Supporting the local decision-making process in England in the context of integrated care





Pamela Gongora Salazar

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy (DPhil)

Hilary 2024

Table of Contents

Acknowledgements 6
Abstract 9
Publications and presentations related to this thesis 10
Declaration of Originality 11
List of Figures 12
List of Tables 13
List of Abbreviations 14
1. Chapter 1. Introduction 18
1.1. Rationale for the thesis 18
1.2. Thesis aim and objectives 25
1.3. Structure of the thesis 26
1.4. Contribution to the literature 28
1.5. Ethics approval 30
1.6. Conclusion 30
2. Chapter 2. Changes and trends in the landscape of commissioning in England in the context of integrated care 31
2.1. Introduction 31
2.2. Methods 33
2.3. Evolution of integrated care in England 35
2.4. Current local commissioning context: The transition towards Integrated Care Systems 46
2.4.1. Understanding of integrated care initiatives 49
2.4.2. Investment decision process 50
2.4.3. Prioritisation of interventions 52
2.4.4. Assessment of interventions 54
2.4.5. Challenges in the commissioning process 55
2.5. Discussion and Conclusion 58
3. Chapter 3. The use of Multi-criteria Decision Analysis to inform decisions in Healthcare: An Updated Systematic Literature Review 65
3.1. Introduction 65
3.2. Methods 67
3.3. Results 71
3.3.1. Study characteristics 72
3.3.2. Decision-context 73
3.3.3. Criteria 77
3.3.4. MCDA approaches 81
3.3.5. Performance scoring 84

3.3.6. Criteria weighting 86
3.3.7. Uncertainty analyses 87
3.4. Discussion and Conclusion 89
4. Chapter 4. A decision-support framework to monitor and evaluate new models of care in England 95
4.1. Introduction 95
4.2. The underpinnings of the decision-support framework 96
4.3. The evaluation component of the framework 99
4.3.1. Decision-problem and alternatives 101
4.3.2. Assessment criteria 104
4.3.3. Measuring performance & scoring alternatives 107
4.3.4. Weighting the assessment criteria 110
4.3.5. The MCDA total score 111
4.3.6. Sensitivity analyses 115
4.4. The monitoring component of the framework 117
4.5. The framework as part of the commissioning cycle 120
4.6. Discussion & Conclusions 122
5. Chapter 5. Using stated preferences choice experiments to elicit MCDA criteria weights in the context of integrated care in England 127
5.1. Introduction 127
5.2. Decision model and descriptive framework 131
5.2.1. Identification and definition of attributes and levels 131
5.2.2. Elicitation task and format 134
5.2.3. Experimental design 135
5.2.4. Survey instrument 140
5.2.5. Data collection, recruitment strategy and sampling 140
5.2.6. Data quality 142
5.2.7. Data analysis 145
5.3. Results 151
5.3.1. Respondents' characteristics 151
5.3.2. Multinomial logit models (MNL) & observed heterogeneity 153
5.3.3. Mixed logit models (MMNL) & unobserved heterogeneity 155
5.4. Discussion and Conclusion 162
6. Chapter 6. The economic evaluation of new models of care in England using multi-criteria decision analysis: A case study on mental health integrated care services 170
6.1. Introduction 170
6.2. Methods 171

6.2.1. Alternatives to be evaluated 172
6.2.2. Assessing performance of EIP services 173
6.2.2.1. Dataset 174
6.2.2.2. Operationalization of the assessment criteria 176
6.2.2.3. Propensity Score Matching and Statistical Analyses 179
6.2.2.4. Standardisation of performance scores 184
6.2.3. MCDA total score 185
6.2.4. Sensitivity analysis 185
6.3. Results 186
6.3.1. Selection of Propensity Score Matching technique 186
6.3.2. Performance scores and MCDA total scores 189
6.3.3. Sensitivity analysis 192
6.4. Discussion & Conclusion 195
7. Chapter 7. Discussion and conclusion 201
7.1. Introduction 201
7.2. Summary of key findings 201
7.3. Challenges and limitations of this research 207
7.4. Implications of this research 212
7.5. Steps forward 217
7.6. Concluding remarks 220
Appendices 221
Appendices – Chapter 2 221
Appendix 2.A. Semi-structured interview guide 221
Appendix 2.B. List of interviewees 221
Appendix 2.C. Local structures under the new commissioning context 222
Appendix 2.D. Supplementary quotes and information provided by interviewees 223
Appendix 2.E. The commissioning decision-context before the abolishment of CCGs 227
Appendices – Chapter 3 228
Appendix 3.A. Results of searches 228
Appendix 3.B. Definition of the decision-contexts 232
Appendix 3.C. Studies included in the review 234
Appendix 3.D. The 10 most used criteria by studies informing prioritisation of interventions in high-income countries 268
Appendices – Chapter 4 270
Appendix 4.A. Selection of assessment and monitoring criteria 270
Appendix 4.B. Graphical representation of assessment and monitoring criteria 279

Appendix 4.C. Example of key performance indicators to operationalise the suggested monitoring criteria in the context of Early Intervention in Psychosis (EIP) services 279
Appendices – Chapter 5 282
Appendix 5.A. Rationale and definition of attributes 282
Appendix 5.B. Attribute levels and rationale 284
Appendix 5.C. BRC event Pre test - DCE task 288
Appendix 5.D. Pilot study 288
Appendix 5.E. Experimental design (full launch) 297
Appendix 5.F. Survey instrument 326
Appendix 5.G. Census balanced quotas 341
Appendix 5.H. Multinomial logit estimates 342
Appendix 5.J.Correlated mixed logit models & comparison of RI scores 348
Appendices – Chapter 6 350
Appendix 6.A. Dataset structure & transformation 350
Appendix 6.B. Selection of performance indicators 352
Appendix 6.C. Performance indicators 355
Appendix 6.D. Selection of Propensity Score Matching (PSM) technique 357
Appendix 6.E. GMLs performance 359
References 364

Acknowledgements

While this thesis is the result of my hard work and dedication, it equally reflects the invaluable support and contributions of many. To all those who have supported me throughout this journey, I am sincerely grateful.

First and foremost, I would like to thank my supervisors, Professor Apostolos Tsiachristas, Professor Ray Fitzpatrick and Professor Rafael Perera, for their guidance and constant support throughout my DPhil. Apo - I owe you a million thanks for the time and effort you invested in guiding me through every stage. Our weekly meetings, always conducted without haste, were all enjoyable and inspiring. Your support for my career and wellbeing, even during periods of personal challenge for you, has been invaluable. Your exceptional supervision and holistic approach to mentorship have been crucial in this journey; I consider myself immensely fortunate to have had you as my primary supervisor. Ray - I am sincerely grateful for your insightful comments and the encouragement to engage in critical reflection throughout my research. Your inputs constantly inspired deeper thought. Rafael - thank you for your support and understanding over these years. Your involvement has also been an integral part of my DPhil journey.

I am also deeply grateful to Professor Oliver Rivero-Arias for his significant contributions and guidance when conducting the systematic literature review and the choice experiment included in this thesis. His support was invaluable.

To my collaborators, Dr Margaret Glogowska, Dr Patrick Fahr, and Stephen Rocks: Many thanks for your contributions. Every meeting was both productive and truly enjoyable. A special note of gratitude to Patrick - your support, not only as a colleague but as a friend, has been priceless.

Similarly, I would like to express sincere gratitude to Dr. John Buckell for his friendship and generosity in every discussion about choice modelling. Additionally, I extend my thanks to Professor Stavros for his invaluable feedback during the confirmation of my thesis. I am also grateful to Edward Penington for his support in understanding the Clinical Record Interactive Search (CRIS)

datasets, and to Professor Marcel Jonker from Erasmus University Rotterdam for his insightful suggestions in the design of the choice experiments.

I would also like to extend my gratitude to members of the Health Economics Research Centre (HERC), the Nuffield Department of Population Health (NDPH), Green Templeton College and the NIHR Oxford and Thames Valley Applied Research Collaboration (OTV ARC). A special mention goes to Marie-Alice Frappat and Barbara Kitchener for their endless kindness and invaluable administrative support. I am also grateful to NDPH and OTV ARC for their generous financial support, which was instrumental in the completion of my DPhil. Special thanks to the Patient and Public Involvement (PPI) group from OTV ARC, particularly to Mary Zacaroli, for their time and feedback when designing the task for the choice experiments.

I cannot forget to thank the local commissioners from South East England and experts who generously gave their time to help me understand the local commissioning process in England and the datasets to assess early intervention in psychosis services. Special thanks to Lulu Kane and Adam Pill from Oxford Health NHS Foundation Trust for their support in accessing the datasets.

I would also like to thank those who attended the 21st anniversary event of the European Health Policy Group as well as the International Health Economics Association (IHEA) 2023 Congress. The discussions were very enriching for my work.

I would like to express my gratitude to Professor Gary Ford, Chief Executive Officer at Health Innovation Oxford and Thames Valley, and Professor Maureen Rutten-van Mölken, Head of the Health Technology Assessment department at Erasmus University Rotterdam, for examining my thesis. Their feedback have allowed me to improve the quality of this work.

The journey of a DPhil is as much about moral support as it is about academic rigour, and in this regard, the role of family and friends has been fundamental. To start, I would like to thank my parents, Martha and Santiago, for their unconditional love and support throughout my life. *Papitos* - You have been so supportive of my dreams and have always believed in me. Each achievement of

mine is as much an achievement of yours. Immense gratitude to my husband, Kevin, for his love, patience and constant support. Kev - meeting you during my DPhil and joining our lives has been the most beautiful surprise of this journey. To my sister, Paula, my nephew, Ollie, and my brother-in-law, Cam, thank you for your ever-present support. I also have to thank you, *Abuela* – you left in the middle of this journey, but I know that even now, from heaven, you are with me.

I also owe my deepest gratitude to my friends in Oxford. James – the best office mate I could have asked for; I value our friendship greatly. Aurelia - we began this journey together, and it has been a joy to see our friendship thrive across borders. To Laura, Martu, Fiorella, Ana, Jack, Koen and Jihyung – thank you for the joyous times we shared and for always cheering me up.

Finally, I would like to acknowledge those who supported me from abroad throughout this journey. Ursula - thank you for being an inspiration and for your constant support. Paul – I will always be grateful for your generosity, both as a colleague and friend. Mafe – thank you for being a constant source of encouragement, regardless of time zone or distance. Pablo – thank you for your understanding and support as I navigated both this DPhil journey and the challenges of embarking on a new job.

This journey, though challenging, has been immensely rewarding, fostering both personal and academic growth. I am grateful for this dream-turned-reality and thank God for this opportunity.

This thesis is for you, Santiago & Martha. *Los amo*.

Abstract

Over the last decade, England has emphasised the implementation of new care models and restructured the local commissioning architecture aiming to strengthen its integration agenda. Although this overhaul has increased the need for evidence-based decisions, research suggests that local commissioners may not have enough evidence to make well-informed investment decisions. Evaluating these models systematically and empirically can be challenging, and traditional economic evaluation methods may not be suitable.

This thesis develops a relevant, practical and theoretically sound decision-support framework, based on multi-criteria decision analysis (MCDA) for evaluating and monitoring new models of care at local level in England. The framework's structure was informed by semi-structured interviews with local stakeholders and a systematic literature review on the application of MCDA in healthcare. Public preferences were integrated through a discrete choice experiment (DCE) with the general public. The framework's feasibility was tested on two early intervention in psychosis (EIP) services using electronic health record data.

The interviews highlighted the importance of rational and transparent local commissioning processes. Through the interviews and literature review, a comprehensive set of value-based elements was identified that can be objectively applied to assess new models of care. Results of the DCE suggest that the public assigns high importance to elements of value that are not captured in conventional value metrics. The EIP evaluation demonstrated the framework's feasibility in mental health, a key priority area for NHS England. The DCE and service evaluation emphasised the importance of reporting the MCDA total score with and without the cost criterion.

The proposed decision-support framework has the potential to assist decision-makers in strengthening the local commissioning process for integrated care. Although further exploration is required to determine its broader applicability and effectiveness, the proposed framework establishes a basis to enhance local population health management in England, underpinned by real-world evidence.

Publications and presentations related to this thesis

Publications

- Gongora-Salazar P, Rocks S, Fahr P, Rivero-Arias O, Tsiachristas A (2023). The Use of Multicriteria Decision Analysis to Support Decision Making in Healthcare: An Updated Systematic Literature Review. Value Health, 26(5):780-790.

- Tsiachristas, A., Vrangbæk, K., Gongora-Salazar, P., & Kristensen, S. (2023). Integrated care in a Beveridge system: Experiences from England and Denmark. Health Economics, Policy and Law, 18(4), 345-361.

- Gongora-Salazar, P., Glogowska, M., Fitzpatrick, R., Perera, R., & Tsiachristas, A. (2022). Commissioning [Integrated] Care in England: An Analysis of the Current Decision Context. International Journal of Integrated Care, 22(4), 3.

Presentations & Posters

- Gongora-Salazar, P "Unravelling the Elements of Value of Health Care and Assessing Their Importance Using Evidence from Two Discrete-Choice Experiments in England" July 10,2022. 15th IHEA World Congress on Health Economics (July 8-12, 2023), Cape Town. [long oral presentation]

- Gongora-Salazar, P. "The Use of Multi-Criteria Decision Analysis (MCDA) to Support Decision Making in Healthcare: an Updated Systematic Literature Review". November 7, 2022. ISPOR Europe 2022, Vienna, Austria [poster]

- Gongora-Salazar, P. "Which integrated care programmes should be funded?. A framework to support the local commissioning of integrated care in England" November 28, 2022. NIHR ARC OxTV Showcase event 2022, Oxford, UK [poster]

- Gongora-Salazar, P and Espinoza, M. "MCDA to design and adjust explicit benefit plans. An introduction with practical examples". May 15, 2022. Inter-American Development Bank. CRITERIA Network [webinar].

Declaration of Originality

I, Pamela Gongora Salazar, declare that this thesis, in its entirety, has not been previously submitted for any academic degree, nor is it concurrently under consideration for any other qualification. I affirm that the work presented in this thesis represents my original research, except where explicitly attributed, and that any assistance received during its preparation, as well as all sources consulted, have been duly acknowledged within this thesis.

