care interventions. *Health Economics, 19,* 1117-1127.

Carr-Hill, R. A. (1989). Assumptions of the QALY procedure. *Social Science & Medicine, 29,* 469-477.

Chen, Y.-F., Jobanputra, P., Barton, P., Jowett, S., Bryan, S., Clark, W. et al. (2006). A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technology Assessment, 10*.

Chen, Y.-F., Jowett, S., Barton, P., Malottki, K., Hyde, C., Gibbs, J. et al. (2009). Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, silaxentan and sildenafil for pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation. *Health Technology Assessment, 13*.

Chiappero-Martinetti, E. (2000). A multidimensional assessment of well-being based on Sen's functioning approach. *Rivista Internazionale di Scienze Sociali, 108,* 207-239.

Chiappero-Martinetti, E. (2003). Unpaid work and household well-being. In A.Picchio (Ed.), *Unpaid Work and the Economy: A Gender Analysis of the Standards of Living* (London: Routledge.

Chiappero-Martinetti, E. & Roche, J. (2009). Operationalization of the capability approach, from theory to practive: a review of techniques and empirical applications. In E.Chiappero-Martinetti (Ed.), *Debating Global Society: Reach and Limits of the Capability Approach* (Milan: Fondazione Giangiacomo Feltinelli.

Chokshi, D. A. & Farley, T. A. (2012). The Cost-Effectiveness of Environmental Approaches to Disease Prevention. *New England Journal of Medicine, 367,* 295-297.

Chuang, L.-H. & Kind, P. (2009). Converting the SF-12 into the EQ-5D: An Empirical Comparison of Methodologies. *Pharmacoeconomics, 27,* 491-505.

Clark, D. & Hulme, D. (2010). Poverty, time and vagueness: integrating the core poverty and chronic poverty frameworks. *Cambridge Journal of Economics, 34,* 347-366.

Clark, D. A. & Qizilbash, M. (2008). Core Poverty, Vagueness and Adaptation: A New Methodology and Some Results for South Africa. *The Journal of Development Studies, 44,* 519-544.

Clark, W., Jobanputra, P., Barton, P., & Burls, A. (2004). The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic evaluation. *Health Technology Assessment, 8*.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S. et al. (2013). *Methods for the Estimation of the NICE Cost Effectiveness Threshold* (Rep. No. CHE Research Paper 81). York: Centre for Health Economics.

Claxton, K., Paulden, M., Gravelle, H., Brouwer, W., & Culyer, A. J. (2011). Discounting and decision making in the economic evaluation of health-care technologies. *Health Economics, 20,* 2-15.

Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M. et al. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics, 14,* 339-347.

Coast, J. (1992). Reprocessing data to form QALYs. *British Medical Journal, 305,* 87-90.

Coast, J. (2004). Is economic evaluation in touch with society's health values? *British Medical Journal, 329,* 1233-1236.

Coast, J. (2009). Maximisation in extra-welfarism: A critique of the current position in health economics. *Social Science & Medicine, 69,* 786-792.

Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J. et al. (2008a). Valuing the ICECAP capability index for older people. *Social Science & Medicine, 67,* 874-882.

Coast, J., Peters, T., Natarajan, L., Sproston, K., & Flynn, T. (2008b). An assessment of the construct validity of the descriptive system for the ICECAP capability measure for older people. *Quality of Life Research, 17,* 967-976.

Coast, J., Smith, R., & Lorgelly, P. (2008c). Should the capability approach be applied in Health Economics? *Health Economics, 17,* 667-670.

Coast, J., Smith, R. D., & Lorgelly, P. (2008d). Welfarism, extra-welfarism and capability: The spread of ideas in health economics. *Social Science & Medicine, 67,* 1190-1198.

Cohen, G. A. (1993). Equaity of What? On Welfare, Goods, and Capabilities. In M.C.Nussbaum & A. Sen (Eds.), *The Quality of Life* (pp. 9-29). Oxford: Oxford University Press.

Comim, F., Qizilbash, M., & Alkire, S. (2008). *The Capability Approach: Concepts, Measures and Applications*. Cambridge, UK: Cambridge University Press.

Connock, M., Juarez-Garcia, A., Jowett, S., Frew, E., Liu, Z., Taylor, R. et al. (2007a). Methadone and buprenorphine for the management of opioid dependence: a systematic review and economic evaluation. *Health Technology Assessment, 11*.

Connock, M., Stevens, C., Fry-Smith, A., Jowett, S., Fitzmaurice, D., Moore, D. et al. (2007b). Clinical effectiveness and cost-effectiveness of different models of managing long-term oral anticoagulation therapy: a systematic review and economic modelling. *Health Technology Assessment, 11*.

Cookson, R., Drummond, M., & Weatherly, H. (2009). Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Economics, Policy and Law, 4,* 231-245.

Cookson, R. (2003). Willingness to pay methods in health care: a sceptical view. *Health Economics, 12,* 891-894.

Cookson, R. (2005a). QALYs and capabilities: a response to Anand. *Health Economics, 14,* 1287-1289.

Cookson, R. (2005b). QALYs and the capability approach. *Health Economics, 14,* 817-829.

Cornelius, N. & Wallace, J. (2010). Cross-Sector Partnerships: City Regeneration and Social Justice. *Journal of Business Ethics, 94,* 71-84.

Couzner, L., Ratcliffe, J., Lester, L., Flynn, T., & Crotty, M. (2013a). Measuring and valuing quality of life for public health research: application of the ICECAP-O capability index in the Australian general population. *International Journal of Public Health, 58,* 367-376.

Couzner, L., Crotty, M., Norman, R., & Ratcliffe, J. (2013b). A Comparison of the EQ-5D-3L and ICECAP-O in an Older Post-Acute Patient Population Relative to the General Population. *Applied Health Economics and Health Policy, 11,* 415-425.

Couzner, L., Ratcliffe, J., & Crotty, M. (2012). The relationship between quality of life, health and care transition: an empirical comparison in an older post-acute population. *Health and Quality of Life Outcomes, 10,* 69.

CRD (2009). *Systematic Reviews: CRD's guidance for undertaking reviews in health care* University of York: Centre for Reviews and Dissemination.

Culyer, A. J. (1989). The Normative Economics of Health Care Finance and Provision. *Oxford Review of Economic Policy, 5,* 34-58.

Culyer, A. J. & Bombard, Y. (2012). An Equity Framework for Health Technology Assessments. *Medical Decision Making, 32,* 428-441.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., & Parkin, D. (2013a). The influence of cost-effectiveness and other factors on NICE decisions. http://news.ohe.org/2013/04/30/nice-decisions-exploring-the-influence-of-cost-effectiveness-and-other-factors/ [On-line].

Dakin, H., Gray, A., & Murray, D. (2013b). Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. *Quality of Life Research, 22,* 683-694.

Dakin, H., Petrou, S., Haggard, M., Benge, S., & Williamson, I. (2010). Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. *Quality of Life Research, 19,* 65-80.

Dakin, H. & Wordsworth, S. (2013). Cost-minimisation analysis versus cost-effectiveness analysis, revisited. *Health Economics, 22,* 22-34.

Daniels, N., Kennedy, B., Kawachi, I., Cohen, J., & Rogers, J. (2000). *Is Inequality Bad for Our Health?* Boston: Beacon Press.

Daniels, N. (2008). *Just health: meeting health needs fairly*. Cambridge: Cambridge University Press.

Davis, J., Liu-Ambrose, T., Richardson, C., & Bryan, S. (2013). A comparison of the ICECAP-O with EQ-5D in a falls prevention clinical setting: are they complements or substitutes? *Quality of Life Research, 22,* 969-977.

de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics, 21,* 145-172.

Dean, H., Bonvin, J. M., Vielle, P., & Farvaque, N. (2005). Developing capabilities and rights in welfare-to-work policies. *European Societies, 7,* 3-26.