List of Figures

Figure 2.1. Evolution of integrated care in England 36

Figure 2.2. The current commissioning decision-context after the abolishment of CCGs 48

Figure 2.3. Bottom-up investment decision process within a CCG or Integrated Care Boards 52

Figure 3.1. PRISMA flowchart 72

Figure 3.2. Publication trend of MCDA studies in healthcare by decision context 73

Figure 3.3. Criteria used by decision context 79

Figure 3.4. MCDA methods and techniques applied 82

Figure 3.5. Stakeholders involved in the scoring and weighting 86

Figure 3.6. Type of uncertainty and methods used to deal with it across decision context 88

Figure 4.1. Principles of the decision-support framework 99

Figure 4.2. The alternatives: Type of healthcare interventions to commission 102

Figure 4.3. Potential comparators for an economic evaluation 104

Figure 4.4. Identification of assessment criteria 107

Figure 4.5. A decision-support framework for the local commissioning of new models of care 121

Figure 5.1. A framework for stated-preference studies 131

Figure 5.2. Example of choice task 135

Figure 5.3. Illustration of generating QALY gains from two attributes 150

Figure 5.4. Comparison of relative importance (RI) scores 157

Figure 6.1. Episodes of care for a subsample of patients with episodes longer than 100 days and first 50 contacts with crisis teams. 176

Figure 6.2. Standardised performance scores, with and without cost 192

Figure 6.3. PSM results of MCDA overall scores 194

Figure 6.4. Conditional multi attribute acceptability curve (CMAC) showing the probability of Oxf EIP to be accepted as the preferred EIP service over a range of available budgets. 194

List of Tables

Table 2.1. Ideal criteria for prioritization of interventions 54

Table 3.1. Search terms used 68

Table 3.2. Decision-contexts that MCDA studies have been applied in 74

Table 3.3. Decision-makers, type of decisions and types of alternatives under evaluation 75

Table 3.4. MCDA methods and techniques applied by decision context 83

Table 5.1. Attributes & levels 132

Table 5.2. Characteristics of experimental designs 139

Table 5.3. Socio-demographic characteristics of the respondents 151

Table 5.4. Joint estimation (No AddTax & AddTax) of multinomail logit models: Main effects & observed heterogeneity 153

Table 5.5 Parameter coefficients from the mixed logit model (MMNL) 156

Table 5.6. MMNL DCE – Add Tax: MRS estimates 159

Table 5.7. MMNL DCE – NoAdd Tax: MRS estimates 160

Table 5.8. Estimates of interaction terms between attributes based on MNL models 161

Table 5.9. Comparison of Relative Importance (RI) scores across estimates 162

Table 6.1. Operationalisation of the assessment criteria 178

Table 6.2. Summary results of performance of PSM and Stabilized Inverse Probability Weighting 187

Table 6.3. Characteristics before and after PSM 188

Table 6.4. Unstandardized performance scores after propensity score matching 190

Table 6.5. Standardised performance scores, weights and MCDA scores 191

Table 6.6. Results of the probabilistic sensitivity analysis (PSM) 193

List of Abbreviations

AddTax	Additional budget needed
AIC	Akaike Information Criterion
AHP	Analytic hierarchy process
AMHO	Adult Mental Health Organisation
BCF	Better Care Fund
BIC	Bayesian Information Criterion
BOB	Buckinghamshire, Oxfordshire and Berkshire West
BRA	Benefit risk assessment
Bucks	Buckinghamshire
CBA	Cost-benefit analysis
CCA	Cost-consequence analysis
CCG	Clinical Commissioning Group
CEA	Cost-effectiveness analysis
CIA	Conditional Independence Assumption
CMAC	Conditional Multi-attribute Acceptability Curve
CMHTs	Community mental health teams
CRIS	Clinical Interactive Search
CSU	Commissioning support unit
CUA	Cost-utility analysis
CVR	Cost-per-value ratio
DCE	Discrete choice experiment
EIP	Early intervention in psychosis
EUnetHTA	European Network for Health Technology Assessment
Equ	Target population

Exp	Patient experience
FYF	Five Year Forward View
GLM	Generalised linear models
GPs	General practitioners
HTA	Health technology assessment
HWB	Health and Wellbeing Board
HoNOS	Health of the Nation Outcome Scales
ICS	Integrated Care System
IMD	Index of multiple deprivation
INAHTA	International Network of Agencies for Health Technology Assessment
IPW	Inverse probability weighting
ISPOR	Professional Society for Health Economics and Outcomes Research
JSNA	Joint Strategic Needs Assessment
KPIs	Key performance indicators
LSOA	Lower layer super output area
MACBETH	Measuring attractiveness through a categorical-based evaluation technique
MRS	Marginal rate of substitution
MAUT	Multi-Attribute Utility Theory
MAVT	Multi-attribute value theory
MCDA	Multi-criteria decision analysis
MLHS	Modified Latin Hypercube Sampling
MMNL	Mixed multinomial logit
MNL	Multinomial logit
MRS	Marginal rate of substitution
NHS	National Health Service
NICE	National Institute for Health and Care Excellence

NIHR	National Institute for Health Research
OECD	Organisation for Economic Co-operation and Development
OHFT	Oxford Health NHS Foundation Trust
OTV ARC	Oxford and Thames Valley Applied Research Collaboration
Oxf	Oxfordshire
PAPRIKA	Potentially all pairwise rankings of all possible alternatives
PCG	Primary care groups
PCTs	Primary care trusts
PPI	Public and patient involvement
PREM	Patient reported experience measure
PSA	Probabilistic sensitivity analysis
PROMETHEE	Preference ranking organization method for enrichment evaluations
PROM	Patient-reported outcome measure
PSM	Propensity score matching
QALY	Quality-adjusted life-year
QoL	Quality of life improvements
RI	Relative importance
RLH	Root likelihood
SELFIE	Sustainable intEgrated chronic care modeLs for multimorbidity: delivery, FInancing, and performance
SHAs	Strategic health authorities
Size	Size of target population
SITA	Strongly ignorable treatment assignment
SLoS	Scale Loss Score
SMD	Standardised mean difference
SMI	Severe mental illness

STP	Sustainability and Transformation Partnership
SUTVA	Stable Unit Treatment Value Assumption
TOPSIS	Technique for order preference by similarity to ideal solution
VCSE	Voluntary, community and social enterprise sector
VAS	Visual analogue scale
WHO	World Health Organisation
WTP	Willingness to pay
YoL	Additional years of life

Chapter 1. Introduction

1.1. Rationale for the thesis

New models of care & the need for evidence-based commissioning at local level in England

In every country, healthcare resources are limited and increasingly fail to meet the growing need and demand for health services. Over the past half century, health expenditure in OECD countries has consistently outpaced economic growth, a trend expected to continue due to factors such as expensive new technologies, an ageing population, and rising public demand and expectation (1). As a consequence, health policy-makers and payers aim to maximise societal welfare by allocating tight healthcare budgets to the most cost-effective interventions (2-4). In making these crucial resource allocation decisions in healthcare, it is essential to adopt a transparent and evidence-based approach. Such an approach ensures that resources are used to an optimal extent, while guaranteeing that the decision-making process remains objective and justifiable. This, in turn, contributes to the fairness and sustainability of the healthcare priority-setting process (5, 6).

In England, local commissioners of Integrated Care Systems (ICSs) are responsible for allocating over 75% of the National Health Service (NHS) budget by commissioning services from local healthcare providers to meet the health and some social care needs of their local communities (7, 8). Forty two ICSs were formally established on 1 July 2022, under the 2022 Health and Care Act, across England (9, 10). They are the result of a decade-long integration agenda, shaped by bottom-up pilots and changes to the local commissioning process. An ICS can be defined as a collaborative network or partnership between various health and care organizations within a geographical region, aiming to coordinate health and social care services, optimize healthcare delivery, and enhance overall population health. ICSs essentially comprise two components: Integrated Care Boards and Integrated Care Partnerships. Integrated Care Boards are responsible for the local NHS planning functions and resource allocation decisions that were previously the responsibility of Clinical

Commissioning Groups (CCGs) (11)[1]. Integrated Care Partnerships are statutory committees that bring together local authorities, local NHS organisations (e.g., Integrated Care Boards, NHS Trusts and Foundation Trusts), the voluntary, community and social enterprise sector (VCSE) and other stakeholders to develop a health and care strategy for the area (12). With the institutionalisation of ICSs, provider collaboratives have emerged as partnership arrangements of at least two trusts, working across multiple places with a shared purpose (13). The Health and Care Act of 2022 set out the primary goals of each ICS, including: (i) to enhance the health of the population and healthcare outcomes; (ii) to reduce disparities in outcomes, experiences, and access to healthcare; (iii) to promote productivity and ensure value for money; and (iv) to contribute to the NHS's wider role in social and economic development (10). In order to achieve these goals, ICSs are developing joint plans and making collaborative decisions on how partners can work together in their respective areas (14).

New models of care, also known as integrated care programmes, are essential in achieving the objectives pursued by ICSs. Originating in England in the early 2000s, these models emerged as bottom-up innovations motivated by the need to reduce care fragmentation, improve patient outcomes, and enhance efficiency. Other high-income countries worldwide have also widely adopted similar care models with the aim to promote integrated care (i.e., improve coordination across care settings and enhance overall health system efficiency) (15-18)[2]. The formal conceptualisation and

[1] CCGs were groups of general practices that had a commissioning directorate. They were introduced by the Health and Social Care Act 2012 as an attempt to enable NHS funding to be spent more effectively and to promote better collaboration across local government and the NHS (9-10).

[2] According to the World Health Organisation (WHO), integrated care is "a concept bringing together inputs, delivery, management and organization of services related to diagnosis, treatment, care, rehabilitation and health promotion. Integration is a means to improve services in relation to access, quality, user satisfaction and efficiency" (22). New models of care, or integrated car programs, vary in terms of the extent of integration. Some tend to focus on integration which brings together various providers of health care within the health system (e.g., integration between primary, secondary, and community care). Other initiatives are rather concerned with integration between health and social care providers. The literature also distinguishes between horizontal integration, which takes places between providers operating at the same level, and vertical integration in which integration takes place across providers working at different stages of care (16-17).

promotion of these new models of care in England can be traced back to the publication of the 2014 NHS Five Year Forward View (2014 FYF), which proposed a transformative vision for service delivery (19). This vision, advocating for an integrated, patient-centred, and efficient healthcare system, aligns closely with the principles of value-based care, an approach gaining momentum worldwide in response to the growing demand for healthcare, rising care costs, and stringent fiscal policies (20, 21)[3]. The 2014 FYF advocated for a range of new care models with a unifying emphasis on breaking down barriers between services and enhancing coordination between different parts of the healthcare system – including primary care, secondary care, and community health services. Although there is no universally agreed definition, new models of care can be defined as innovative approaches to organizing and delivering health services. They aim to meet the changing needs of patients and communities, and focus on improving the health of the population, patient experience of care and operational efficiency (22). These models are also recognized as complex interventions, as they feature multiple interacting components and target diverse outcomes, extending beyond gains in longevity and quality of life (23).

To ensure that new models of care have the desired impact and consolidate the newly established local commissioning structures, enhancing the local evidence-based decision-making process is crucial. NHS England has emphasized the importance of evidence-based local commissioning, which involves making decisions based on reliable evidence while recognising public preferences, to deliver effective and efficient healthcare interventions (7, 24). The use of disaggregated data is essential for taking action and promptly evaluating its impact (25). Furthermore, the evolving nature of new models of care underscores the need for ongoing monitoring and dynamic evaluation. This approach is essential not only for assessing the effectiveness and efficiency of these models but also for ensuring that they continue to meet the changing needs of the healthcare system and its

[3] "Value-based healthcare is the equitable, sustainable and transparent use of the available resources to achieve better outcomes and experiences for every person" (21)

beneficiaries (23). Without relevant and timely evidence, ICSs risk inefficiently allocating health budgets, potentially leading to suboptimal care and adverse outcomes.

However, challenges remain in realizing the full potential of these new models of care. There is evidence suggesting that local healthcare commissioners in England may lack a comprehensive understanding of the 'value for money' provided by these models. Despite the limited research on local healthcare commissioning in England, some studies suggest that local commissioning decisions are not always drawn on evidence. It is argued that local commissioners do not have enough capacity to use all routinely collected data and make well-informed decisions (26-29). This is unlike priority-setting at national level, where healthcare spending decisions are routinely informed by economic evaluations conducted by the National Institute for Health and Care Excellence (NICE) (30).

Multi-criteria decision analysis for assessing new models of care

Due to the challenges of local decision-making and the complexity of new care models, traditional health economic evaluation methods may not be suitable to assess them. Cost-effectiveness analysis (CEA), for instance, focuses on assessing an intervention's impact on a single health outcome, aiming to maximise health outcomes (e.g., life-years gained) and to minimise opportunity costs (3). This is problematic as ICSs and new models of care target outcomes that go beyond health-related gains. A similar problem arises with cost-utility analysis (CUA), a variant of CEA, as it also incorporates a single measure of health gain (i.e., quality-adjusted life year (QALY)) (ibid). Cost-benefit analysis (CBA), another widely used economic evaluation method, measures both the costs and outcomes of interventions in monetary terms, and thus allows non-health benefits to be considered (ibid). However, it may not be possible to monetise all outcomes of interest (e.g., patient experience), and even if possible it could potentially be costly and time-consuming (23).

Cost-consequence analysis (CCA) is another type of economic evaluation that allows for the inclusion of all relevant outcomes when assessing healthcare interventions (31). In principle, CCA could assist local commissioners in evaluating new models of care as it would allow them to account for the several outcomes in which they are interested in. However, while CCA can provide a comprehensive summary of the different costs and outcomes, the results are presented separately, making their interpretation more subjective than in other types of economic evaluation. This creates room for selectively highlighting positive outcomes (32)[4]. Program budgeting and marginal analysis (PBMA) has also been suggested as an approach for assessing interventions and setting priorities in healthcare. It involves evaluating the additional benefits and costs of incremental changes in program funding. Although benefits may consist of multiple attributes, PBMA often relies on subjective judgments (32, 33). The Social Return on Investment (SROI) is another approach that has been proposed for measuring and accounting for a much broader concept of value in healthcare (34). However, like CBA, SROI aims to assign a monetary value to non-financial outcomes which can be subjective (35). Moreover, the application of this approach demands significant time, expertise, and resources, which can pose challenges for local-level evaluations where commissioners are presented with multiple and diverse initiatives and require timely economic evidence for short-term decision-making (23, 35-37).

A framework based on a multi-criteria decision analysis (MCDA) could potentially support local commissioners to evaluate new models of care, combining economic evidence with public values. MCDA is a systematic and theory-based approach designed to perform a comparative analysis of several competing investment options based on their performance on multiple criteria. It stands out as an approach that allows for a systematic and impartial evaluation of healthcare interventions,

[4] Drummond and colleagues, in their widely cited work, "Methods for the Economic Evaluation of Health Care Programmes," define CCA as a form of partial economic evaluation as it does not aggregate or combine costs and consequences into a single metric (3).

including new models of care, bridging the gaps that characterise traditional health economics methods (23, 36, 37). Beyond measuring health impacts, MCDA enables the inclusion of additional decision-maker interests, such as equity and patient experience. Furthermore, its versatility allows for the explicit consideration of public preferences for multiple outcomes by assigning weights to each outcome of interest. This is particularly valuable as incorporating public preferences in the decision-making process enhances the legitimacy and accountability of the local commissioners' decisions (6). MCDA synthesizes the performance of interventions across multiple criteria into a single, comprehensive measure of effectiveness, allowing local decision-makers to develop a systematic ranking of various healthcare initiatives. Due to its comprehensive, transparent, and inclusive nature, MCDA has been suggested as a suitable approach for local commissioning (23, 36, 37), and the Medical Research Council in the United Kingdom has recognized it as an appropriate methodology for the economic evaluation of complex interventions (38).

MCDA is gaining widespread application in various healthcare decision contexts, especially in priority-setting, with decision-makers recognising its potential to enhance the quality and effectiveness of decision-making processes (39). The literature reviews conducted in this field have identified a relatively recent and significant increase in the use of MCDA in healthcare, with most of these MCDAs undertaken to support health care reimbursement and coverage decisions (39-42). In England, at national level, MCDA has been used to develop a prioritisation framework to assist local authorities make transparent, evidence-based spending decisions across public health programmes (43, 44). Additionally, some ICSs are beginning to use this approach to evaluate health interventions, including robotic surgery (45). While its application at the local level is still in the early stages, MCDA is increasingly recognized as a valuable technique to support local decision-makers in making investment decisions driven by data and population health needs (46).

The performance of MCDA involves eight steps, starting with the identification of the decision problem and the healthcare interventions to be compared (step 1) (47, 48). Once the decision goal has been set, the assessment criteria are identified along with the corresponding performance indicators (steps 2 and 3). Afterwards, performance indicators are converted onto a common scale or score (step 4) and weighting techniques are applied to elicit people' preferences between criteria (step 5). Performance values and criteria weights are aggregated in a MCDA total score (step 6), sensitivity analyses are conducted (step 7) and results are reported and examined (step 8).

A decision-support framework to support the local commissioning of new models of care

An MCDA-based decision-support framework could substantially improve the transparency, consistency, accountability, credibility, and acceptability of local health policy decisions in England. First, within each ICS, such a framework would enable local commissioners to systematically evaluate new models of care against the multiple outcomes that embody the elements of value central to integrated care. These outcomes, or criteria, selected to align with the objectives pursued by ICSs, would ensure that local commissioning decisions are directly tied to ICS goals, which extend beyond health-related outcomes. Second, with this framework, local commissioners would be able to explicitly assess trade-offs between the various outcomes of interest and clarify the primary drivers of final investment decisions. Third, the framework would enable consideration of public preferences in a systematic manner, explicitly recognising public values in the local priority-setting process. Fourth, an MCDA-based decision-support framework could facilitate a comparative analysis of many different healthcare interventions, including new models of care run by different provider collaboratives, enabling local commissioners to identify inefficiencies and make better-informed decisions when replicating or adapting care models.

The establishment of ICSs has created a sustained 'window of opportunity' to support local commissioners with evidence-based priority-setting in England (14). In a context where healthcare expenditure consistently outpaces growth and expectations and demands for integration are higher than ever, it is crucial to promote greater efficiency, while improving healthcare quality and outcomes. The development of a robust MCDA-based decision-support framework could be instrumental in achieving this goal.

1.2. Thesis aim and objectives

The aim of this thesis is to develop a systematic and accessible framework that incorporates evidence on multiple outcomes of interest to ICSs, to support the commissioning of new models of care in England. This thesis has six main objectives:

1. To describe the local commissioning decision context in England and identify the main challenges in that process.
2. To summarise the MCDA approaches that have been employed to generate evidence across different healthcare decision-contexts.
3. To identify the elements of value of ‘new models of care’ in England.
4. To construct an MCDA-based framework for the prioritization of new models of care in England.
5. To determine the relative importance of the criteria for assessing ‘new models of care’, based on preferences from the general public in England.
6. To explore the feasibility of the decision-support framework using a case study in mental health care.

1.3. Structure of the thesis

To effectively develop a decision-support framework for the commissioning of new models of care, a thorough understanding of the decision-making context is essential. **Chapter 2** addresses this by describing the evolution of integrated care in England and how the local commissioning of health services works. The chapter starts by examining the reforms and policy initiatives implemented by the government in recent years, summarising the development of the integrated care agenda in the country from 1990 to the present day. It is then followed by findings from 26 semi-structured interviews with stakeholders from South East England. Based on these interviews, the chapter describes the transition from CCGs to ICSs, the type of new models of care in place, the local investment decision process, the assessment of interventions and challenges in the local commissioning process. The prioritisation of interventions is also described, including a list of ideal criteria that according to the local stakeholders, should be guiding the local investment decision process in the country.

Chapter 3 presents the results of an updated systematic literature review on the use of MCDA in different healthcare decision-contexts, aiming to inform the decision-support framework that this thesis develops. The chapter provides a comprehensive overview of the different decision-contexts in healthcare. Afterwards, by decision-context, the chapter details the criteria used; the data sources and methods applied to derive performance scores; as well as the applied scoring and weighting techniques. Additionally, it discusses the approaches used to obtain overall values and rank the compared alternatives, and the type of sensitivity analysis conducted. The criteria most commonly used for priority-setting in high-income countries are used to inform the assessment criteria for the MCDA in the decision-support framework for assessing new models of care.

Drawing on the insights from policy document reviews, interviews, and the updated systematic literature review outlined in earlier chapters, **Chapter 4** develops the structure of a decision-support framework for the systematic monitoring and evaluation of healthcare interventions, including new models of care. The chapter starts by defining the fundamental principles underpinning the framework, essential for effectively supporting the local commissioning of new care models. The subsequent section describes the evaluation component of the framework, which is based on a MCDA. It details each one of the eight steps typically followed for the development of an MCDA, including the definition of the assessment criteria which is based on results of the semi-structured interviews and the systematic literature review conduced. Then, the monitoring component of the framework is described, including an example of a suggested monitoring dashboard. The chapter concludes with an exposition of how the proposed decision-support framework could be integrated into the local commissioning cycle to enhance the local decision-making processes.

Chapter 5 reports the methods and results of a stated preferences discrete choice experiment conducted, aiming to elicit MCDA criteria weights, based on preferences from the general public in England. The chapter begins by describing the decision model and descriptive framework of the experiment, including the elicitation task, the experimental design, the survey instrument and the data collection strategy. The following section describes the econometric models used to model public preferences and obtain criteria weights. These weights are elicited to be used in the assessment component –the MCDA– of the decision-support framework. The chapter concludes with a report and discussion of the results.

Chapter 6 explores the potential feasibility of the assessment component of the decision-support framework. The performance of early intervention in psychosis (EIP) services in Oxfordshire and Buckinghamshire is evaluated using electronic health records. The two alternatives under evaluation and the dataset used are described, followed by an explanation of the operationalization of the

assessment criteria. The chapter explains the application of the quasi-experimental method to assess the performance of two EIP services on the assessment criteria. It then describes the standardisation of the performance scores, how MCDA total scores are obtained, and the sensitivity analyses. Finally, the chapter presents and discusses the results of the service evaluation.

Chapter 7 summarises the findings of the analyses conducted in this thesis, considers the relevance of these findings for research, reflects on the added value of the decision-support framework that this thesis develops, and suggests potential implications for policy-making. The limitations of the analyses conducted and the decision-support framework developed are also discussed, along with proposed future areas for research.

1.4. Contribution to the literature

This thesis contributes to the literature in four research areas: the application of MCDA in healthcare, discrete choice experiments in health economics, the evaluation of new models of care in mental health, and local priority-setting in the context of integrated care.

The contribution to the application of MCDA in healthcare is made by the updated systematic literature review (**Chapter 3**). It is the first review in the field to include grey literature and summarise key methodological information by decision-context. Furthermore, the review consolidates the criteria used in healthcare-related MCDA studies, which could serve as a basis for developing a uniform set of top-level criteria by decision-context.

The second contribution of this thesis, in the field of discrete choice experiments (DCE) in health economics, is made with the experiment summarised in **Chapter 5**. To date, no study has explored

the factors that local healthcare commissioners and members of the public in England would give precedence to inform decisions in the context of integrated care. This thesis fills this gap by using a DCE to determine the relative importance of attributes of new models of care, based on preferences from the public. Additionally, the choice experiment conducted is the first to use a split-sample design that expresses costs in terms of income tax increase, contributing to the existing literature in the field.

The service evaluation conducted in **Chapter 6** contributes to the field of evaluating new models of care in mental health. The evaluation of two EIP services, using MCDA, quasi-experimental approaches and electronic health records from Oxford Health NHS Foundation Trust (OHFT), adds to the relatively few comprehensive and robust evaluations in the mental health sector. This adds to existing literature in the MCDA field that has demonstrated the usefulness of MCDA to assess healthcare interventions, particularly in the context of complex interventions. This evaluation is particularly relevant considering the significant burden of mental health conditions in the country and the current lack of comprehensive rationing frameworks for assessing mental health programmes.

Finally, this thesis as a whole contributes to the field of local priority-setting in the context of integrated care. The description of changes in the local commissioning landscape and how local commissioners in England invest in integrated care (**Chapter 2**) contributes to the limited literature and the current decision context. This is particularly relevant as ICSs are still consolidating their priority-setting processes and no up-to-date study has explored how local commissioners in England invest in integrated care. The overview of the assessment criteria that have been used in the context of priority-setting in healthcare (**Chapter 3**) can be further disaggregated by the type of country (e.g., high income or low-middle income countries) and is the first of its kind in the literature. Additionally, although several frameworks have been proposed for evaluating complex

interventions, such as new models of care, this thesis introduces the first decision-support framework that integrates the monitoring and evaluation stages inherent in the commissioning cycle, as described in **Chapter 4**. The proposed framework defines a comprehensive yet applicable set of value-based elements (derived from results of **Chapter 2** and **Chapter 3**), and allows for the explicit recognition of public preferences (see **Chapter 5**) in the local priority-setting process. Finally, the case study reported in **Chapter 6**, demonstrates that it is possible to use the MCDA-based framework developed in this thesis to generate evidence on the performance of new models of care in England.

1.5. Ethics approval

Ethics approval was granted by the University of Oxford's Medical Sciences Interdivisional Research Ethics Committee (Reference: R74765/RE001).

1.6. Conclusion

This introductory chapter has provided a comprehensive overview of the thesis, outlining its rationale, aims, and objectives. It has also detailed the structure of the subsequent chapters, elucidating how each contributes to the broader academic literature. The need for evidence-based prioritization at local level in England is clear, particularly as ICSs seek for effective decision-support frameworks to inform healthcare planning and resource allocation decisions. Given the limitations inherent in traditional economic evaluation methods, an MCDA-based framework emerges as a promising alternative, offering a more holistic and inclusive approach. The development and empirical testing of such a framework, which promises to enhance the decision-making processes in England, forms the core focus of this thesis.

Chapter 2. Changes and trends in the landscape of commissioning in England in the context of integrated care

Part of the research presented in this chapter has been published as peer-reviewed papers, and co-authors have given permission to include the material in this thesis:

Sections 2.2, 2.4 and 2.5 of this chapter have been published with the following citation: Gongora-Salazar, P., Glogowska, M., Fitzpatrick, R., Perera, R., & Tsiachristas, A. (2022). Commissioning [Integrated] Care in England: An Analysis of the Current Decision Context. International Journal of Integrated Care, 22(4), 3.

Section 2.3 of this chapter has been published with the following citation: Tsiachristas, A., Vrangbæk, K., Gongora-Salazar, P., & Kristensen, S. (2023). Integrated care in a Beveridge system: Experiences from England and Denmark. Health Economics, Policy and Law, 18(4), 345-361.

2.1. Introduction

A clear understanding of the decision-making context is essential for the development of a decision-support framework to effectively support the commissioning of integrated care. To ensure the usefulness and applicability of the framework, it needs to be aligned and tailored to the specific needs of decision-makers. This entails identifying the societal values that should inform investment decisions and recognising the challenges commissioners face in translating these values into practical implementation. It also requires an understanding of the complexity of the decision-making process, the governance structures and the nature of the interventions on which local commissioners must make decisions. In the context of integrated care, this understanding involves examining how integrated care programmes have emerged and how they are currently defined.

Since the 1990s, England has implemented numerous reforms and initiatives to promote integration within the NHS and between the NHS, local authorities and other organisations (49, 50). The Health and Care Act 2022 is the latest and most far-reaching reform, abolishing Clinical Commissioning Groups (CCGs), which have been responsible for commissioning the majority of hospital and community NHS services in local areas since 2012 (10, 51, 52), and replacing them with Integrated

Care Systems (ICSs). These newly established local commissioning structures are the result of a decade-long integration agenda, shaped by bottom-up pilots and changes to the local commissioning. Integration has become a key priority for NHS England, with all efforts towards promoting more coordinated and integrated forms of care provision. Expectations and demands for integration are higher than ever, and local commissioners are being asked to increase efficiency while achieving the best possible outcomes for the populations that they serve within a highly constrained budget (24, 53). This aligns with a global trend in which health systems are gradually moving towards value-based care[5], seeking to improve patient outcomes, enhance patient experience, and achieve better overall population health while ensuring the sustainable use of health resources (20, 54).

Despite the various studies on healthcare commissioning in England (28, 29, 55, 56), it is still unclear how local commissioners allocate budgets to healthcare services, particularly in the context of integrated care. It is not yet clear what factors drive the local decision-making process, who is involved, and what are the main challenges that commissioners face when investing in one intervention over others. This is in contrast to priority-setting at national level, where decisions on health spending are routinely informed by economic evaluations carried out by the National Institute for Health and Care Excellence (NICE) (28, 57). Integration of care is being promoted as a key driver of value in healthcare, but a clear and applicable definition of the elements of value is needed to implement this agenda in practice (25). Although some initiatives have proposed a route to defining value-based healthcare in the NHS[5], there is no agreed definition and its application in local healthcare priority-setting is still unclear (58).

[5] In a 2014 report, a group of scholars from the University of Oxford define value-based healthcare in the NHS as "*the equitable, sustainable and transparent use of the available resources to achieve better outcomes and experiences for every person*" (21).

The aim of this chapter is to describe the development of integrated care in England and how local commissioning of health services is working and making the transition to ICSs. To achieve this, (i) the various reforms and policy initiatives that the government has implemented in England over the past year are examined and (ii) semi-structured interviews with stakeholders are conducted to understand the investment decision process in integrated care. In the discussion section some recommendations to inform ICSs in their healthcare commissioning role are provided.

2.2. Methods

An analysis of the evolution of integrated care is based on a review and analysis of policy documents describing reforms and policy initiatives with implications for integrated care in England. Relevant documents were identified from the NHS website and through searching Google using a combination of the following search terms: “commissioning”, “priority-setting”, “investment”, “healthcare”, “integrated care”, “new models of care”, “local”, and “England”. Reports published by independent organisations in the UK that focus on integrated care and commissioning, such as the King’s Fund, were also considered.

To understand the investment decision process in integrated care amid the transition toward ICSs in the country, semi-structured interviews were conducted with local commissioners and other relevant stakeholders in South East England (59). The topic guide was designed to identify how is integrated care defined by commissioners, who are involved in the commissioning of integrated care programmes, how these programmes are assessed, how are financial flows structured and how are investment priorities set. The topic guide was tested in two pilot interviews, adjusted accordingly and further refined after the first few interviews to ensure that questions and concepts were not ambiguous. The topic guide is presented in **Appendix 2**.**A**.

A purposive-snowballing sampling approach was adopted to recruit interviewees from two large ICSs located in South East England (59). Local leaders of commissioning processes were identified through mapping work and suggestions from health experts working at the Oxford Academic Health Science Network and the National Institute for Health Research (NIHR) Oxford and Thames Valley Applied Research Collaboration. Potential participants were selected based on their experience or knowledge of local healthcare commissioning. Roles and type of stakeholders were pre-specified to guarantee equal representation of all levels and seniority. **Appendix 2.B** provides a list of interviewees, their anonymous codes, and their roles. **Appendix 2.C** lists and defines the local structures under the new commissioning context.