Decancq, K. & Lugo, M. a. A. (2012). Weights in Multidimensional Indices of Wellbeing: An Overview. *Econometric Reviews, 32,* 7-34.

Devlin, N. J. & Sussex, J. (2011). *Incorporating multiple criteria in HTA: methods and processes* London: Office of Health Economics.

Di Tommaso, M. L., Shima, I., Strøm, S., & Bettio, F. (2009). As bad as it gets: Well-being deprivation of sexually exploited trafficked women. *European Journal of Political Economy, 25,* 143-162.

Distaso, A. (2007). Well-being and/or quality of life in EU countries through a multidimensional index of sustainability. *Ecological Economics, 64,* 163-180.

Dolan, P. (1997). Modeling Valuations for EuroQol Health States. *Medical Care, 35,* 1095-1108.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1995). *A Social Tariff for EuroQol: Results from a UK General Population Survey* (Rep. No. 138). York: Centre for Health Economics.

Dolan, P. & Kahneman, D. (2008). Interpretations Of Utility And Their Implications For The Valuation Of Health. *The Economic Journal, 118,* 215-234.

Dolan, P., Lee, H., King, D., & Metcalfe, R. (2009). Valuing health directly. *British Medical Journal, 339*.

Dolan, P., Loomes, G., Peasgood, T., & Tsuchiya, A. (2005a). Estimating the Intangible Victim Costs of Violent Crime. *British Journal of Criminology, 45,* 958-976.

Dolan, P., Shaw, R., Tsuchiya, A., & Williams, A. (2005b). QALY maximisation and people's preferences: a methodological review of the literature. *Health Economics, 14,* 197-208.

Dolan, P. & Tsuchiya, A. (2009). The social welfare function and individual responsibility: Some theoretical issues and empirical evidence. *Journal of Health Economics, 28,* 210-220.

Donaldson, C., Mason, H., & Shackley, P. (2012). Contingent valuation in health care. In A.M.Jones (Ed.), *The Elgar Companion to Health Economics* (2nd ed., pp. 425-437). Cheltenham, UK: Edward Elgar Publishing Limited.

Dowswell, G., Ryan, A., Taylor, A., Daley, A., Freemantle, N., Brookes, M. et al. (2012). Designing an intervention to help people with colorectal adenomas reduce their intake of red and processed meat and increase their levels of physical activity: a qualitative study. *BMC Cancer, 12,* 255.

Draper, N. & Smith, H. (1998). *Applied Regression Analysis*. (3rd ed.) New York: Wiley.

Drèze, J. & Sen, A. K. (1989). *Hunger and public action*. Oxford: Clarendon Press.

Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., & Nord, E. (2009). Towards a Consensus on the QALY. *Value in Health, 12,* S31-S35.

Drummond, M., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). *Methods for the Economic Evaluation of Health Care Programmes*. (3rd ed.) Oxford: Oxford University Press.

Drummond, M., Torrance, G., & Mason, J. (1993). Cost-effectiveness league tables: More harm than good? *Social Science & Medicine, 37,* 33-40.

Edlin, R., McCabe, C., Round, J., Wright, J., Claxton, K., Sculpher, M. et al. (2013). Understanding Harris' understanding of CEA: is cost effective resource allocation undone? *Journal of Health Services Research & Policy, 18,* 34-39.

Evans, D. B., Chisholm, D., & Tan-Torres Edejer, T. (2006). Generalized cost-effectiveness analysis: principles and practice. In A.M.Jones (Ed.), *The Elgar comapnion to health economics* (pp. 479-491). Cheltenham, UK: Edward Elgar.

Feeny, D., Furlong, W., Boyle, M., & Torrance, G. W. (1995). Multi-attribute health status classification systems. Health Utilities Index. *Pharmacoeconomics, 7,* 490-502.

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., Depauw, S. et al. (2002). Multiattribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Med Care, 40,* 113-128.

Ferrer, R. L. & Carrasco, A. V. (2010). Capability and Clinical Success. *The Annals of Family Medicine, 8,* 454-460.

Fertig, M. (2012). Educational leadership and the Capabilities Approach: evidence from Ghana. *Cambridge Journal of Education, 42,* 391-408.

Fischer, K. E. (2012). A systematic review of coverage decision-making on health technologies–Evidence from the real world. *Health Policy, 107,* 218-230.

Fletcher, K., Mant, J., McManus, R., Campbell, S., Betts, J., Taylor, C. et al. (2010). Protocol for Past BP: a randomised controlled trial of different blood pressure targets for people with a history of stroke of transient ischaemic attack (TIA) in primary care. *BMC Cardiovascular Disorders, 10,* 37.

Floro, M. S. & Pichetpongsa, A. (2010). Gender, Work Intensity, and Well-Being of Thai Home-Based Workers. *Feminist Economics, 16,* 5-44.

Flynn, T. N., Chan, P., Coast, J., & Peters, T. J. (2011). Assessing Quality of Life among British Older People Using the ICEPOP CAPability (ICECAP-O) Measure. *Applied Health Economics and Health Policy, 9,* 317-329.

Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics, 26,* 171-189.

Forder, J., Towers, A. M., Caiels, J., Beadle-Brown, J., & Netten, A. (2008). *Measuring Outcomes in Social Care: Second Interim Report* (Rep. No. 2542). University of Kent: Canterbury: PSSRU.

Foster, J., Greer, J., & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica, 52,* 761-766.

Foster, J. E., McGillivray, M., & Seth, S. (2012). Composite Indices: Rank Robustness, Statistical Association, and Redundancy. *Econometric Reviews, 32,* 35-56.

Foster, N., Tooth, S., & Hill, J. (2010). The BEEP study: a trial aiming to improve the effectiveness of physiotherapy-led exercise for knee pain in older adults in primary care. http://www.controlled-trials.com/ISRCTN93634563 [On-line].

Fox-Rushby, J. & Hanson, K. (2001). Calculating and presenting disability adjusted life years (DALYs) in cost-effectiveness analysis. *Health Policy and Planning, 16,* 326-331.

Fox-Rushby, J. A. (2002). *Disability Adjusted Life Years (DALYs) for decision-making? An overview of the literature* London: Office of Health Economics.

Frew, E. J. (2010). Benefit assessment for cost-benefit analysis studies in health care using contingent valuation methods. In E.McIntosh, P. M. Clarke, E. J. Frew, & J. J. Louviere (Eds.), *Applied Methods of Cost-Benefit Analysis in Health Care* (pp. 97-118). Oxford, UK: Oxford University Press.

Fries, J. F., Spitz, P., Kraines, R. G., & Holman, H. R. (1980). Measurement of patient outcome in arthritis. *Arthritis & Rheumatism, 23,* 137-145.

Fryback, D., Lawrence, W., Martin, P., Klein, R., & Klein, B. (1997). Predicting Quality of Well-being Scores from the SF-36: Results from the Beaver Dam Health Outcomes Study. *Medical Decision Making, 17,* 1-9.

Fukuda-Parr, S. (2003). The human development paradigm: operationalizing Sen's ideas on capabilities. *Feminist Economics, 9,* 301-317.

Furlong, W. J., Feeny, D. H., Torrance, G. W., & Barr, R. D. (2001). The Health Utilities Index (HUI) system for assessing health-related quality of life in clinical studies. *Annals of Medicine, 33,* 375-384.

Gafni, A. (1994). The standard gamble method: what is being measured and how it is inerpreted. *Health Services Research, 29,* 207-224.

Gafni, A. (2006). Economic Evaluation of Health-care Programmes: Is CEA Better than CBA? *Environmental & Resource Economics, 34,* 407-418.

Gafni, A., Walter, S., & Birch, S. (2013). Uncertainty and the decision maker: assessing and managing the risk of undesirable outcomes. *Health Economics, 22,* 1287-1294.