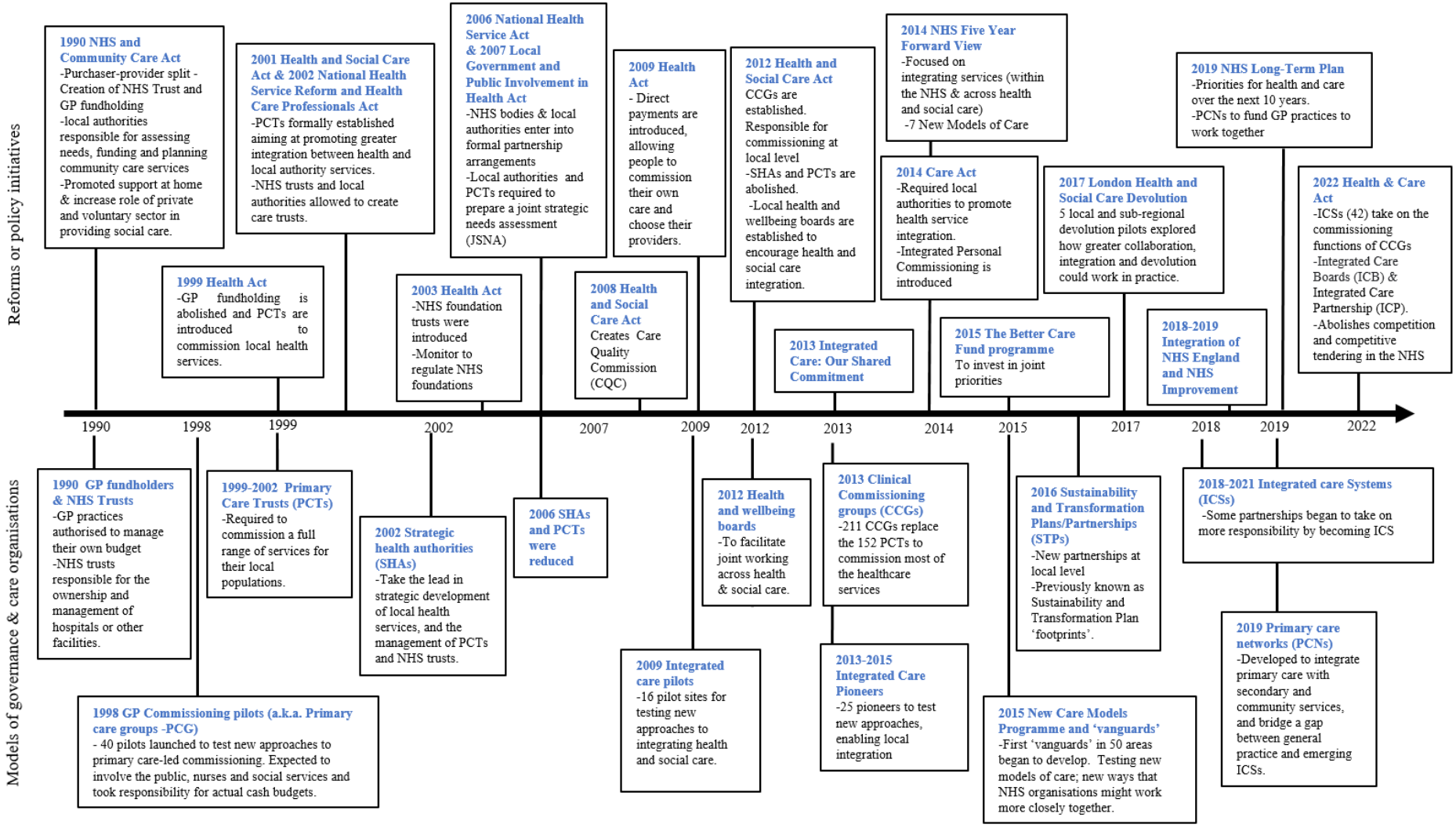
As a result of the above process, thirty-seven stakeholders were invited by email that included a participant information sheet about the research study, of whom 26 took part in a semi-structured interview (70% response rate). From the interviewees, 6 were directors/chief officers, 5 were chair, accountable officers or executive leads, 4 were head/lead commissioners, 3 were programme managers, 3 were clinicians (pathway leads), 3 were CSU members, and 2 were public health registrars. More than 80% of the interviewees were CCG members. All interviews were conducted online via Microsoft Teams 2017 (60) between April and July 2021, and were audio-recorded with the participants' consent. Interviews were then transcribed verbatim by a professional transcriber and one of the researchers verified the transcriptions against the audio-recordings (61). NVivo 12 software (62) was used to categorise and analyse data collected. When developing the thematic coding framework a hybrid approach of inductive and deductive coding was adopted (63), with themes being identified based on analysis of raw data and the four main interview topics. A selection of interview excerpts are shown in quotation marks, although supplementary quotes are provided in **Appendix 2.D**.

2.3. Evolution of integrated care in England

Promoting integrated care has been a long-standing concern in England, although it started to be strongly emphasised by the government in the late 90's through multiple initiatives and new forms of health care organisation (64). The timeline below (

Figure 2.1) attempts to summarise the interventions and structural changes associated with the promotion of integrated care. The upper part of the figure illustrates the legislation and policy initiatives, while the lower part shows the evolution of governance and care organisation. The following sections describe each of these policy reforms and the associated models of care that have emerged, and their relevance to integrated care.

Figure 2.1. Evolution of integrated care in England



1990-2000: The emergence of an integration agenda

The first policies aiming at encouraging integration of services emerged in the 1990's, while the government simultaneously implemented market-based healthcare delivery policies to increase efficiency and productivity in the NHS (50, 65).

Firstly, the NHS and Community Care Act 1990 formally laid the foundations of the internal market within the NHS by introducing the purchaser/provider split and giving greater autonomy to the newly created bodies (66-68). Healthcare commissioners were handed budgets to purchase services from health and social care providers. This was accompanied by the creation of NHS Trusts, which were responsible for owning and managing health facilities (e.g. hospitals) previously run by regional, county or special health authorities, and were allowed to borrow money and generate income from the provision of services (69).

The first pilots exploring new forms for the commissioning of health service emerged with the NHS and Community Care Act 1990, with forty GPs commissioning pilots, labelled as primary care groups (PCG) (69). These new bodies were responsible of a cash-limited prescribing budget for the planning and delivery of primary care, taking over the role of family practitioner committees. PCGs were also responsible for commissioning actual hospital and community health services budgets, and elements of general medical services budgets (70). In their role, they were encouraged to involve the public, nurses and social services, although no formal structures were put in place to enforce this collaboration.

Following the positive results of the GP commissioning pilots, and aiming at improving quality of care via better coordination at local level, the Health Act 1999 established primary care trusts (PCTs) to take on the local commissioning functions undertaken by health authorities and GP fundholders, with the

existing PCGs pilots evolving into PCTs (68). These new structures, covering a population of 165,000, were also allowed to provide some services directly to their local populations, although they were accountable to the local health authority and subject to direction from the Secretary of State (69, 71). To promote coordination between health and social care, the NHS Act 1999 allowed local authorities and PCTs to pool budgets, and the NHS and local authorities to integrate health and social care staff into a single organisation and delegate commissioning to a single lead organisation (69, 72).

2000-2010: Creating a context receptive to partnership working

Reforms implemented during this decade aimed at increasing central control, via 'command-and-control elements', and promoting quality of care and integration within the NHS and between health and social care.

The Health and Social Care Act 2001 allowed the NHS and local authorities to set up Care Trusts responsible for planning and delivering integrated services, covering health and community care, becoming the first model close to a full merger of health and social care in the country (50, 73). To support this, the government increased funding for health and social care between 2000 and 2007, and health authorities or PCTs that outperformed against pre-determined criteria received additional funding from the Secretary of State (68, 69, 73). Central control was strengthened by giving the Secretary of State the power to intervene in an NHS body that was not performing adequately, and with the creation of the National Care Standards Commission to regulate health and social care services and promote improvements in the quality of services (69).

A year later, the 8 regional offices and 95 regional health authorities were renamed and merged to form 28 strategic health authorities (SHAs), covering 1.2–2.7 million people, via the National Health Service

Reform and Health Care Professions Act 2002 (74). The SHAs were responsible for managing performance, issuing directives and implementing health policy in their respective areas, ensuring that improvements in health and health services were delivered locally by PCTs and NHS Trusts (67). With the SHAs, a degree of local discretion was allowed as to how different organisations would work together (50).

The act also introduced two new bodies to strengthen patient and public involvement, and encourage integration on the workforce component (74). An independent patients' forum for every NHS trust and PCT in England was created, with monitoring, inspection and representation functions on behalf of the public. In addition, the Council for the Regulation of Health Care Professionals was created aiming at overseeing the activities of different regulatory bodies of the healthcare professions. Promoting cooperation between regulatory bodies and coordination of good practice guidance was one of the functions of this Council.

One year after, the Health Act 2003 established NHS foundation trusts and a new regulatory body known as Monitor (75). NHS foundation trusts were given greater autonomy and allowed to retain surpluses, invest in new service provisions, and introduce flexible staff pay, allowing them to increase integration across services (69). Unlike PCTs, performance of NHS foundation trust was not regulated by the Secretary of state but by Monitor.

Aiming at encouraging integration between health and social care, the National Health Service Act 2006 required NHS bodies (e.g., PCTs, SHAs) and local authorities to enter into formal partnership arrangements and to pool funds when necessary (76). A year later, this was complemented by the introduction of the Health Act 2007, which required PCTs and local authorities to work together to identify the relevant health and social care needs of their populations and to produce a Joint Strategic

Needs Assessment (JSNA) on a regular basis (69, 77). The same act replaced the patients' forums, created in 2002, with ‘local involvement networks’ aiming at giving communities the opportunity to influence the planning and delivery of both health and social care.

In the subsequent years, two acts aimed at improving the quality of health and social care services and increase patient’s choice. In 2008, with the Health and Care Act 2008, a new act established the Care Quality Commission, responsible for protecting the health, safety and welfare of users of both health and social care services (78). This new independent regulator was required to carry out inspection and enforcement activities to ensure that registered providers were meeting standards of quality and safety (69). In congruence with this, in 2009, the Health Act 2009 required all providers to publish information relevant to the quality of NHS services (79). The same act introduced direct payments, aiming at giving patients greater scope to choose their providers and commission their own health and social care (68, 80).

New integrated care pilots were launched in 2009 to test different approaches to service delivery that would enable greater integration of care (69, 81). These included variations on case management, multidisciplinary team working, and the development of new organisational structures, with a particular focus on horizontal integration and services for older people and those with co-morbidities (e.g., integration between general practice, community nursing services and social services) (82). Sixteen sites were designated across the country, and ran for two years. At the time, evidence on the effectiveness of new integrated care structures was somewhat limited. Early studies from experiences in Australia, Germany and the United States provided insights into their potential effectiveness, with indications that integrated care could enhance patient satisfaction and hospitalisation rates for patients with heart failure (83, 88). However, the evidence on the impact of integration overall on quality of care, patient outcomes and cost savings was inconclusive (82, 88). There were also calls for a focus on

clinical and service integration, as organisational integration alone may not lead to better outcomes (85). The effectiveness of integration was found to be highly dependent on contextual factors and the presence of shared values, roles, training and effective communication to facilitate multidisciplinary teamwork (81, 82).

The evaluation programme of the new integrated care pilots in England found no improvements in patient experience or hospital use, despite positive perceptions from staff (82). However, the same evaluation concluded that although well led and managed pilots can improve the quality of care, improvements are unlikely to be seen in the short term.

2010-2022: Integration at the core of the local commissioning structures

The Health and Social Care Act 2012 introduced an extensive governance and service delivery reorganisation, aimed at closer integration both within the NHS and across health and social care (11). The new act established Clinical Commissioning Groups (CCGs), made up of local general practitioners (GPs), to take over the functions held by the PCTs and the SHAs, being responsible for commissioning GP, community health, acute care and mental health services for their local communities. The new CCGs were responsible to control the majority of the overall NHS budget (i.e., £60 billion and £80 billion), with specialised services commissioned by NHS England, and health improvement services commissioned by local authorities. The new Act also encouraged them to improve services quality, promote patient choice and reduce inequalities by overseeing local markets and securing that health services were provided in an integrated way (50, 84). In parallel, the same Act established local Health and Wellbeing Boards (HWB), as a forum responsible for facilitating joint commissioning across health and social care organisations, with a particular focus on adult social care and public health (50). From a regulatory perspective, the role of the Care Quality Commission was strengthened and required to

work in cooperation with Monitor, which was given additional duties, including overseeing aspects of access and competition in the NHS (11).

In the following years, the government introduced a number of action plans and pilots to support integrated care. In 2013, the National collaboration for Integrated Care and Support, a partnership of organisations including NHS England and the Local Government Association, published a framework document on integration, 'Integrated care and support: our shared commitment'. The document outlined strategies for local areas to use established structures like HWBs as a means to facilitate collaboration among local authorities, the NHS, care and support providers, education, housing services, public health, and various stakeholders to advance integration efforts (69, 86).

In the same year, 25 health and social care integration pioneers were set up to test new approaches, working across health, social care and public health, as well as with local authorities and local voluntary organisations (67). A year later, the Care Act 2014 promoted integrated personal commissioning and legislated for the right to a personal budget. It set new standards for social care, gave carers the right to receive the same level of support as those they care for, and required local authorities to promote the integration of care with health services and ensure continuity of care when a provider ceases to provide a service (69). In line with this, local authorities were required to provide comprehensive information and advice about social care services in their local area (87). In the same year, the NHS Five Year Forward view (2014 FYF) was published, setting up the challenges facing the NHS and how the delivery of health service needed to change, including a range of so-called “new models of care” (19). These new models of care advocate for the dissolution of the traditional boundaries across levels of care (i.e., primary care, community services, and hospitals) and call for a more personalised and coordinated

provision of care to patients[6]. With this policy document the notion of competition starts to debilitate, while the integration agenda is strongly emphasised.

To deliver the integration agenda set out in the 2014 FYF, 50 'vanguard' sites have been selected to implement the proposed 'new models of care' and a new fund has been announced to support integration between health and social care (69). In 2015, 'integrated primary and acute care systems', 'enhanced health in care homes' and 'multi-specialty community providers' were the first new models to be tested in 29 vanguards, followed by 8 more vanguards to improve coordination of urgent and emergency care. Not much later, more radical approaches to the organisation of hospital care were implemented in 13 new vanguards (89). These vanguards were one of the milestones in the development of new models of care in England and the structural arrangements introduced by the government shortly afterwards (69). In the same year, the Better Care Fund (BCF) was launched to support local health and social care providers to work together and invest in shared priorities. To access BCF funding (£3.8 billion), CCGs, local authorities and health and wellbeing boards had to work together to agree a joint area plan (50).

To continue the implementation of the 2014 FYF, in December 2015 local authorities, CCGs and providers were asked to work together to produce Sustainability and Transformation Plans (STPs), which can be described as 5-years plans that outline the changes to be introduced to accelerate the implementation of the FYF (90). Forty-four STPs were submitted and used as the basis for 44

[6] In principle, seven new models of care were specified (19): (i) Multispecialty community providers under which the leadership of primary care is expended, including nurses, therapists and other community based professionals; (ii) "Vertically" integrated Primary and Care Acute care systems, which allows single organisations to provide NHS list-based GP and hospital services, together with mental health and community care services; (iii) Urgent and Emergency Care networks, which organises and simplifies the system via networks of linked hospitals; (iv) Viable smaller hospitals, which includes new organisational models for smaller acute hospitals that enable them to gain the benefits of scale without necessarily having to centralise services; (v) Specialised care, with networks of services over a geography, integrating different organisations and services around patients; (vi) Modern maternity services, including strategies that make it easier for groups of midwives to set up their own NHS-funded midwifery services; and (vii) Enhanced health in care homes, promoting new shared models of in-reach support, including medical reviews, medication reviews, and rehab services to improve quality of life and reduce hospital bed use.

Sustainability and Transformation Partnerships (STPs). England was divided into 44 STPs, through which local authorities and local NHS organisations, including CCGs, were asked to develop proposals to improve health and quality of care, to deliver better services for patients in the areas they serve, and to address the challenges set out in the FYF (69). Although the STPs had no legal status, several mergers between local hospitals, community health and mental health providers took place during these years (50). Some of these STPs voluntarily evolved into Integrated Care Systems (ICSs), where commissioners and providers, often in partnership with local authorities, took collective responsibility for the health of their population and the health (and in some cases social) care resources required (51, 91). Along this, five local and sub-regional devolution pilots across London explored how greater collaboration, integration and devolution could work in practice, including impacts within and beyond the London system (69, 92). These devolution pilots included joint workforce strategic planning with the creation of a London Workforce Board. This board was responsible for seeking agreement amongst member organisations to pool resources where appropriate (92).

In 2018 and 2019, the government introduced further developments to reinforce the integration agenda. In terms of governance, NHS England and NHS Improvement[7] (the formerly known as Monitor), merged into a single organisation, aiming at improving efficiency and effectiveness of collective NHS resources, by removing unnecessary duplication, and breaking down boundaries between purchasers and providers (93). In 2019, the NHS long-term plan is published, setting out the priorities for health and care over the following 10 years (24). To incentivise systems-based working and support the delivery of integrated care, the plan committed to giving STPs and ICSs more flexibility, allowing them to make financially neutral changes to control totals for individual organisations within their systems. The plan also encouraged GPs to join together to form 'primary care networks', working as 'truly integrated teams of GPs, community health and social care staff' covering communities of around

[7] NHS Improvement was responsible for overseeing providers and was the result of the merger between Monitor and the NHS Trust Development Authority in 2016.