Gandjour, A. (2010). Theoretical Foundation of Patient v. Population Preferences in Calculating QALYs. *Medical Decision Making, 30,* E57-E63.

Garber, A. M., Weinstein, M. C., Torrance, G. W., & Kamlet, M. S. (1996). Theoretical Foundations of Cost-Effectiveness Analysis. In M.R.Gold, J. E. Siegel, L. B. Russell, & M. C. Weinstein (Eds.), *Cost-Effectiveness in Health and Medicine* (pp. 25-53). Oxford, UK: Oxford University Press.

Gardoni, P. & Murphy, C. (2010). Gauging the societal impacts of natural disasters using a capability approach. *Disasters, 34,* 619-636.

Gasper, D. (2007). What is the capability approach?: Its core, rationale, partners and dangers. *The Journal of Socio-Economics, 36,* 335-359.

Gerard, K. (1993). Setting priorities in the new NHS: can purchasers use cost-utility information? *Health Policy, 25,* 109-125.

Gerard, K. & Mooney, G. (1993). Qaly league tables: Handle with care. *Health Economics, 2,* 59-64.

Ghomrawi, H., Mandl, L., Rutledge, J., Alexiades, M., & Mazumdar, M. (2011). Is there a role for expectation maximization imputation in addressing missing data in research using WOMAC questionnaire? Comparison to the standard mean approach and a tutorial. *BMC Musculoskeletal Disorders, 12,* 109.

Gold, M. R., Siegel, J. E., Russell, L. B., & Weinstein, M. C. (1996). *Cost-Effectiveness in Health and Medicine*. Oxford: Oxford University Press.

Gravelle, H., Brouwer, W., Niessen, L., Postma, M., & Rutten, F. (2007). Discounting in economic evaluations: stepping forward towards optimal decision rules. *Health Economics, 16,* 307-317.

Gravelle, H. & Smith, D. (2001). Discounting for health effects in cost–benefit and cost-effectiveness analysis. *Health Economics, 10,* 587-599.

Gray, A. M., Rivero-Arias, O., & Clarke, P. M. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical Decision Making, 26,* 18-29.

Greco, G., Skordis, J., & Mills, A. (2009). *Methodological challenges for developing an outcome measure based on Sen's Capability approach for assessing women's wellbeing in rural Malawi* Sheffield, UK: Health Economists' Study Group.

Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: Preferences or capabilities? *Social Science & Medicine, 62,* 1891-1901.

Gries, T. & Naudé, W. (2011). Entrepreneurship and human development: A capability approach. *Journal of Public Economics, 95,* 216-224.

Grieve, R., Grishchenko, M., & Cairns, J. (2009). SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility. *European Journal of Health Economics, 10,* 15-23.

Grootendorst, P., Marshall, D., Pericak, D., Bellamy, N., Feeny, D., & Torrance, G. W. (2007). A model to estimate health utilities index mark 3 utility scores from WOMAC index scores in patients with osteoarthritis of the knee. *The Journal of Rheumatology, 34,* 534-542.

Grunfeld, H., Hak, S., & Pin, T. (2011). Understanding benefits realisation of iREACH from a capability approach perspective. *Ethics and Information Technology, 13,* 151-172.

Gudex, C. & Kind, P. (1988). *The QALY toolkit* (Rep. No. 38). York: Centre for Health Economics (CHE).

Hadorn, D. C. (1991a). Setting health care priorities in Oregon: Cost-effectiveness meets the rule of rescue. *JAMA: The Journal of the American Medical Association, 265,* 2218-2225.

Hadorn, D. C. (1991b). The role of public values in setting health care priorities. *Social Science & Medicine, 32,* 773-781.

Haefeli, M., Elfering, A., McIntosh, E., Gray, A., Sukthankar, A., & Boos, N. (2008). A Cost-Benefit Analysis Using Contingent Valuation Techniques: A Feasibility Study in Spinal Surgery. *Value in Health, 11,* 575-588.

Hammitt, J. K. (2002). QALYs Versus WTP. *Risk Analysis, 22,* 985-1001.

Hanley, N. & Barbier, E. B. (2009). *Pricing nature: cost-benefit analysis and environmental policy*. Cheltenham, UK: Edward Elgar.

Harreveld, R. & Singh, M. (2008). Amartya Sen's Capability Approach and the Brokering of Learning Provision for Young Adults. *Vocations and Learning, 1,* 211-226.

Harris, J. (1987). QALYfying the value of life. *Journal of Medical Ethics, 13,* 117-123.

Hartley, R. J., Keen, E. M., Large, J. A., & Tedd, L. A. (1990). Search Strategies. In *Online Searching: Principles and Practice* (pp. 153-173). London: Bowker-Saur.

Hatakka, M. & Lagsten, J. (2011). The capability approach as a tool for development evaluation – analyzing students' use of internet resources. *Information Technology for Development, 18,* 23-41.

Hauber, A. B. (2009). Healthy-years equivalent: wounded but not yet dead. *Expert Review of Pharmacoeconomics & Outcomes Research, 9,* 265-269.

Hawthorne, G., Richardson, J., & Osborne, R. (1999). The Assessment of Quality of Life (AQoL) instrument: a psychometric measure of Health-Related Quality of Life. *Quality of Life Research, 8,* 209-224.

Henderson, C., Knapp, M., Fernández, J.-L., Beecham, J., Hirani, S. P., Cartwright, M. et al. (2013). Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *British Medical Journal, 346*.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M. F., Kind, P., Parkin, D. et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research, 20,* 1727-1736.

Hicks, J. R. (1939). The Foundations of Welfare Economics. *The Economic Journal, 49,* 696-712.

HM Treasury (2003). *The Green Book: Appraisal and Evaluation in Central Government* London: TSO: HM Treasury.

Hobson, B. & Fahlén, S. (2009). Competing Scenarios for European Fathers: Applying Sen's Capabilities and Agency Framework to Work–Family Balance. *The ANNALS of the American Academy of Political and Social Science, 624,* 214-233.

Hobson, B., Fahlén, S., & Takács, J. (2011). Agency and Capabilities to Achieve a Work-Life Balance: A Comparison of Sweden and Hungary. *Social Politics: International Studies in Gender, State & Society, 18,* 168-198.

Holland, R., Matheson, C., Anthony, G., Roberts, K., Priyardarshi, S., MacRae, A. et al. (2012). A pilot randomised controlled trial of brief versus twice weekly versus standard supervised consumption in patients on opiate maintenance treatment. *Drug and Alcohol Review, 31,* 483-491.

Holmes, D. (2013). Report triggers quibbles over QALYs, a staple for health metrics. *Nature Medicine, 19*.

Hopper, K. (2007). Rethinking social recovery in schizophrenia: What a capabilities approach might offer. *Social Science & Medicine, 65,* 868-879.

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health and Quality of Life Outcomes, 1,* 54.

Hulme, D. & McKay, A. (2008). Identifying and Measuring Chronic Poverty: Beyond Monetary Measures? In N.Kakwani & J. Silber (Eds.), *The Many Dimensions of Poverty* (Basingstoke: Palgrave Macmillan.

Human Development Report (2013). *The Rise of the South: Human Progress in a Diverse World* United Nations Development Programme.

Hunt, E. K. (2002). *History of economic thought: a critical perspective*. (2nd ed.) New York: M. E. Sharpe.

Hurley, J. (1998). Welfarism, Extra-Welfarism and Evaluative Economic Analysis. In M.L.Barer, T. E. Getzen, & G. L. Stoddart (Eds.), *Health, health care and health economics: perspectives on distribution* (pp. 373-395). Chicester, England: John Wiley & Sons Ltd.

Hurley, J. (2000). An overview of the normative economics of the health sector. In (1 ed., pp. 55-118). Elsevier.

Hurst, N. P., Kind, P., Ruta, D., Hunter, M., & Stubbings, A. (1997). Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Rheumatology, 36,* 551-559.

Husereau, D., Drummond, M., Petrou, S., Carswell, C., Moher, D., Greenberg, D. et al. (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *BMC Medicine, 11,* 80.

Hutubessy, R., Chisholm, D., & Tan-Torres Edejer, T. (2003). Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost Effectiveness and Resource Allocation, 1,* 8.

Ibrahim, S. & Alkire, S. (2007). Agency and Empowerment: A Proposal for Internationally Comparable Indicators. *Oxford Development Studies, 35,* 379-403.

James, J. (2006). The Internet and poverty in developing countries: Welfare economics versus a functionings-based approach. *Futures, 38,* 337-349.

Jit, M., Barton, P., Chen, Y.-F., Uthman, O., Aveyard, P., & Meads, C. (2009). *School-based interventions to prevent the uptake of smoking among children and young people: cost-effectiveness model* London: National Institute for Health and Clinical Excellence.

Jobanputra, P., Barton, P., Bryan, S., & Burls, A. (2002). The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation. *Health Technology Assessment, 6*.

Johnson, F. R. (2009). Editorial: Moving the QALY Forward or Just Stuck in Traffic? *Value in Health, 12,* S38-S39.

Jordan, K., Bulloch, H., & Buchanan, G. (2010). Statistical equality and cultural difference in Indigenous wellbeing frameworks: A new expression of an enduring debate. *Australian Journal of Social Issues, 45,* 333-362.

Kaambwa, B., Bryan, S., Barton, P., Parker, H., & Martin, G. (2006). Relationship between the EuroQol-5d and Barthel Index - examining the use of proxy outcome measures for older people. In *Health Economists' Study Group (HESG)* York, UK.

Kaambwa, B., Billingham, L., & Bryan, S. (2013). Mapping utility scores from the Barthel index. *European Journal of Health Economics, 14,* 231-241.

Kaldor, N. (1939). Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *The Economic Journal, 49,* 549-552.

Kelly, A. (2012). Sen and the art of educational maintenance: evidencing a capability, as opposed to an effectiveness, approach to schooling. *Cambridge Journal of Education, 42,* 283-296.

Kelly, M., Morgan, A., Ellis, S., Younger, T., Huntley, J., & Swann, C. (2010). Evidence based public health: A review of the experience of the National Institute of Health and Clinical Excellence (NICE) of developing public health guidance in England. *Social Science & Medicine, 71,* 1056-1062.

Kerstenetzky, C. L. & Santos, L. (2009). Poverty as Deprivation of Freedom: The Case of Vidigal Shantytown in Rio de Janeiro. *Journal of Human Development and Capabilities, 10,* 189-211.

Kinghorn, P. B. (2010). *Developing a Capability Approach to Measure and Value Quality of Life: An application to chronic pain.* PhD Thesis University of East Anglia.

Kivunike, F. N., Ekenberg, L., Danielson, M., & Tusubira, F. F. (2011). Perceptions of the role of ICT on quality of life in rural communities in Uganda. *Information Technology for Development, 17,* 61-80.

Klarman, H. E., Francis, J. O., & Rosenthal, G. D. (1968). Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. *Medical Care, 6,* 48-54.

Klasen, S. (2000). Measuring poverty and deprivation in South Africa. *Review of Income and Wealth, 46,* 33-58.

Kleine, D. (2010). ICT4WHAT?−Using the choice framework to operationalise the capability approach to development. *Journal of International Development, 22,* 674-692.

Kleine, D. (2011). The capability approach and the 'medium of choice': steps towards conceptualising information and communication technologies for development. *Ethics and Information Technology, 13,* 119-130.

Kleine, D., Light, A., & Montero, M. J. (2012). Signifiers of the life we value? − considering human development, technologies and Fair Trade from the perspective of the capabilities approach. *Information Technology for Development, 18,* 42-60.

Klugman, J. (2010). *Human Development Report 2010. The Real Wealth of Nations: Pathways to Human Development* New York, NY 10017: United Nations Development Programme.

Klugman, J. (2011). *Human Development Report 2011. Sustainability and Equity: A Better Future for All* New York, NY10017: United Nations Development Programme.

Kreps, D. M. (1979). A Representation Theorem for "Preference for Flexibility". *Econometrica, 47,* 565-577.

Krishnakumar, J. & Ballon, P. (2008). Estimating Basic Capabilities: A Structural Equation Model Applied to Bolivia. *World Development, 36,* 992-1010.

Krishnakumar, J. & Nagar, A. L. (2008). On Exact Statistical Properties of Multidimensional Indices Based on Principal Components, Factor Analysis, MIMIC and Structural Equation Models. *Social Indicators Research, 86,* 481-496.

Kuklys, W. (2005). *Amartya Sen's Capability Approach: Theoretical Insights and Empirical Applications*. Berlin: Springer-Verlag.

Kuklys, W. & Robeyns, I. (2005). Sen's Capability Approach to Welfare Economics. In *Amartya Sen's Capability Approach* (pp. 9-30). Springer-Verlag, Berlin.

Kvien, T. K. (2004). Epidemiology and Burden of Illness of Rheumatoid Arthritis. *Pharmacoeconomics, 22,* S1-S12.

Lavaque-Manty, M. (2001). Food, Functioning and Justice: From Famines to Eating Disorders. *Journal of Political Philosophy, 9,* 150-167.

Le, Q. A. & Doctor, J. N. (2011). Probabilistic Mapping of Descriptive Health Status Responses Onto Health State Utilities Using Bayesian Networks: An Empirical Analysis Converting SF-12 Into EQ-5D Utility Index in a National US Sample. *Medical Care, 49,* 451-460.

Lessmann, O. (2009). Conditions of Life, Functionings and Capability: Similarities, Differences and Complementary Features. *Journal of Human Development and Capabilities, 10,* 279-298.

Lewis, F. (2012a). Auditing Capability and Active Living in the Built Environment. *Journal of Human Development and Capabilities, 13,* 295-315.

Lewis, F. (2012b). Toward a general model of built environment audits. *Planning Theory, 11,* 44-65.

Lewis, J. & Giullari, S. (2005). The adult worker model family, gender equality and care: the search for new policy principles and the possibilities and problems of a capabilities approach. *Economy and Society, 34,* 76-104.

Lin, F. J., Longworth, L., & Pickard, A. (2013). Evaluation of content on EQ-5D as compared to disease-specific utility measures. *Quality of Life Research, 22,* 853-874.

Lipscomb, J., Drummond, M., Fryback, D., Gold, M., & Revicki, D. (2009). Retaining, and Enhancing, the QALY. *Value in Health, 12,* S18-S26.

Liu, L., Strawderman, R. L., Cowen, M. E., & Shih, Y. C. (2010). A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics, 29,* 110-123.

Longworth, L. & Rowen, D. (2013). Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value in Health, 16,* 202-210.

Longworth, L. & Rowen, D. (2011). *NICE DSU Technical Support Document 10: The use of mapping methods to estimate health state utility values*.

Loomes, G. & McKenzie, L. (1989). The use of QALYs in health care decision making. *Social Science & Medicine, 28,* 299-308.

Lord, J., Laking, G., & Fischer, A. (2004). Health care resource allocation: is the threshold rule good enough? *Journal of Health Services Research & Policy, 9,* 237-245.

Lorgelly, P. K., Lawson, K. D., Fenwick, E. A., & Briggs, A. H. (2010a). Outcome measurement in economic evaluations of public health interventions: a role for the capability approach? *International journal of environmental research and public health, 7,* 2274-2289.

Lorgelly, P. K., Lorimer, K., Fenwick, E. A., & Briggs, A. H. (2008). *The Capability Approach: Developing an Instrument for Evaluating Public Health Interventions: Final Report* Glasgow, UK: Glasgow Centre for Population Health.

Lorgelly, P. K., Coast, J., & Smith, R. D. (2010b). Concepts of capability and overlooked applications. *American Journal of Public Health, 100,* 1823-1824.