30,000 to 50,000 people (24). Other commitments included (i) the development of 'new models of care' through a comprehensive offer for children and young people and the redesign of core community mental health services; (ii) specific allocations for primary and community care to improve out-of-hospital care; and (iii) the development of 'fully integrated community health care' through multidisciplinary teams including GPs, pharmacists, district nurses and allied health professionals working across primary care and hospital settings. (24, 94). All commitments and actions were conceived to ensure that by 2021 every area of England would be part of an ICS, with primary and community services funded to do more.

In line with the intentions and commitments outlined in the 2019 Long Term Plan, the Health and Care Act 2022 replaced the CCGs with ICSs, giving them statutory status and formally dividing the country into 42 ICSs, covering populations of around 500,000 to 3 million people (10). Each ICS has an Integrated Care Board, which takes on the NHS planning functions previously held by the CCGs. Within each ICS, Integrated Care Boards, health and social care providers, and the local authorities, who are responsible for commissioning social care services (51), are expected to work under a new statutory partnership to promote integration and deliver the best health and social care to their local populations (10, 53, 95, 96). For that to happen, each Integrated Care Board and their partner local authorities are being required to establish an Integrated Care Partnership (97). Integration between health and social care is also supported with a new standalone legal basis for the BCF, with the government committing with a minimum of £7.2 billion to the BCF in 2022 (98). Some CCGs started to merge and work as non-statutory ICSs since 2020, with ‘integrated care programmes’ or ‘new models of care’ as the spearhead to promote the integration agenda (24, 99, 100). The new Act also abolished competition and competitive tendering within the NHS, reducing transaction costs and giving commissioners greater flexibility (101). The Health and Care Act 2022 is arguably the most radical reform in recent years. However, its impact on the local commissioning process of integrated care

remains to be fully understood and assessed. The section below elaborates further on the transition that the local commissioning in England is currently undergoing.

2.4. Current local commissioning context: The transition towards Integrated Care Systems

Integrated care systems (ICSs) are partnerships of organisations that work together to plan and deliver integrated health and care services[8] to (i) improve outcomes in population health and healthcare in their area; (ii) reduce inequalities in outcomes, experience and access; and (iii) promote productivity and value for money (12). Although the design and implementation of ICSs is locally led, it is done within a broad national framework and is accountable to the NHS England. While some of the ICSs emerged since 2016, others are still building on the foundations of the previous CCGs to fully implement the new governance structures (14).

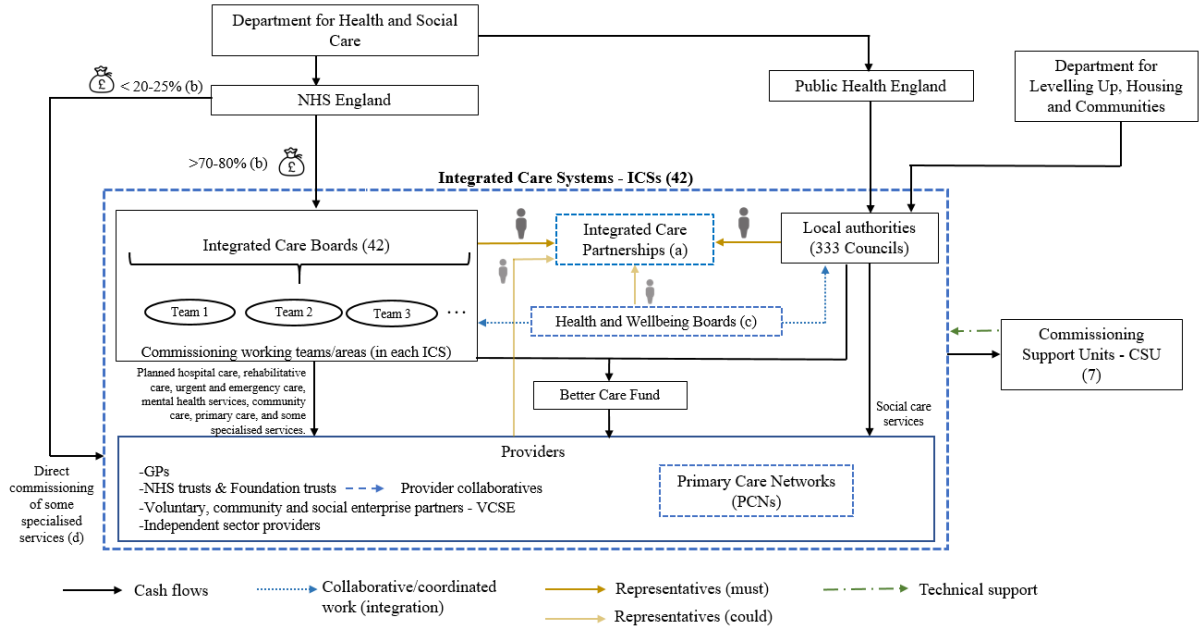
Within each ICS, Integrated Care Boards are assuming the local NHS planning and taking over the commissioning functions previously carried out by CCGs, in addition to some of the direct commissioning functions that NHS England used to have. Local authorities remain responsible for social care services in the ICS area, including home care and residential care services (12). They maintain a separate administrative and decision-making structure from the ICSs, although they are expected to work closely with integrated care boards and providers within the ICS framework. To this purpose, Integrated Care Partnerships are being established as a broad alliance between the NHS, local authorities and providers (10, 53, 102, 103).

[8] There is no universally agreed definition of integrated care, with over 170 variations identified in the literature (17). The World Health Organization provides a broad definition, stating that integrated care involves coordinating input, delivery, management, and organization of services across healthcare settings to improve access, quality, user satisfaction, and efficiency (22). The service-based definition of integrated care refers to collaboration between two or more professionals from different disciplines, starting with a holistic assessment of patients in their local environment (17). The focus of this section is primarily on organizational and systemic integration (ibid).

Each Integrated Care Partnership has developed or is in the process of developing an 'integrated care strategy' for its whole population (104). Providers are becoming constituent members of the Integrated Care Partnership and the Integrated Care Boards, and GP practices are being grouped into primary care networks to operate in specific neighbourhoods within the ICS (51). Along with this, some 'provider collaboratives' are emerging as partnership arrangements involving two or more NHS Trusts (13). Under this new architecture, HWBs remain as a formal statutory committee instilling mechanisms for joint working across health and care organisations (105). Commissioning support units (CSUs), which were operating as external consultancy advisory, providing business intelligence, health & clinical procurement, and contract management services to support the CCGs, are expanding their services to support Integrated Care Boards and, local authorities, NHS trusts, among other organisations (106, 107).

Similarly as when CCGs were in place, under the new architecture, the money comes down in cascade from the Department of Health & Social Care and NHS England to all the Integrated Care Boards, following and adjusted allocation formulae (108). These resources represent over 70-80% of the NHS England budget and are tied to specific funding streams determined by NHS (e.g., maternity, community services, mental health) (107, 108). The BCF remains as a pooled budget between the NHS England and local authorities, aiming at supporting local partnerships, and with this, the implementation of integrated health and social care services (109, 110). **Figure 2.2** summarises the new local commissioning decision context, including the main cash flows. **Appendix 2.C** provides a summary of the new local commissioning structures, and **Appendix 2.E** summarises the former local commissioning structure, with the CCGs as the main actors in the commissioning decisions.

Figure 2.2. The current commissioning decision-context after the abolishment of CCGs



Notes: Percentages (%) in the figure refer to the proportion of the NHS England budget.
(a) Integrated care partnership members must include representatives from the local authorities and from the Integrated Care Boards. Beyond this, members may be from other organisations, including housing and education providers (12).
(b) Financial allocations to each Integrated Care Board include resources for a range of functions that the NHS England and NHS Improvement used to have (e.g. other primary care budgets) (12).
(c) In some ICSs there might be more than one HWBs (95, 111)
(d) With the 2022 Health and Care Act, within a framework of continued national accountability, national standards, Integrated Care Boards are now allowed to take on delegated responsibility for commissioning specialised services (112)

ICS arrangements vary significantly, depending on the size and scale of each system. Each ICS is defining its place-based partnership arrangements, with small ICSs operating mainly as a single place-based partnership (113, 114)[9]. NHS England is already making financial allocations to each one of the ICSs (115), and each system is deciding how to distribute these resources across their place-based partnership and the different providers (116). Integrated Care Boards are responsible for deciding

[9] Place-based partnership arrangements and contracts are designed to serve specific geographic areas within defined boundaries. This approach can potentially limit the choice of people living at the borders of the defined ICSs as individuals may find themselves in a situation where they are served by one ICS but are geographically closer to providers affiliated with another. Furthermore, individuals may encounter administrative obstacles when trying to access services across ICS boundaries due to varying contractual arrangements and funding streams, which can complicate referrals and care coordination.

whether allocations are routed through a series of contracts, or whether they prefer delegating the responsibility for securing services for a population (10). Contracts between Integrated Care Boards and providers may be managed by place-based partnerships or provider collaboratives. This transformation is not entirely new since, in some places, providers already took commissioning powers for specialised mental health services (51). As was the case with the CCGs, Integrated Care Boards are able to commission jointly with local authorities. More pooled budgets and joint appointments across the two organisations are likely to emerge as a result of this transition (12, 51).

A more in detailed description of the transition that the system is undergoing is presented in the following sub-sections, based on the analysis of the interviews.

2.4.1. Understanding of integrated care initiatives

When interviewees were asked to define integrated care, most of them provided a person-centred definition, highlighting the importance of adopting a holistic approach when addressing individuals' needs, and consequently the relevance of services working together regardless of the organisation the teams or care professionals belong to.

> "it's literally everybody in all aspects working together to identify what the challenges are, what it is we need to overcome, and how we work together" (No. 8)

There are multiple integrated care programmes in place, although there is not a formal terminology to refer to these initiatives (e.g., programmes, interventions, work-streams). Participants distinguished between two types of integrated care programmes that in practice tend to overlap: (i) condition-specific interventions with elements of integration, such as the 'Integrated diabetes care programme', and (ii) interventions that look at the wider picture and overlap with the condition-specific programmes, such as the 'Aging well programme', which looks at individuals with 3 or 4 long-term conditions. Examples of integrated care programmes in place, mentioned by the participants, are provided in **Appendix 2.D**.

Despite the existence of these programmes, some of them are not yet fully integrated in practice. According to some interviewees, most of the integration has arisen in the form of partnerships, via memorandums of understanding, rather than in the provision of care sphere. Patients' information is not yet shared with all teams involved, limiting the coordination in the planning and provision of care. Among other limitations, participants argued that the lack of flexibility in the contractual structure precludes the reorganisation of care services to operate under a common structure. Some participants were also sceptical about the idea of integration, arguing that it is an over-used phrase or an unfilled promise made 20 years ago.

2.4.2. Investment decision process

According to the interviewees, there was a top-down and bottom-up prioritization process within the CCGs. On the one hand, every year NHS England set national priorities and provided guidance to the CCGs, indicating the outcomes to be achieved. Based on this, CCGs tailored local patient pathways and introduced changes in the provision of care. Some arrangements were also directly introduced at national level, irrespective of what was happening locally, or what was seen as priorities for the CCG governing body. Some examples are the introduction of the remote 'total triage' model in 2020 (117) or the six new models of care outlined in the 2014 FYF (19). Based on this, the CCG governing body or the HWBs discussed and set priorities, which were subsequently communicated to the different commissioning working teams (104, 118).

On the other hand, priorities also emerged from needs or gaps that commissioners identified locally, as illustrated in **Figure 2.3**. Each commissioning working team worked together with GPs and clinical leads to identify improvements needed. If a service-specific need was detected, the commissioning working team worked on the basic structure and requirements of that service, and prepared a business

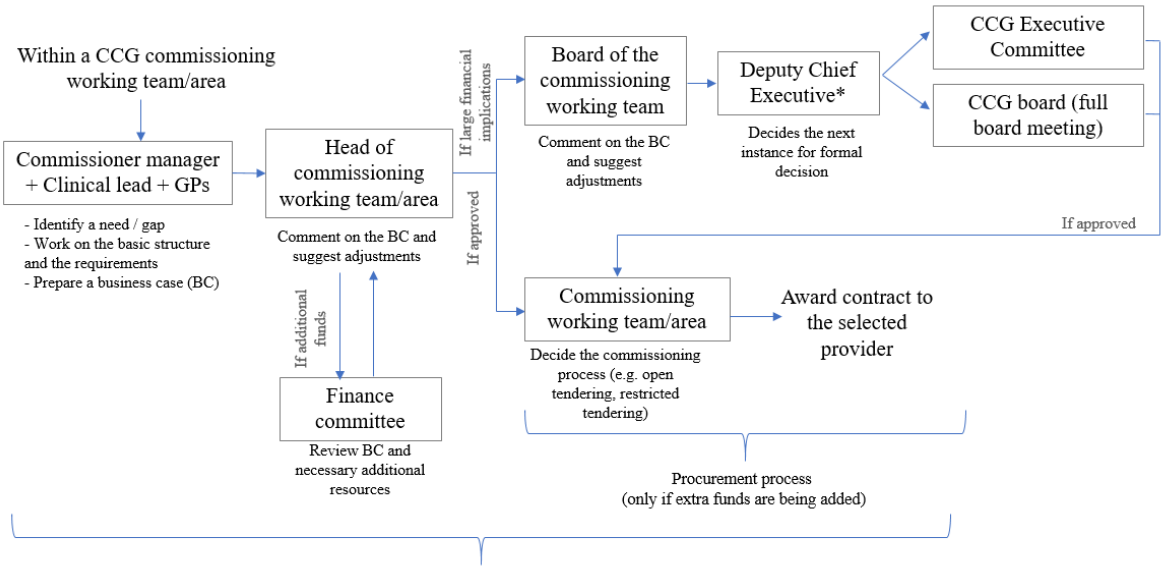
case. Managers costed up the proposed improvement and checked whether it was following all the equality and diversity legislation (119). The finance team took part in the approval decision if the proposal required additional funds. If the new service or arrangement in the proposed business case had a substantial investment implication, the board of the corresponding commissioning working team reviewed the business case. Subsequently, the business plan was examined by the Deputy Chief Executive for approval, who then decided whether to send it to the CCG Executive Committee or to the CCG board for formal decision.

The NHS England was normally not involved in the local commissioning of services, unless the proposal had a direct or indirect impact on specialised services. If the business case involved arrangements around the provision of social care or the use of funds from pooled budgets between the council and the CCG, public health teams from the council also got involved in the shaping and implementation of the business case. Although patients and the public were sometimes involved in the commissioning process, it seems that there was a lack of clear arrangement for making their engagement more systematic.

> "For patient involvement, that depends very much on the project manager in the area of work. Some are very good at involving patients and making sure there's an input and engagement, and others not" (No. 18)

Participants did not mention major differences between the commissioning of individual interventions and integrated care programmes.

Figure 2.3. Bottom-up investment decision process within a CCG or Integrated Care Boards



* When explaining this process, some participants referred to the integrated care board while others to the CCG role.

Despite the new management structures and committees, in the opinion of various interviewees, most of the CCG staff will be transferred to the Integrated Care Boards. This goes in line with NHS guidance, indicating that each CCG's staff, assets and liabilities will be transferred to the relevant Integrated Care Boards (120). No substantial changes were expected in the commissioning process in itself, as many features of CCG were believed to remain within the new place-based partnership arrangements. By the time the interviews were conducted, however, there was a great deal of uncertainty on how each of the ICSs will develop and how the organisational landscape will look like.