Lovibond, K., Jowett, S., Barton, P., Caulfield, M., Heneghan, C., Hobbs, F. R. et al. (2011). Cost-effectiveness of options for the diagnosis of high blood pressure in primary care: a modelling study. *The Lancet, 378,* 1219-1230.

Mabsout, R. (2011). Capability and Health Functioning in Ethiopian Households. *Social Indicators Research, 101,* 359-389.

Maddox, B. & Esposito, L. (2011). Sufficiency Re-examined: A Capabilities Perspective on the Assessment of Functional Adult Literacy. *The Journal of Development Studies, 47,* 1315-1331.

Maguire, C., Donovan, C., Mishook, J., Gaillande, G. d., & Garcia, I. (2012). Choosing a life one has reason to value: the role of the arts in fostering capability development in four small urban high schools. *Cambridge Journal of Education, 42,* 367-390.

Makai, P., Brouwer, W., Koopmanschap, M., & Nieboer, A. (2012). Capabilities and quality of life in Dutch psycho-geriatric nursing homes: an exploratory study using a proxy version of the ICECAP-O. *Quality of Life Research, 21,* 801-812.

Makai, P., Koopmanschap, M., Brouwer, W., & Nieboer, A. (2013). A validation of the ICECAP-O in a population of post-hospitalized older people in the Netherlands. *Health and Quality of Life Outcomes, 11*.

Malottki, K., Barton, P., Tsourapas, A., Uthman, O., Liu, Z., Routh, K. et al. (2011). Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. *Health Technology Assessment, 15*.

Mann, J. M., Gostin, L., Gruskin, S., Brennan, T., Lazzarini, Z., & Fineberg, H. V. (1994). Health and Human Rights. *Health and Human Rights, 1,* 6-23.

Mann, R., Brazier, J., & Tsuchiya, A. (2009). A comparison of patient and general population weightings of EQ-5D dimensions. *Health Economics, 18,* 363-372.

Manning, W. G., Morris, C. N., Newhouse, J. P., Orr, L. L., Duan, N., Keeler, E. B. et al. (1981). A Two-Part Model of the Demand for Medical Care: Preliminary Results from the Health Insurance Study. In J.van der Gaag & M. Perlman (Eds.), *Health, Economics, and Health Economics* (pp. 103-123). Amsterdam: North-Holland.

Marra, C., Bansback, N., Anis, A., & Shojania, K. (2011). Introduction to economic modeling for clinical rheumatologists: application to biologic agents in rheumatoid arthritis. *Clinical Rheumatology, 30,* S9-S18.

Matsuyama, J. & Mori, K. (2011). Freedom And Achievement Of Well-Being And Adaptive Dynamics Of Capabilities. *Metroeconomica, 62,* 494-511.

Mauskopf, J., Paul, J., Grant, D., & Stergachis, A. (1998). The Role of Cost–Consequence Analysis in Healthcare Decision–Making. *Pharmacoeconomics, 13,* 277-288.

Maynard, A. (1991). Developing the Health Care Market. *The Economic Journal, 101,* 1277-1286.

McAllister, M., Dunn, G., Payne, K., Davies, L., & Todd, C. (2012). Patient empowerment: The need to consider it as a measurable patient-reported outcome for chronic conditions. *BMC Health Services Research, 12,* 157.

McCabe, C., Claxton, K., & Culyer, A. J. (2008). The NICE Cost-Effectiveness Threshold: What it is and What that Means. *Pharmacoeconomics, 26,* 733-744.

McCabe, C., Claxton, K., & Tsuchiya, A. (2005a). Orphan drugs and the NHS: should we value rarity? *British Medical Journal, 331*.

McCabe, C., Stevens, K., Roberts, J., & Brazier, J. (2005b). Health state values for the HUI 2 descriptive system: results from a UK survey. *Health Economics, 14,* 231-244.

McIntosh, E., Clarke, P. M., Frew, E. J., & Louviere, J. J. (2010). *Applied Methods of Cost-Benefit Analysis in Health Care*. Great Clarendon Street, Oxford: Oxford University Press.

McKie, J. & Richardson, J. (2003). The Rule of Rescue. *Social Science & Medicine, 56,* 2407-2419.

McLean, M. & Walker, M. (2012). The possibilities for university-based public-good professional education: a case-study from South Africa based on the 'capability approach'. *Studies in Higher Education, 37,* 585-601.

Mehrez, A. & Gafni, A. (1989). Quality-adjusted Life Years, Utility Theory, and Healthy-years Equivalents. *Medical Decision Making, 9,* 142-149.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine, 55,* 2149-2158.

Mishan, E. J. (1988). *Cost-benefit analysis: an informal introduction*. (4th ed.) Unwin Hyman, London.

Mitchell, P. M., Roberts, T. E., Barton, P. M., Pollard, B. S., & Coast, J. (2013). Predicting the ICECAP-O Capability Index from the WOMAC Osteoarthritis Index: Is Mapping onto Capability from Condition-Specific Health Status Questionnaires Feasible? *Medical Decision Making, 33,* 547-557.

Mooney, G. (2009). *Challenging Health Economics*. Oxford, UK: Oxford University Press.

Mooney, G. & Ryan, M. (1993). Agency in health care: Getting beyond first principles. *Journal of Health Economics, 12,* 125-135.

Morris, S., Devlin, N., & Parkin, D. (2007). *Economic Analysis in Health Care*. John Wiley & Sons, Ltd.

Morris, S., Devlin, N., Parkin, D., & Spencer, A. (2012). *Economic Analysis in Health Care*. (2nd ed.) Cichester: John Wiley and Sons, Ltd.

Mortimer, D. & Segal, L. (2008). Comparing the Incomparable? A Systematic Review of Competing Techniques for Converting Descriptive Measures of Health Status into QALY-Weights. *Medical Decision Making, 28,* 66-89.

Mrcela, A. K. & Sadar, N. C. (2011). Social Policies Related to Parenthood and Capabilities of Slovenain Parents. *Social Politics, 18,* 199-231.

Murphy, C. & Gardoni, P. (2008). The Acceptability and the Tolerability of Societal Risks: A Capabilities-based Approach. *Science and Engineering Ethics, 14,* 77-92.

Murphy, C. & Gardoni, P. (2010). Assessing capability instead of achieved functionings in risk analysis. *Journal of Risk Research, 13,* 137-147.

Murray, C. & Lopez, A. (1996). *The Global Burden of Disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020*. Cambridge: Harvard University Press.

Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A.-M., Brazier, J. et al. (2012). Outcomes of social care for adults: developing a preference-weighted measure. *Health Technology Assessment, 16,* 1-184.

Netten, A., McDaid, D., Fernàndez, Forder, J., Knapp, M., Matosevic, T. et al. (2005). *Measuring and understanding social services outputs* (Rep. No. 2132/3). University of Kent: Canterbury: PSSRU.

Netten, A., Ryan, M., Smith, P., Skatun, D., Healey, A., Knapp, M. et al. (2002). *The Development of a Measure of Social Care Outcome for Older People* (Rep. No. 1690/2). University of Kent: Canterbury: PSSRU.

Neumann, P. J., Chi-Hui, F., & Cohen, J. T. (2009). 30 Years of Pharmaceutical Cost-Utility Analyses: Growth, Diversity and Methodological Improvement. *Pharmacoeconomics, 27,* 861-572.

Neumann, P. J. & Greenberg, D. (2009). Is The United States Ready For QALYs? *Health Affairs, 28,* 1366-1371.

Nguefack-Tsague, G., Klasen, S., & Zucchini, W. (2011). On Weighting the Components of the Human Development Index: A Statistical Justification. *Journal of Human Development and Capabilities, 12,* 183-202.

NICE (2001). *Guidance for Manufacturers and Sponsors* (Rep. No. 5). London: National Institute for Clinical Excellence.