> "So I'd say the main difference would be that we would have providers in the conversations much earlier than we used to and sitting alongside us" (No. 23)

2.4.3. Prioritisation of interventions

Participants were asked about the approaches taken by local commissioners to prioritise health interventions, including integrated care programmes, and the factors that drive local investment

decisions. Various participants maintained that there was no structured framework but rather a mixture of subjective and objective approaches, with no differences between simple or single interventions and integrated care programmes. The many objectives to fulfil and the multiple sources of information to account for, hindered the development of a robust prioritization process using explicit criteria. Some interviewees, however, highlighted the existence of some frameworks or reports, such as the JSNA (104) or the products developed by the RightCare Intelligence (121) Programme which allowed them to identify local needs or quality issues. In spite of this, decisions were not always guided by evidence. Almost all of the participants argued that NHS England directives and budgetary considerations were the main drivers of prioritization of health interventions at local level. Although to a minor extent, some participants also mentioned leadership and ability to lobby for particular interventions as potential determinants of final investment decisions.

> "I think we are heavily influenced by national directives […] And the second thing […] I would say the finance, I'm afraid, does influence it big time" (No. 20)

These investment drivers contrast with the criteria that, according to the respondents, should be guiding the local investment decision process. **Table 2.1** provides an overview of these ideal criteria and the frequency of interviewees that mentioned them. Health outcomes at individual and population level were highlighted by over 90% of the interviewees that addressed this question. Similarly, quality of care, including patient experience, was underlined by the majority (70%). The third most popular criteria was the 'cost/financial considerations'. These first criteria are known in the literature as the 'Triple aim' of integrated care (17, 22, 122), a framework directly mentioned by some of the participants.

Table 2.1. Ideal criteria for prioritization of interventions

Criteria	# Participants
Health outcomes e.g. *'mortality and morbidity assessments'(No.5) 'QALYs'(No.17), 'HbA1c' (No.2)*	24
Quality of care e.g. '*Best possible health and care experience' (No. 1), 'better access, like reducing waiting times' (No. 16); ''getting the pathway right'(No.22)*	18
Cost/budget considerations e.g. '*financial implications of the programmes'(No.14); 'financial burden of the condition'(No.2); 'saving money'(No.15)*	12
Efficiency e.g. *'value for money'(No.12); 'cost effectiveness of that service'(No.25)*	7
Equity e.g. '*how well resources are distributed to different groups in the population' (No.10); 'reduce unwarranted variation'(No.3); 'equality'(No.24)*	6
Compliance to policies/priorities/guidance e.g. '*national directives' (No.20), 'integration with social care services' (No.1);*	5
Size of the population e.g. '*make a difference to a large amount of people'(No.15); 'high prevalence conditions' (No.12)*	3
Others e.g. *'meeting a completely unmet need'(No.18); 'whole system benefit'(No.26); 'focuses more on prevention'(No.8)*	

Note: In total, the 26 participants responded to this question.

2.4.4. Assessment of interventions

The assessment of healthcare interventions, including integrated care programmes, did not seem to be a systematic process and varied across interventions. For certain specialities, there were some population health monitoring tools in place, mostly in the form of dashboards, with developments at national and CCG levels (123). Some one-off evaluations were also conducted, using a mixture of qualitative and quantitative methods, aiming at identifying service failures or assessing the efficiency or effectiveness of interventions. Most of these evaluations, however, apparently lacked methodological rigorousness and there were only few counterfactual analyses (124).

> "It hasn't been systematic. It's been driven by looking at the integrated care program that we put in place, trying to work through what we think the impact has been, uhm… and using data to back that up. The problem has been demonstrating cause and effect" (No.6)

To compare healthcare interventions against standards, national trends, or other regions, participants mentioned multiple national (121, 125, 126) and local (104, 127) sources of information. Examples of these are the clinical audits and registries run by NHS Digital (126) and the CCG Integrated Performance Report (127), which looks at the overall performance of the CCG and the associated organisations . These tools, however, do not include in detail information of interventions' performance, but rather a summary of performance and quality issues across broad areas within the CCG. Finance, quality and service issues with current providers used to be reviewed based on key performance indicators (KPIs). Despite the multiplicity of indicators, no metric has been developed to measure the degree or success of integrated care, and few KPIs can be used for comparison purposes. This is particularly relevant when health inequalities is one of the major NHS England concerns (24).

Prospective evaluations were also mentioned by participants, and were normally included in the business cases that working teams prepared. Some CCGs conducted quality and equity impact assessments, in which they considered the potential impact of the new programme or service on patient safety, clinical outcomes, patient experience of care delivery, and equality of access.

2.4.5. Challenges in the commissioning process

Access and use of information was one of the most common barriers highlighted by interviewees. Relevant data was sometimes held by different organisations, with different IT systems, making it difficult for commissioners to access timeliness linked databases. In some cases, commissioners had to pay or obtain permission to use certain databases (e.g. NHS primary care). Participants also mentioned data quality issues and information overload, which prevented commissioners from monitoring intervention's performance and conducting a comprehensive analysis of local population needs (128). The need to gather more data on patient experience or satisfaction was also highlighted (122). Some interviewees also argued that CCGs lacked the capacity to use all data available to inform the

commission process. The CSUs offered support in this regard, although various participants saw them as bureaucratic and unresponsive organisations. This might partially explain why CCGs bought fewer services from CSUs over the last years while investing in developing more in-house capabilities (129).

> "I do think it's access to information. I think it's an access to data and being able to pull the data and being able to understand the data" (No.1)

Financial constraints were also underlined as a challenge by most interviewees. The combination of increasing demand for care and costs, achieving new targets, and dealing with reduced funding, made it difficult to finance integrated care. An increasing number of CCGs were overspending against their planned expenditure (129). Furthermore, each one of the commissioning working teams had to work within the funding allocation to different services set by the centre (130, 131), which limited CCGs' ability to redirect funds and reprocess investment.

Operational challenges were also highlighted. Most of the contracts with providers were long and had different termination dates, hindering the planning and implementation of service transformations. The procurement of services itself was also stressed as a barrier due to the limited staff, the complexity of all associated rules and policies, and with this, the involvement of many actors. Workforce constraints not only impaired the procurement process but all other commissioning tasks. Recruiting and retaining skilled health and social care workers was constantly a challenge, particularly in rural areas. Inequalities in access to healthcare due to limited supply of services or quality issues in under-served areas of the country have been documented in the literature (132-137) and acknowledged by the NHS England (24, 107, 138).

From a governance point of view, nearly all participants considered that CCGs had little flexibility to set local priorities. According to some interviewees, the pressure to comply with the multiplicity of NHS England constitutional targets, together with the constant changes introduced at national level, obstructed the prioritisation of local needs. The lack of a truly integrated system approach was also stressed by various participants. Despite all organisational changes introduced over recent years, the different teams and organisations worked in silos, with competing financial priorities and a focus on acute care. In some cases, there was also a disconnection between the executive committee and those on the ground doing the commissioning. In this context, working relationships played a crucial role in the planning and delivering of integrated care.

All these financial and human resources limitations, together with contractual and procurement constraints, hindered the commissioning of integrated care. It seems that most of the commissioning was about introducing changes around the edges, rather than radical modifications to the way services are being provided to the population.

According to some interviewees, the Health and Care reform might have the potential to address some of the commissioning challenges mentioned above (53, 96). Some participants indicated that ICSs were starting to invest in new software developments and analytics (102, 139) to aggregate and analyse data from multiple care providers. Others also mentioned that changes to the system architecture are expected to offer commissioners more flexibility in terms of budget allocations and the use of their workforce across larger geography, with the possibility to gain some efficiency. Under the new procurement system, the role of competition and tendering is being reduced, and local commissioners are allegedly having more flexibility in the arrangement of services (12).

Some respondents were however more sceptical, and saw this transformation as a top-down reorganisation of the commission system. According to some of them, ICSs will not solve the problem of providers being financially constrained to operate on a system basis. Bringing all stakeholders together will probably be the biggest challenge, particularly for ICSs merging with big or multiple CCGs. Others also suggested that national priorities will still dominate over the local ones. Nevertheless, by the time the interviews were conducted, there was still uncertainty around the statutory framework, and the NHS was in the process of publishing guidance on the implementation of ICSs (102).

2.5. Discussion and Conclusion

Over the past three decades, England has introduced numerous policy initiatives that have gradually shaped the integration agenda that underpins current commissioning structures. While the initial emphasis was on competition and the separation of commissioners and providers, a transformation has occurred over time, with a noticeable shift towards prioritizing collaboration and integration. This may in part be the result of lessons learned from experimenting with competition in health care, with some evidence suggesting that market-based interventions have potentially led to fragmented care and unequal access to services, although this remains a subject of debate (50, 140).

Local health commissioning in England is currently undergoing a transition to a new organisational architecture that aims to further strengthen the NHS integration agenda. Although there was a broad consensus among respondents about the benefits of integrated care (141), most integrated care programmes were still in the process of implementation and their commissioning had not yet been finalised. This may be due to the various barriers in the commissioning process identified when conducting the interviews. Professional and organisational cultures as well as ways of working also

play a significant role. Skills in change management, engagement, and leadership are essential for supporting and developing integrated care initiatives in this changing landscape (29, 81).

The new ICSs are building on the CCG legacy and inheriting most local commissioning processes and structures that the CCGs developed. Findings from this study can potentially inform local commissioners and the NHS England, and shed light on potential areas of improvement or strategies to address the identified commissioning barriers.

Analysis of the last 30 years in England reveals a system that is constantly experimenting, evolving and learning from local experience to promote integration. However, frequent changes in local commissioning structures can hinder the consolidation of the proposed changes. In some cases, reforms have taken place in less than two years, preventing the desired outcomes from being achieved. If the promise of integrated care is to be realised, it is essential to stop the constant shifting of responsibilities.

While greater flexibility allows better responsiveness to context-specific needs, attention should be paid to potential inequalities that may arise between different ICSs. Bottom-up local initiatives have helped to shape the integration agenda that the national government is now promoting, and in the same spirit the reform gives local commissioners greater discretion in the arrangements they put in place, including flexibility to delegate commissioning place-based partnerships. Systems that are more advanced in consolidating partnerships (e.g., STPs operating as ICSs before 2022) are more likely to deliver on the promise of integrated care. However, to ensure success across all ICSs and reduce the variation that is already observed due to differences in size and structure, support from the central level remains crucial, particularly for those ICSs at earlier stages of development.

Findings from the semi-structure interviews highlighted the need to strengthen the evaluation culture and capacity, and ensure that evaluations are effectively embedded in the local commissioning process. Robust evaluations, particularly of integrated care programmes, have proved to be not only beneficial but also highly valued by commissioners and providers (142, 143). Even though current evidence indicates that integrated care is likely to reduce cost and improve outcomes (including health and patient experience) (55, 144, 145), more rigorous economic evaluations are needed as the evidence is of poor methodological quality (145, 146). In evaluating integrated care programmes, it is essential to acknowledge the complex nature of these interventions and the consequent limitation of traditional analytical methods (17, 22, 23, 147, 148). As indicated in the previous chapter, multi-criteria decision analysis has been suggested as a sound and comprehensive method that can overcome the limitations of traditional frameworks in the context of integrated care (36, 149, 150).

The interviews revealed an interest in data-driven commissioning of integrated care and the need to improve the availability and quality of data for this purpose (29, 142, 151-153). Routine collection and use of data related to the ideal criteria, particularly on patient experience, would enable the monitoring and performance evaluation of integrated care programmes. Although CSUs were partly conceived to support commissioners with the access and use of data, there was a general preference –among commissioners- for CSUs services to be carried out directly by the CCGs or ICSs. This is in line with the findings of Petsoulas et al (2014) (153). There is therefore a need for ICBs to invest in improving data collection, aggregation and curation, while developing in-house analytical capabilities. At the same time, this also suggests the need for CSUs to redefine their roles in the new commissioning landscape.

Although it is unclear how the top-down and bottom-up prioritization processes interacted, national directives and financial constraints played a predominant role in local healthcare investment decisions. This contrasts with the multiplicity of reports and data sources available. These results are in line with findings from Currie et al (2018), who concluded that commissioners had insufficient capacity to use

evidence to inform decisions and powerful actors may influence knowledge absorption processes in the CCGs, aiming to achieve their goals (29). The disconnection between evidence and final investment decisions can be partially explained by the absence of a structured prioritisation framework. Findings of this chapter revealed what seems to be a set of desirable prioritisation criteria that could be used to develop a clear prioritisation framework that acknowledges the complexity of integrated care programmes. However, it is important to acknowledge that while such a framework could inform local priority setting, commissioners' flexibility is limited as ICSs are still bounded by NHS England directives and budgetary constraints. As interviews took place during the nascent stages at a time when ICSs were just starting to emerge, it remains to be seen whether local commissioners have gained greater room for setting priorities while navigating within the broader parameters established by NHS England.

With the ICSs now in place, it is also worth reflecting on the distinction between the commissioning of simple interventions and integrated care programmes. In contrast with simple interventions, integrated care requires, for instance, the involvement of multiple actors and joint or more flexible contracting tools (28). Simple interventions are, however, progressively disappearing as multidisciplinary teams and integrated care pathways are becoming part of usual practice through updated guidelines (154-156, 160). This might explain why interviewees did not raise differences between the commissioning of simple interventions and integrated care. In light of this, it is desirable to promote commissioning processes and tools, including contractual models, suitable for integrated care.

Although it remains to be seen how the Health and Care Act 2022 will address all challenges in the commissioning of integrated care, the findings reported in this chapter suggest that the majority of commissioners see ICSs as an opportunity to pool budgets, implement shared-decision making and materialize the idea of integrated care. This enthusiasm was also identified by the King's Fund when exploring how ICSs are starting to develop across the country (100). However, it is too early to draw

conclusions as there is still uncertainty about how each of the forty-two ICSs will consolidate their governance structure and partnerships at local level (100, 101, 142).

Strengths and Limitations

One of the major strengths of the study is the diversity of the 26 interviews conducted, with professionals with different roles and varying levels of seniority. Due to the plurality of interviewees, it was possible to build a more comprehensive view of the commissioning challenges and bring commissioners' voices at all levels to the ongoing discussion on improving the commissioning of integrated care. Conducting this research in the midst of the transition toward ICSs is the second strength. This increases the chances that results are taken into consideration by local commissioners, as they are currently in the process of defining how the new commissioning structures will work in practice, and how they will relate to one another. Results of this study have been already presented to the local commissioners involved in the interviews, and were used for discussions within some of the ICS' teams. The novelty of this study is the third strength as this is the first study that explores changes in the local commissioning landscape and how local commissioners in England invest in integrated care. Previous studies have focussed on the use of evidence in the commissioning process (157), the prioritisation of public health interventions (27), the development of integrated care programmes (100, 142), the contractual options for integrated care (28, 158) or the effectiveness of clinical leadership in the commissioning of healthcare (151).

The external validity of these findings is limited by the focus of this study on only two ICSs in South East England. Future research would benefit from the inclusion of more ICSs, as there is great variation in terms of the size, complexity and development of the different ICSs across the country (100). Nevertheless, the two ICSs selected for the study have different governance structures and are at different stages of maturity. These differences and the similarities found in previous literature make the

findings of this chapter likely to be transferable to other ICSs. Another limitation arises from the lack of interviewees from NHS England or the Department of Health and Social Care. Their perceptions would have enriched this study with their perspective and views, as they are the main precursors of the transformation in the local commissioning process. Interviews with provider professionals could have also provided a valuable perspective on the arrangements for delivering integrated care, but they were not possible due to time and ethical approval constraints. Similarly, only two public registers from the local authorities were interviewed and the local commissioning of social care was not examined in depth.