NICE (2004). *Guide to the methods of technology appraisal* London: National Institute for Clinical Excellence (NICE).

NICE (2008). *Guide to the methods of technology appraisal (2nd ed.)* London: National Institute for Health and Clinical Excellence.

NICE (2009a). Assessing cost effectiveness. In *The guidelines manual* (3rd ed., pp. 81-91). London: National Institute for Health and Clinical Excellence (NICE).

NICE (2009b). *Methods for the development of NICE public health guidance*. (2nd ed.) London, UK: National Institute for Health and Clinical Excellence (NICE).

NICE (2011). *Supporting investment in public health: Reivew of methods for assessing cost effectiveness, cost impact and return on investment* London: National Institute for Health and Clinical Excellence (NICE).

NICE (2013). *Guide to the methods of technology appraisal 2013* London: National Institute for Health and Care Excellence (NICE).

Nikiema, B., Haddad, S., & Potvin, L. (2012). Measuring women's perceived ability to overcome barriers to healthcare seeking in Burkina Faso. *BMC Public Health, 12,* 1-12.

Norcia, M., Rissotto, A., & Noci, E. (2012). Measuring Poverty Through Capabilities: Prelminiary Results of a Resarch in Italy. *OIDA International Journal of Sustainable Development, 4,* 39-46.

Nord, E. (1992). An Alternative To QALYs: The Saved Young Life Equivalent (SAVE). *British Medical Journal, 305,* 875-877.

Nord, E. (1995). The Person-trade-off Approach to Valuing Health Care Programs. *Medical Decision Making, 15,* 201-208.

Nord, E. (1999). *Cost-Value Analysis in Health Care: Making Sense out of QALYs*. Cambridge, UK: Cambridge University Press.

Nord, E. (2012). Measuring concerns for severity: Re-examination of a health scale with purported equal interval properties. *Health Policy, 105,* 312-316.

Nord, E., Enge, A. U., & Gundersen, V. (2010). QALYs: is the value of treatment proportional to the size of the health gain? *Health Economics, 19,* 596-607.

Nord, E., Pinto, J. L., Richardson, J., Menzel, P., & Ubel, P. (1999). Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Economics, 8,* 25-39.

Nordenfelt, L. (1995). *On the Nature of Health: An Action-Theoretic Approach*. (2nd ed.) Dordrecht, The Netherlands: Kluwer Academic Publishers.

Nordenfelt, L. (2007). The concepts of health and illness revisited. *Medicine, Health Care and Philosophy, 10,* 5-10.

Notten, G. & Roelen, K. (2012). A New Tool for Monitoring (Child) Poverty: Measures of Cumulative Deprivation. *Child Indicators Research, 5,* 335-355.

Nussbaum, M. C. (2000). *Women and human development: The capabilities approach*. Cambridge: Cambridge University Press.

Nussbaum, M. C. (2011). *Creating Capabilities: The Human Development Approach*. Harvard University Press.

Nussbaumer, P., Bazilian, M., & Modi, V. (2012). Measuring energy poverty: Focusing on what matters. *Renewable and Sustainable Energy Reviews, 16,* 231-243.

O'Shea, E., Gannon, B., & Kennelly, B. (2008). Eliciting preferences for resource allocation in mental health care in Ireland. *Health Policy, 88,* 359-370.

Olsen, J. A. & Donaldson, C. (1998). Helicopters, hearts and hips: Using willingness to pay to set priorities for public sector health care programmes. *Social Science & Medicine, 46,* 1-12.

Olson, K. (2002). Recognizing Gender, Redistributing Labor. *Social Politics: International Studies in Gender, State & Society, 9,* 380-410.

Palmer, S. & Torgerson, D. J. (1999). Definitions of efficiency. *British Medical Journal, 318,* 1136.

Parks, S. (2012). Divergent pathways of development: a comparative case study of human well-being in two Thai provinces. *Environment and Planning C: Government and Policy, 30,* 891-909.

Payne, K., McAllister, M., & Davies, L. M. (2013). Valuing the economic benefits of complex interventions: when maximising health is not sufficient. *Health Economics, 22,* 258-271.

Peris, J., Farinas, S., Lopez, E., & Boni, A. (2012). Expanding collective agency in rural indigenious communities in Guatemala: a case for El Almanario approach. *International Development Planning Review, 34,* 83-102.

Perrons, D. (2012). Regional performance and inequality: linking economic and social development through a capabilities approach. *Cambridge Journal of Regions, Economy and Society, 5,* 15-29.

Petrou, S. & Gray, A. (2011). Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *British Medical Journal, 342*.

Pharmaceutical Benefits Advisory Committee (2008). *Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee*. (4.3 ed.) Canberra: Australian Government: Department of Health and Ageing.

Phipps, S. (2002). The Well–Being of Young Canadian Children in International Perspective: A Functionings Approach. *Review of Income and Wealth, 48,* 493-515.

Pollard, B., Dixon, D., Dieppe, P., & Johnston, M. (2009). Measuring the ICF components of impairment, activity limitation and participation restriction: an item analysis using classical test theory and item response theory. *Health and Quality of Life Outcomes, 7,* 1-20.

Qizilbash, M. (2002). A note on the measurement of poverty and vulnerability in the South African context. *Journal of International Development, 14,* 757-772.

Qizilbash, M. (2003). Vague language and precise measurement: the case of poverty. *Journal of Economic Methodology, 10,* 41-58.

Rawles, J. (1989). Castigating QALYs. *Journal of Medical Ethics, 15,* 147.

Reinhardt, U. E. (2011). Is There a Market for Ruger's 'Right to Health'? *Journal of Human Development and Capabilities, 12,* 557-563.

Reitinger, C., Dumke, M., Barosevcic, M., & Hillerbrand, R. (2011). A conceptual framework for impact assessment within SLCA. *The International Journal of Life Cycle Assessment, 16,* 380-388.

Renouard, C. (2011). Corporate Social Responsibility, Utilitarianism, and the Capabilities Approach. *Journal of Business Ethics, 98,* 85-97.

Richardson, J., McKie, J., & Bariola, E. (2011). *Review and Critique of Health Related Mulit Attribute Utility Instruments* (Rep. No. 64). Melbourne, Australia: Monash University.

Rivero-Arias, O., Ouellet, M., Gray, A., Wolstenholme, J., Rothwell, P. M., & Luengo-Fernandez, R. (2010). Mapping the Modified Rankin Scale (mRS) Measurement into the Generic EuroQol (EQ-5D) Health Outcome. *Medical Decision Making, 30,* 341-354.

Roberts, J., Brazier, J., & Tsuchiya, A. (2005). Mapping the Overactive Bladder Questionnaire to SF6D indices stage 2: final results.
Ref Type: Unpublished Work

Roberts, T. E., Robinson, S., Barton, P. M., Bryan, S., McCarthy, A., Macleod, J. et al. (2007). Cost effectiveness of home based population screening for Chlamydia trachomatis in the UK: economic evaluation of chlamydia screening studies (ClaSS) project. *British Medical Journal, 335*.

Roberts, T. E., Tsourapas, A., Middleton, L. J., Champaneria, R., Daniels, J. P., Cooper, K. G. et al. (2011). Hysterectomy, endometrial ablation, and levonorgestrel releasing intrauterine system (Mirena) for treatment of heavy menstrual bleeding: cost effectiveness analysis. *British Medical Journal, 342*.

Roberts, T., Henderson, J., Mugford, M., Bricker, L., Neilson, J., & Garcia, J. (2002). Antenatal ultrasound screening for fetal abnormalities: a systematic review of studies of cost and cost effectiveness. *British Journal of Obstetrics and Gynaecology, 109,* 44-56.

Robeyns, I. (2003). Sen's capability approach and gender inequality: selecting relevant capabilities. *Feminist Economics, 9,* 61-92.