It is important to note that the analysis of the development of integrated care was limited to care integration within the health care system. Exploring the commissioning process from the local authorities' perspective would have offered a more comprehensive understanding of the local commissioning of integrated care. Nevertheless, it remains to be seen how Integrated Care Partnerships and HWBs will work together. Finally, it could be argued that the transitional nature of the context may render some of the findings of this study in the near future. The interviews were also conducted in 2021, and since then ICSs have been consolidating their commissioning arrangements (14). However, most of the CCGs' staff have been transferred to the corresponding integrated care boards (120) and latest reviews on ICSs are consistent with several of the findings and recommendations reported in this chapter (14,159).

Conclusion

The new local commissioning structures provide an opportunity for commissioners to strengthen and streamline the commissioning of integrated care. Despite the many reforms implemented over the years, the last decade has increasingly confirmed the focus on integration and collaboration and laid the foundations for ICSs to potentially deliver on the longstanding promise of integrated care. ICSs and the NHS England can learn from the evolution of integrated care, and in particular, from the challenges

faced by the CCGs. Improvements in the collection and use of performance data and the adoption of an evidence-based priority-setting framework could contribute to overcoming some of the challenges identified during the interviews and improving local investment decisions in England. Findings from this chapter can inform the development of such a framework. In particular, the ideal criteria identified can inform the definition of the elements of value to guide the local commissioning of integrated care in England. The increasing demand and interest in data-driven commissioning and the need for ongoing monitoring and evaluation, and timely decision-making, also highlights the importance of underpinning this framework with routinely collected data.

Chapter 3. The use of Multi-criteria Decision Analysis to inform decisions in Healthcare: An Updated Systematic Literature Review

The research presented in this chapter has been published as a peer-reviewed paper with the following citation, and co-authors have given permission to include the material in this thesis:

Gongora-Salazar, P., Rocks, S., Fahr, P., Rivero-Arias, O., & Tsiachristas, A. (2023). The Use of Multi-Criteria Decision Analysis (MCDA) to Support Decision-Making in Healthcare: An Updated Systematic Literature Review. Value in Health, 26(5): 780–790

3.1. Introduction

Multi-criteria decision analysis (MCDA) is defined as a systematic and theory-based approach to perform a comparative analysis of several competing options (e.g., healthcare interventions) based on their performance on multiple and often conflicting criteria (161, 162). It facilitates the structuring of complex decision problems and explicitly defines the importance of criteria through the assignment of weights, adding transparency and legitimacy to the decision-making process (47).

The growing need for MCDA in healthcare is reflected by the steep increase of empirical studies across different decision-contexts. MCDA has been employed by health technology assessment (HTA) agencies to inform regulatory decisions on medicines or medical devices (47, 163-166) and in shared decision-making to bring together clinicians' expertise and patients' preferences (47, 167, 168). In the context of priority-setting, governments are increasingly employing MCDA to identify the most beneficial public health interventions (43, 44), to appraise and rank unmet medical needs (169, 170), and to guide the design of national health benefit packages (171, 172). In the context of COVID-19, numerous MCDAs were also conducted to guide pandemic government responses (173-177).

As indicated in previous chapters, an MCDA-based decision-support framework has the potential to support local commissioners in evaluating new models of care, combining economic evidence with public values. It stands out as an approach that allows for a systematic and impartial evaluation of healthcare interventions, including new models of care, bridging the gaps that characterise traditional health economics methods (23, 36, 37). Due to its ability to offer comprehensive, transparent, and inclusive decision-making by considering multiple criteria and stakeholder perspectives, tailored to specific local healthcare needs and priorities, MCDA has been suggested as a suitable approach for local commissioning (ibid). Additionally, the UK Medical Research Council guidance has recommended MCDA for the economic evaluation of complex interventions in healthcare, such as new models of care (38, 150).

Although the application of MCDA in healthcare has been summarised in several review studies, none of them has systematically detailed how MCDA has been applied within and across different healthcare decision-contexts. The most comprehensive review of MCDA studies was published nine years ago by Marsh et al (2014) (39) and did not examine the methodological differences by decision-context. In addition, all previous review studies did not include grey literature, such as technical documents from HTA agencies and policy reports from healthcare authorities or multi-lateral bodies, thereby increasing the risk of missing important applications of MCDA in healthcare.

This chapter presents an updated systematic literature review to provide a comprehensive overview of MCDA studies in different healthcare decision-contexts and inform the decision-support framework that this thesis develops, aiming at supporting the local commissioning of new models of care in England. To achieve this, the review provides a detailed overview of (a) the different decision-contexts in healthcare; and then by decision-context; (b) the criteria used; (c) the data sources and methods used to derive performance scores; (d) the applied scoring and weighting techniques; (e) the approaches used

to obtain overall values and rank the compared alternatives; and (f) the type of sensitivity analysis conducted.

3.2. Methods

This review was registered in PROSPERO [CRD42020219093] in November 2020. The results were reported in accordance with the PRISMA[10] 2020 statement (178, 179), the guidance for updated systematic reviews (180), and recently published systematic literature reviews in the Value in Health Journal.

Search strategy

To update the review by Marsh et al (2014) (39), a database search was conducted for scientific literature published between August 2013 and November 2020. This review was selected for updating due to its extensive nature, including a large number of studies, detailed information extraction, and thorough data synthesis. Since Marsh et al. (2014) covered studies up to 2013, the search also included more recent MCDA studies found in the reviews by Adunlin et al (2015) (41) and Frazão et al (2018) (40), which also examined the application of MCDA in different decision-contexts in healthcare. Additionally, grey literature (181) published between January 1990 and November 2020 was examined to broaden the scope of the original review by Marsh et al. All searches were conducted by PG under the guidance of a librarian who advised on the search strategy and selection of databases.

Variations in search terms were added[11] and terms were searched as free text keywords in the title and abstract fields (182) (see

[10] Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[11] Truncation, spelling differences and abbreviations.

Table 3.1). The search terms were based on Marsh et al (2014) (39), Adunlin et al (2015) (41) and Frazão et al (2018) (40). Scientific studies were identified from the following databases: MEDLINE, EMBASE and Google Scholar (100 first hits).

Table 3.1. Search terms used

'Multi-criteri* decision*'	`Risk benefit assessment*´	`Analytical hierarchy process OR AHP´
`Multicriteri* decision*´	'health*', 'medicine*'	'Analytic network process OR ANP´
`Multiple criteri* decision*´	`Multi-attribute utility´	'measuring attractiveness by a categorical based evaluation technique OR MACBETH'
`Multicriteri* analysis´	`Multiattribute utility theory´	'technique for order preference by similarity to ideal solution OR TOPSIS'
`MCDM´	'Weighted product method´	`Preference ranking organization method of enrichment evaluation OR PROMETHEE´
`MCDA´	`Goal programming'	`Elimination and choice expressing reality´
`Benefit risk assessment*´	`Multi-objective optim*´	'weighted product model'

It is expected that all relevant academic studies are identified from MEDLINE and EMBASE, because they are high-level academic databases. Nevertheless, Google Scholar was used as an additional tool to locate potentially relevant studies that might not have been included in MEDLINE or EMBASE. While for primary review searches Google Scholar is unsuitable, it is considered a suitable supplementary source of evidence when conducting systematic literature reviews (183, 184). Due to the vast number of results displayed and the inability to directly export results in bulk as citations, systematic reviews typically screen the first 50 to 100 search records within Google Scholar (184, 185).

To identify grey literature – online resources – searches were conducted in Trip Medical Database[12] and Google. In Trip Medical Database records classified as evidence-based synopses and guidelines

[12] Trip Medical Database is a clinical search engine designed to find and use high-quality research evidence. In Trip Medical Database results can be filtered by the following categories: Evidence-based synopses, Systematic reviews, Guidelines, Clinical Q&A, and Primary research (i.e. academic studies). In the search for grey

were reviewed, using the title and short text underneath. Relevant titles were selected for further screening by a reviewer using the eligibility criteria. All previously defined search terms were used. Google (184) was used to search within 53 websites of governmental entities, networks, societies and international bodies in health economics, using the "site" option in the search box (e.g. site:nhs.uk) and all previously defined search terms[13]. For Google, the search was limited to four search terms as searches cannot be limited to abstract and title, and it was expected that relevant grey literature will have any of the following terms: "Multi-criteria decision" health; "Multicriteria decision" *health*; "Multiple criteria decision" health; "Multicriteria analysis" health.

Selection process and eligibility criteria

All study titles and abstracts were exported to EndNoteX9 (186). Before the screening of titles and abstracts, all reviewers met to finalise the eligibility criteria. Titles and abstracts were doubled screened independently by two reviewers (PG, SR and an Anonymous Reviewer (AR)): PG screened titles and abstracts of all records, and the same studies were equally split and screened by AR and SR. Disagreements between reviewers were resolved through discussion, and the fourth reviewer (AT) was consulted when necessary. All studies were assessed and included based on the following eligibility criteria:

Inclusion criteria:

(i) Full-text studies in English language;

literature of empirical MCDA studies, the guidance of the involved librarian was followed to ensure the inclusion of only evidence-based synopses and guidelines in the Trip search.

[13] This includes: The International Network of Agencies for Health Technology Assessment (INAHTA); active websites of INHATA members; World Health Organization (WHO); the Professional Society for Health Economics and Outcomes Research (ISPOR); the European Network for Health Technology Assessment (EUnetHTA); the EuroScan International Network; Health Technology Assessment international (HTAi), and; International Symposium on the Analytic Hierarchy Process.

(ii) Studies that reported an empirical application of MCDA to inform decisions in healthcare and described the MCDA methods used;

(iii) Studies that specified the MCDA method implemented.

Exclusion criteria:

(i) Studies that did not apply MCDA;

(ii) Studies that did not complete all the steps in the MCDA definition stated in the introduction (e.g., relevant criteria were identified but weights were not elicited or defined);

(iii) Studies that did not aim to inform a decision in healthcare (e.g., studies that used MCDA to derive health status utility values and with this calculate QALYs.

Similar to Marsh et al's study (2014) (39), studies were not excluded on the basis of methodological quality. However, a broader definition of healthcare was adopted, using a system perspective. This approach included all studies addressing decisions within the healthcare sector or the entire healthcare system, encompassing production, sourcing, organization, and delivery of healthcare. Studies focusing on interventions delivered outside the health sector (e.g., social, occupational, or environmental interventions), despite their potential health impact, were excluded.

Data extraction and analysis

A comprehensive template for data extraction was developed and piloted to ensure high level of consistency in extracting data between reviewers (PG, PF and AR). The data were grouped in decision-contexts, criteria, weighting and scoring techniques, sources of data, stakeholders involved and methods to address uncertainty. The categorization of decision-contexts was based on Marsh et al (2014) (39), but another layer of detail was included by providing sub-categories in each decision-context. Then, the

selected studies were allocated to each sub-category using the reported (often in the study aim/objectives) information on what decision each MCDA study was aiming to inform. To define the different criteria categories, the taxonomies from Tsiachristas et al. (2013) (187) and Marsh et al. (2014) (39) were used, together with an explicit code book with definitions and examples for each category. To identify the type of uncertainty addressed by the MCDA studies, Broekhuizen et al's classification was used (188).

Descriptive statistics were used for analysing data with regards to the decision-context, the applied MCDA techniques, and the type of stakeholders that participated in each of the MCDA steeps. Microsoft Excel (189) was used for constructing all tables, figures and descriptive statistics, and summarised the results for each of the MCDA steps (47, 48).

Reviewers' consistency

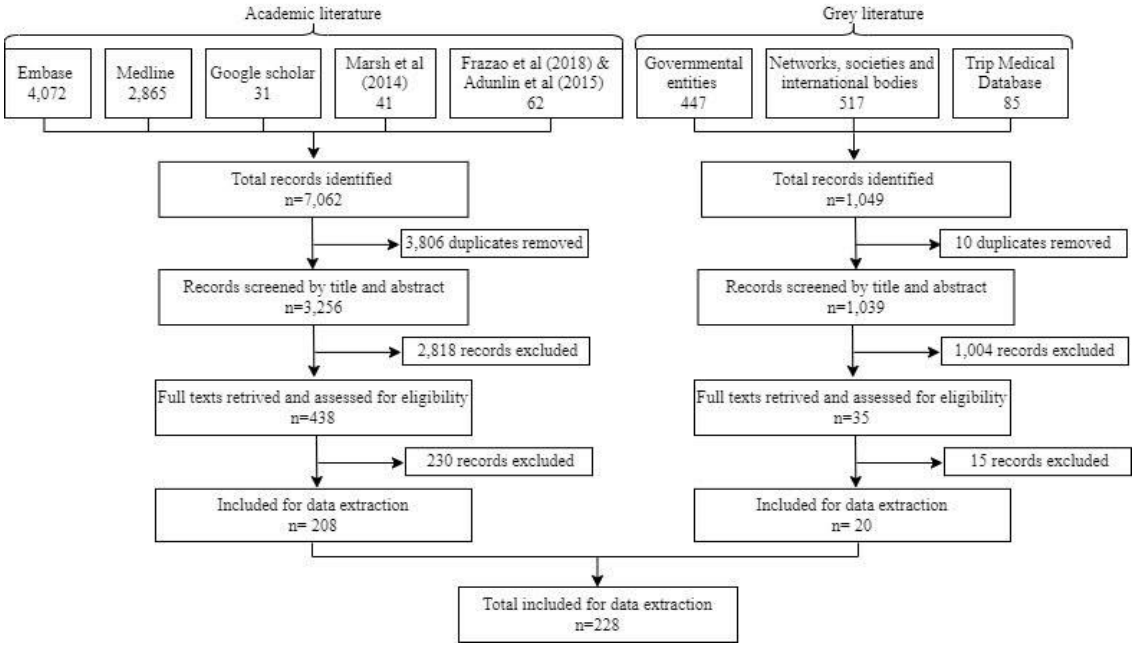
For reducing potential biases during the assessment and data extraction process, and following good practice guidelines (190), a selection of papers was discussed between reviewers during the agreement meeting. PG reviewed and extracted data from a sample of 10% of the studies reviewed by AR and PF, respectively. AR and PF did the same with 10% of the records reviewed by PG. The extracted data was then compared and a strong consensus was achieved between reviewers for both the assessment and extraction process (over 95% and over 85% inter-reviewer agreement, respectively).

3.3. Results

The PRISMA flowchart of this review is presented in **Figure 3.1**. After removing duplicates, 4,295 studies (3,256 and 1,039 studies from the database and grey literature search, respectively) were retained. After the screening of all records, 438 academic titles and 35 grey literature documents met

inclusion criteria. The full-text examination of the retrieved records excluded 230 academic studies and 15 grey literature documents. The analysis included in the end 228 eligible studies (219 MCDA studies). A summary of the results of all searches conducted, with the list of studies included in the review, can be found in **Appendix 3.A and Appendix 3.C**.