Robeyns, I. (2005a). Assessing global poverty and inequality: income, resources, and capabilities. *Metaphilosophy, 36,* 30-49.

Robeyns, I. (2005b). The Capability Approach: a theoretical survey. *Journal of Human Development, 6,* 93-117.

Robeyns, I. (2006). The Capability Approach in Practice. *Journal of Political Philosophy, 14,* 351-376.

Robinson, J. C. (1986). Philosophical Origins of the Economic Valuation of Life. *The Milbank Quarterly, 64,* 133-155.

Roelen, K., Gassmann, F., & Neubourg, C. (2010). Child Poverty in Vietnam: Providing Insights Using a Country-Specific and Multidimensional Model. *Social Indicators Research, 98,* 129-145.

Rosano, A., Mancini, F., & Solipaca, A. (2009). Poverty in People with Disabilities: Indicators from the Capability Approach. *Social Indicators Research, 94,* 75-82.

Round, J. (2012). Is a QALY still a QALY at the end of life? *Journal of Health Economics, 31,* 521-527.

Rouse, P. C., Ntoumanis, N., Duda, J. L., Jolly, K., & Williams, G. C. (2011). In the beginning: Role of autonomy support on the motivation, mental health and intentions of participants entering an exercise referral scheme. *Psychology & Health, 26,* 729-749.

Rowen, D., Brazier, J., Tsuchiya, A., & Alava, M. n. H. (2012). Valuing states from multiple measures on the same visual analogue sale: a feasibility study. *Health Economics, 21,* 715-729.

Ruger, J. P. (1998). *Aristotelian Justice and Health Policy: Capability and Incompletely Theorized Agreements.* Ph.D. dissertation Harvard University.

Ruger, J. P. (2004). Health and social justice. *The Lancet, 364,* 1075-1080.

Ruger, J. P. (2006). Ethics and governance of global health inequalities. *Journal of Epidemiology and Community Health, 60,* 998-1002.

Ruger, J. P. (2010a). *Health and Social Justice*. Oxford: Oxford University Press.

Ruger, J. P. (2010b). Health Capability: Conceptualization and Operationalization. *American Journal of Public Health, 100,* 41-49.

Ruger, J. P. (2011). Shared Health Governance. *The American Journal of Bioethics, 11,* 32-45.

Ruger, J. P. (2012). Fair Enough? Inviting Inequities in State Health Benefits. *New England Journal of Medicine, 366,* 681-683.

Ruggeri Laderchi, C. (1999). The many dimensions of poverty in Peru: theoretical debates and empirical evidence. Queen Elizabeth House Working Paper Series 29.
Ref Type: Unpublished Work

Ruggeri Laderchi, C. (1997). Poverty and its many dimensions: The role of income as an indicator. *Oxford Agrarian Studies, 25,* 345-360.

Ruggeri Laderchi, C., Saith, R., & Stewart, F. (2003). Does it Matter that we do not Agree on the Definition of Poverty? A Comparison of Four Approaches. *Oxford Development Studies, 31,* 243-274.

Sach, T., Smith, R. D., & Whynes, D. K. (2007). A 'League Table' of Contingent Valuation Results for Pharmaceutical Internveions: A Hard Pill to Swallow? *Pharmacoeconomics, 25,* 107-127.

Saith, R. (2011). A Public Health Perspective on the Capability Approach. *Journal of Human Development and Capabilities, 12,* 587-594.

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A. et al. (2012). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet, 380,* 2129-2143.

Schlosser, R. W., Wendt, O., Bhavnani, S., & Nail-Chiwetalu, B. (2006). Use of information−seeking strategies for developing systematic reviews and engaging in evidence−based practice: the application of traditional and comprehensive Pearl Growing. A review. *International Journal of Language & Communication Disorders, 41,* 567-582.

Schokkaert, E. & Van Ootegem, L. (1990). Sen's concept of the living standard applied to the Belgian unemployed. *Recherches Economiques de Louvain, 56,* 429-450.

Sculpher, M. J., Claxton, K., Drummond, M., & McCabe, C. (2006). Whither trial-based economic evaluation for health care decision making? *Health Economics, 15,* 677-687.

Segall, S. (2010). *Health, luck and justice*. Woodstock, Oxfordshire: Princeton University Press.

Sen, A. (1977). Social choice theory: A re-examination. *Econometrica, 45,* 53-89.

Sen, A. (1970). The Impossibility of a Paretian Liberal. *The Journal of Political Economy, 78,* 152-157.

Sen, A. (1976). Poverty: An Ordinal Approach to Measurement. *Econometrica, 44,* 219-231.

Sen, A. (1979). Personal Utilities and Public Judgements: Or What's Wrong With Welfare Economics. *The Economic Journal, 89,* 537-558.

Sen, A. (1980). Equality of What? In *The Tanner Lectures*.

Sen, A. (1985). *Commodities and Capabilities*. Amsterdam, North-Holland.

Sen, A. (1992). *Inequality Reexamined*. (1st ed.) Oxford University Press.

Sen, A. (1993). Capability and Well-Being. In M.Nussbaum & A. Sen (Eds.), *The Quality of Life* (1st ed., pp. 30-53). Oxford University Press.

Sen, A. (1999). *Development as Freedom*. Oxford University Press, Oxford.

Sen, A. (2002). Health: Perception versus observation. *British Medical Journal, 324,* 860-861.

Sen, A. (2004). Capabilities, Lists, and Public Reason: Continuing the Conversation. *Feminist Economics, 10,* 77-80.

Sen, A. (2009). *The Idea of Justice*. Harvard University Press.

Sendi, P., Gafni, A., & Birch, S. (2002). Opportunity costs and uncertainty in the economic evaluation of health care interventions. *Health Economics, 11,* 23-31.

Simon, J., Anand, P., Gray, A., Rugkåsa, J., Yeeles, K., & Burns, T. (2013). Operationalising the capability approach for outcome measurement in mental health research. *Social Science & Medicine, 98,* 187-196.

Smith, A. (1776). *An Inquiry into the Nature and Causes of the Wealth of Nations*. Oxford: Clarendon Press.

Smith, M. & Barrett, A. M. (2011). Capabilities for learning to read: An investigation of social and economic effects for Grade 6 learners in Southern and East Africa. *International Journal of Educational Development, 31,* 23-36.

Smith, R., Lorgelly, P., Al-Janabi, H., Venkatapuram, S., & Coast, J. (2012). The capability approach: an alternative evaluation paradigm for health economics? In A.Jones (Ed.), *The Elgar Companion to Health Economics* (2nd ed., pp. 415-424). Cheltenham, UK: Edward Elgar Publishing, Inc.

Sprangers, M. A. G. & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine, 48,* 1507-1515.

Stafinski, T., McCabe, C. J., & Menon, D. (2010). Funding the Unfundable: Mechanisms for Managing Uncertainty in Decisions on the Introduction of New and Innovative Technologies into Healthcare Systems. *Pharmacoeconomics, 28,* 113-142.

Stafinski, T., Menon, D., & McCabe, C. J. (2011). Health Technology Funding Decision-Making Processes Around the World: The Same, Yet Different. *Pharmacoeconomics, 29,* 475-495.

Stahl, J. (2008). Modelling Methods for Pharmacoeconomics and Health Technology Assessment. *Pharmacoeconomics, 26,* 131-148.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the Commission on the Measurement of Economic Performance and Social Progress* Paris: Commission of the Government of France.

Stinnett, A. A. & Paltiel, A. D. (1996). Mathematical programming for the efficient allocation of health care resources. *Journal of Health Economics, 15,* 641-653.

Stolk, E., van Donselaar, G., Brouwer, B., & Busschbach, J. (2004). Reconciliation of Economic Concerns and Health Policy: Illustration of an Equity Adjustment Procedure Using Proportional Shortfall. *Pharmacoeconomics, 22,* 1097-1107.

Sugden, R. (1993). Review: Welfare, Resources, and Capabilities: A Review of Inequality Reexamined by Amartya Sen. *Journal of Economic Literature, 31,* 1947-1962.

Sugden, R. & Williams, A. (1978). *The Principles of Practical Cost-Benefit Analysis*. Oxford University Press.

Sullivan, P. W. (2011). Are Utilities Bounded at 1.0? Implications for Statistical Analysis and Scale Development. *Medical Decision Making, 31,* 787-789.

Sullivan, P. W. & Ghushchyan, V. (2006). Mapping the EQ-5D Index from the SF-12: US General Population Preferences in a Nationally Representative Sample. *Medical Decision Making, 26,* 401-409.

Sunstein, C. R. (1995). Incompletely Theorized Agreements. *Harvard Law Review, 108,* 1733-1772.

Sutton, E. J. & Coast, J. (2013). Development of a supportive care measure for economic evaluation of end-of-life care using qualitative methods. *Palliative Medicine*.

Tan-Torres Edejer, T., Baltussen, E. R., Adam, T., Hutubessy, R., Acharya, A., Evans, D. B. et al. (2003). *Making choices in health: WHO guide to cost-effectiveness analysis*. Geneva: World Health Organisation.

The World Bank (2010). Extreme Poverty Continues to Fall. http://data.worldbank.org/news/extreme-poverty-rates-continue-to-fall [On-line].

Thokala, P. & Duenas, A. (2012). Multiple Criteria Decision Analysis for Health Technology Assessment. *Value in Health, 15,* 1172-1181.

Tikly, L. & Barrett, A. M. (2011). Social justice, capabilities and the quality of education in low income countries. *International Journal of Educational Development, 31,* 3-14.

Torrance, G. W., Thomas, W. H., & Sackett, D. L. (1972). A Utility Maximisation Model for Evaluation of Health Care Programs. *Health Services Research, 7,* 118-133.

Tosh, J. C., Longworth, L. J., & George, E. (2011). Utility values in National Institute for Health and Clinical Excellence (NICE) Technology Appraisals. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research, 14,* 102-109.

Trani, J. F., Bakhshi, P., & Nandipati, A. (2012). 'Delivering' education; maintaining inequality. The case of children with disabilities in Afghanistan. *Cambridge Journal of Education, 42,* 345-365.

Tsuchiya, A., Brazier, J., McColl, E., & Parkin, D. (2002). Deriving preference-based single indices from non-preference based condition specific instruments: converting AQLQ into EQ-5D indices. *HEDS Discussion Paper*.

Tsuchiya, A. & Dolan, P. (2005). The QALY Model and Individual Preferences for Health States and Health Profiles over Time: A Systematic Review of the Literature. *Medical Decision Making, 25,* 460-467.

Turner, N., Campbell, J., Peters, T., Wiles, N., & Hollinghurst, S. (2013). A comparison of four different approaches to measuring health utility in depressed patients. *Health and Quality of Life Outcomes, 11,* 81.

Ubel, P., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of Life Research, 12,* 599-607.

ul Haq, M. (1990). *Concept and measurement of human development: Human Development Report* New York: United Nations Development Programme.

van de Wetering, E., Stolk, E., van Exel, N., & Brouwer, W. (2013). Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. *The European Journal of Health Economics, 14,* 107-115.

Van Ootegem, L. & Verhofstadt, E. (2012). Using Capabilities as an Alternative Indicator for Well-being. *Social Indicators Research, 106,* 133-152.

Vaughan, D. (2011). The importance of capabilities in the sustainability of information and communications technology programs: the case of remote Indigenous Australian communities. *Ethics and Information Technology, 13,* 131-150.

Venkatapuram, S. (2011). *Health Justice: An Argument from the Capabilities Approach*. Cambridge, UK: Polity Press.

Verkerk, M. A., Busschbach, J. J. V., & Karssing, E. D. (2001). Health-related quality of life research and the capability approach of Amartya Sen. *Quality of Life Research, 10,* 49-55.

von Neumann, J. & Morgenstern, O. (1953). *Theory of games and economic behaviour*. (3rd ed.) Princeton, New Jersey: Princeton University Press.

Wagle, U. R. (2008). Multidimensional poverty: An alternative measurement approach for the United States? *Social Science Research, 37,* 559-580.

Wagle, U. (2009). Capability Deprivation and Income Poverty in the United States, 1994 and 2004: Measurement Outcomes and Demographic Profiles. *Social Indicators Research, 94,* 509-533.

Walby, S. (2012). Sen and the Measurement of Justice and Capabilities: A Problem in Theory and Practice. *Theory, Culture & Society, 29,* 99-118.

Walker, M. (2006). Towards a capability-based theory of social justice for education policy-making. *Journal of Education Policy, 21,* 163-185.

Walker, M. (2008). A human capabilities framework for evaluating student learning. *Teaching in Higher Education, 13,* 477-487.

Walker, M. (2012). A capital or capabilities education narrative in a world of staggering inequalities? *International Journal of Educational Development, 32,* 384-393.

Wang, L. (2011). Social exclusion and inequality in higher education in China: A capability perspective. *International Journal of Educational Development, 31,* 277-286.

Ware, J. E., Kolinski, M., & Keller, S. (1995). *How to Score the SF-12 Physical and Mental Health Summaries: A User's Manual*. New England Medical Centre, Boston, MA: Boston: The Health Institute.

Ware, J. E., Jr. & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection. *Medical Care, 30,* 473-483.

Weinstein, M. C. & Stason, W. B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *The New England journal of medicine, 296,* 716-721.

Weinstein, M. C. (1990). Principles of Cost-Effective Resource Allocation in Health Care organizations. *International Journal of Technology Assessment in Health Care, 6,* 93-103.

Weinstein, M. C., Torrance, G., & McGuire, A. (2009). QALYs: The Basics. *Value in Health, 12,* S5-S9.

Whitehead, M. (1990). *The concepts and principles of equity and health* Copenhagen: World Health Organization.

Whitehurst, D. G. T. & Bryan, S. (2011). Another Study Showing that Two Preference-Based Measures of Health-Related Quality of Life (EQ-5D and SF-6D) are Not Interchangeable. But Why Should We Expect Them to be? *Value in Health, 14,* 531-538.

Williams, A. (1985). Economics of Coronary Artery Bypass Grafting. *British Medical Journal, 291,* 326-329.

Williams, A. (1997). Intergenerational Equity: An Exploration of the 'Fair Innings' Argument. *Health Economics, 6,* 117-132.

Winkelmann, R. (2012). Copula bivariate probit models: with an application to medical expenditures. *Health Economics, 21,* 1444-1455.

Wolfe, F. (2001). Which HAQ is best? A comparison of the HAQ, MHAQ and RA-HAQ, a difficult 8 item HAQ (DHAQ), and a rescored 20 item HAQ (HAQ20): analyses in 2,491 rheumatoid arthritis patients following leflunomide initiation. *The Journal of Rheumatology, 28,* 982-989.

Wust, K. & Volkert, J. (2012). Childhood and Capability Deprivation in Germany: A Quantitative Analysis Using German Socio-Economic Panel Data. *Social Indicators Research, 106,* 439-469.

Xie, F., Pullenayegum, E. M., Li, S. C., Hopkins, R., Thumboo, J., & Lo, N. N. (2010). Use of a Disease-Specific Instrument in Economic Evaluations: Mapping WOMAC onto the EQ-5D Utility Index. *Value in Health, 13,* 873-878.

Young, M. (2009). Basic Capabilities, Basic Learning Outcomes and Thresholds of Learning. *Journal of Human Development and Capabilities, 10,* 259-277.

Zaidi, A. & Burchardt, T. (2005). Comparing incomes when needs differ: equivalization for the extra costs of disability in the U.K. *Review of Income and Wealth, 51,* 89-114.