Figure 3.1. PRISMA flowchart

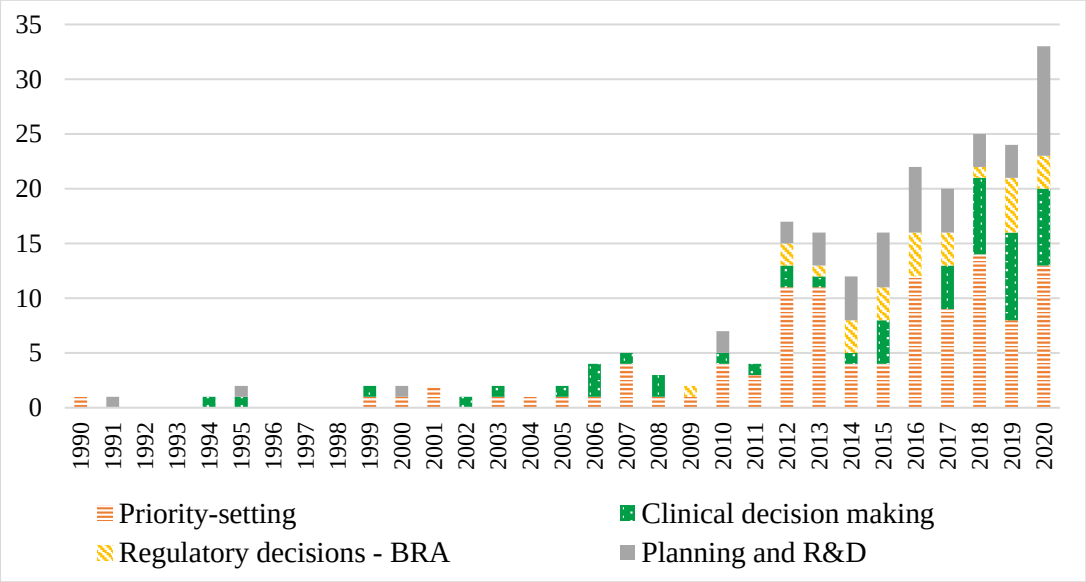


3.3.1. Study characteristics

Figure 3.2 illustrates the number of MCDA studies (i.e. 219 MCDA studies in total) in healthcare between 1990 and 2020. Although the first studies were conducted in the early 90s, it is since 2010 that a sharp upward trend can be observed, with no grey literature found before 2010. Studies informing priority-setting decisions were the first and have been the most predominant, although the share of studies informing other decision-contexts has increased over the last ten years. A definition of each one of the decision-contexts identified is provided in **Appendix 3.B**.

In almost 90% (n=192) of the MCDA studies, academia was involved either solely (62%, n=135) or in conjunction with the public sector (18%, n=39), the private sector (10%, n=22), or international organisations (3%, n=7).

Figure 3.2. Publication trend of MCDA studies in healthcare by decision context



Notes: BRA indicates benefit-risk assessment; MCDA, multi-criteria decision analysis; R&D, research and development.

3.3.2. Decision-context

As presented in **Table 3.2**, the majority of studies have informed priority-setting decisions (48%, n=107). Roughly 22% (n=48) MCDA studies were concerned with clinical-decision making, and 11% (n=25) with guiding regulatory decisions such as marketing authorizations. Although to a lesser extent, MCDA studies have also informed hospitals or health systems when making capital investment decisions (4%, n=8) or decisions on the location or reallocation of healthcare services or facilities (4%, n=8). MCDA studies have also been conducted to support manufacturers in the development of new medicines (4%, n=8), to guide research decisions (2%, n=5), to inform the regulation of addictive substances (1%, n=3), and to support the selection of best health-care wastes management method (1%, n=3).

Table 3.2. Decision-contexts that MCDA studies have been applied in

Decision context		**N**	**% of 219[(a)]**
Priority-setting	Prioritization of interventions for coverage / reimbursement /funding /future developments	85	39
	Prioritization of diseases for coverage / reimbursement / funding	12	5
	Prioritization of patients to access health care	10	5
Clinical decision making	Treatments or Prescription	34	16
	Screening	14	6
Regulatory decisions	Benefit-risk assessment (BRA) to issue recommendations[(b)]	25	11
Planning and R&D	Capital investment decisions	8	4
	Location / Reallocation decisions	8	4
	Guide pharmaceutical developments	8	4
	Research purposes	5	2
	Regulation of addictive substances	3	1
	Decisions on hospital wastes management decisions	3	1
	Other investing planning decisions[(c)]	9	3

Notes: BRA indicates benefit-risk assessment; MCDA, multi-criteria decision analysis; R&D, research and development. (a) 5 MCDA studies inform more than one type of decision. (b) For example, market authorization. (c) For example, MCDA to guide the selection of healthcare providers or to choose best length of stay policy alternatives.

As presented in **Table 3.3**, most of the MCDA studies (63%, n=137) have taken place in high-income countries, principally in Europe (37%, n=82) and North America (19%, n=41). With exception of regulatory decision-contexts, where almost all studies have informed high-income countries (96%, n=22), approximately a third of MCDA studies applied in the other decision-contexts have taken place in low-middle income countries.

Over a third (35%, n=76) of the MCDA studies have been developed to inform decisions at national level, followed by studies conducted to inform decisions at local level (14%, n=31). Nevertheless, 21% (n=47) of the studies did not report the decision-level. Regulators or budget holders are the most common users of MCDA in healthcare, with over 40% (n=90) of the studies informing ministries of

health, commissioners, public health institutes or local authorities. However, in clinical decision-making, 56% (n=27) of the MCDA studies aimed to inform physicians, 29% (n=14) informed patients or carers, and only 10% (n=5) informed regulators or budget holders.

MCDA studies in healthcare have evaluated diverse types of alternatives. More than 40% (n=93) of the studies have assessed pharmaceuticals, followed by public health (14%, n=31) and screening interventions (12%, n=27). Pharmaceuticals are the most evaluated alternative (41%, n=90) across all decision-contexts, a frequency that goes up to 92% (n=23) in the regulatory decision-context. The second most frequent are public health interventions among priority-setting studies (26%, n=28) and medical screening programmes in the case of clinical-decision making (25%, n=12). Nearly 16% (n=35) of the eligible studies evaluated more than one type of alternative.

In terms of the number of alternatives assessed, on average, MCDA studies have evaluated 15 alternatives (median=5, SD=39.8), with a minimum of one and a maximum of 500. Priority-setting studies evaluated on average 20 alternatives (median=8, SD=51, min=1, max=500). This contrasts with the lower number of alternatives assessed when informing diagnosis or prescription decisions (mean=5.1, median=4, SD=4.6, min:1, max:22), regulatory-decision (mean=4.9, median=4, SD=3.6, min:1, max:15) or planning and R&D decisions (mean=18.7, median=6, SD=39.8, min=1, max=194).

Table 3.3. Decision-makers, type of decisions and types of alternatives under evaluation

	Priority-setting		**Clinical decision making**		**Regulatory decisions - BRA**		**Planning and R&D decisions**		**All contexts**	
	No	**%**	**No**	**%**	**No**	**%**	**No**	**%**	**No**	**%**
a. Type of economies (a)										
High-income economy	68	64%	32	67%	22	96%	20	45%	137	63%
Upper-middle income economy	17	16%	13	27%	1	4%	9	20%	40	19%
Lower-middle income economy	13	12%	3	6%	0	0%	10	23%	26	12%
Global	7	7%	0	0%	0	0%	4	9%	11	5%
Low-income economy	1	1%	0	0%	0	0%	1	2%	2	1%

b. Level of decision										
National	56	52%	10	21%	3	12%	9	20%	76	35%
Not reported	8	7%	20	42%	12	48%	9	20%	47	21%
Local (b)	15	14%	7	15%	0	0%	10	23%	31	14%
Institution (c)	7	7%	6	13%	1	4%	10	23%	24	11%
Regional (> 1 country)	13	12%	2	4%	8	32%	2	5%	25	11%
Global	7	7%	0	0%	0	0%	4	9%	11	5%
Individual person	1	1%	3	6%	1	4%	0	0%	5	2%
c. Type of decision-maker										
Regulators / Budget holders	62	58%	5	10%	7	28%	18	41%	90	41%
Physicians	11	10%	27	56%	9	36%	3	7%	49	22%
Health care Institutions	13	12%	5	10%	0	0%	13	30%	31	14%
HTA bodies (d)	16	15%	1	2%	2	8%	3	7%	21	10%
Patients / Carers	1	1%	14	29%	2	8%	0	0%	17	8%
Not reported / Not clear	5	5%	5	10%	4	16%	3	7%	16	7%
Private companies	2	2%	2	4%	4	16%	5	11%	13	6%
Research institutions	5	5%	2	4%	1	4%	3	7%	11	5%
Other (e)	9	8%	0	0%	1	4%	1	2%	11	5%
General population	1	1%	1	2%	0	0%	0	0%	2	1%
d. Type of alternatives										
Pharmaceuticals	42	39%	28	58%	23	92%	4	9%	93	42%
Public health interventions	28	26%	0	0%	1	4%	2	5%	31	14%
Screening	13	12%	12	25%	1	4%	2	5%	27	12%
Non-surgical interventions & Non pharmacological methods	15	14%	5	10%	0	0%	0	0%	20	9%
Diseases / conditions	15	14%	0	0%	0	0%	3	7%	18	8%
Surgical interventions	10	9%	6	13%	0	0%	1	2%	17	8%
Medical devices	10	9%	6	13%	0	0%	0	0%	16	7%
Other (f)	1	1%	1	2%	0	0%	17	39%	19	9%
Strategies/Project alternatives(g)	4	4%	0	0%	0	0%	6	14%	10	5%
Locations / Relocations	0	0%	0	0%	0	0%	8	18%	8	4%
Capital investment	1	1%	0	0%	0	0%	8	18%	9	4%
Population groups	8	7%	0	0%	0	0%	0	0%	8	4%
Integrated care programmes	3	3%	0	0%	0	0%	0	0%	3	1%
Not specified	1	1%	0	0%	0	0%	0	0%	1	0%

Notes: (a) Four studies covered different types of economies, and were not classified. (b) One or more cities or counties within a country. (c) e.g. hospital or healthcare facility. (d) Health Technology Assessment (HTA) bodies

(e) e.g. WHO (f) e.g. Ingredients for pharmaceuticals developments. (g) e.g. Areas for efficiency gains within a hospital.

3.3.3. Criteria

Different sources have been used to identify and define the MCDA criteria. Almost half of the MCDA studies have chosen criteria based on literature (n=103). From these, 50% (n=52) have solely used literature as a source, while the rest (n=51) have additionally used other sources (e.g., specialists or experts in the field). A quarter of the studies (n=56) have chosen the criteria based on expert opinion, of which 15% (n=32) used literature as well. More than 20% (n=48) have adopted an established set of criteria (published or not published), most frequently from the Evidence and Value: Impact on DEcision-Making (EVIDEM) framework (7%, n=16) (191).

Less than 35% (n=76) have considered stakeholders' views in the process of defining the criteria set. This includes providers of medical care (13%, n=28), health authorities (8%, n=18), patients (6%, n=13), the private sector (3%, n=7), social care authorities (2%, n=5), government entities for social care services (2%, n=4), and NGOs (0%, n=1).

In terms of the ways to agree on the criteria, approximately 37% (n=80) of the studies have resorted to group discussions, 18% (n=39) have relied on experts' opinions, and 11% (n=23) have based the final selection of criteria on related literature. A considerable number of studies, however, have failed to report the method adopted to agree on the criteria (39%, n=86).

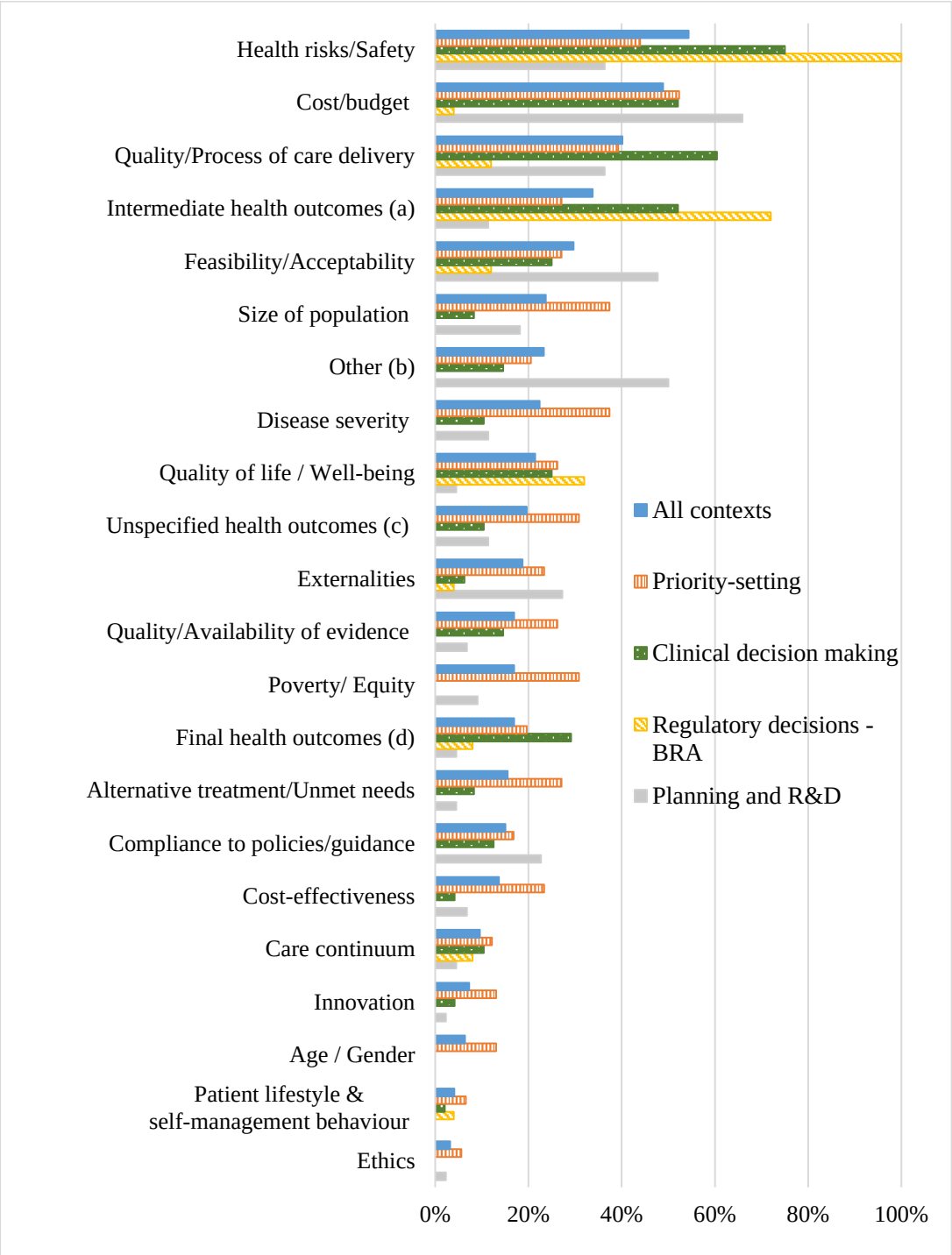
When structuring the criteria, 41% (n=89) of studies have adopted a hierarchical structure. On average, 4.8 criteria (SD=2.1, min=2.0, max=18) and 13.7 sub-criteria (SD=8.5, min=3, max=45) have been included in MCDA studies in healthcare that used a hierarchical structure. Studies that adopt a non-hierarchical structure have used on average 7.1 criteria (SD=3.2, min=2, max=18).

As presented in **Figure 3.3.a**, several criteria have been used to inform decisions in healthcare, with 'health risks/safety' (54%, n=119), 'cost' (49%, n=107), 'quality of care delivery' (40%, n=88), 'intermediate health outcomes' (34%, n=74) and 'feasibility or acceptability' (30%, n=65) as the five most frequently used. While in regulatory decisions more than 70% of the studies have used 'health risks/safety' (100%, n=25) and 'intermediate health outcomes' criteria, in priority-setting a wider variety of criteria have been used, with 'cost' (52%, n=56) and 'safety' (44%, n=47) at the front. Studies informing clinical decisions have mainly used 'safety' (75%, n=36) and 'quality or process of care delivery' (60%, n=29), while in the context of planning and R&D, 'cost' (66%, n=29) and 'other' (50%, n=22) are the two most predominant criteria. If the number of studies that have used intermediate, final and unspecified health outcomes as criteria are aggregated, 'health outcomes' becomes the most common criterion in all decision-contexts, but planning and R&D.

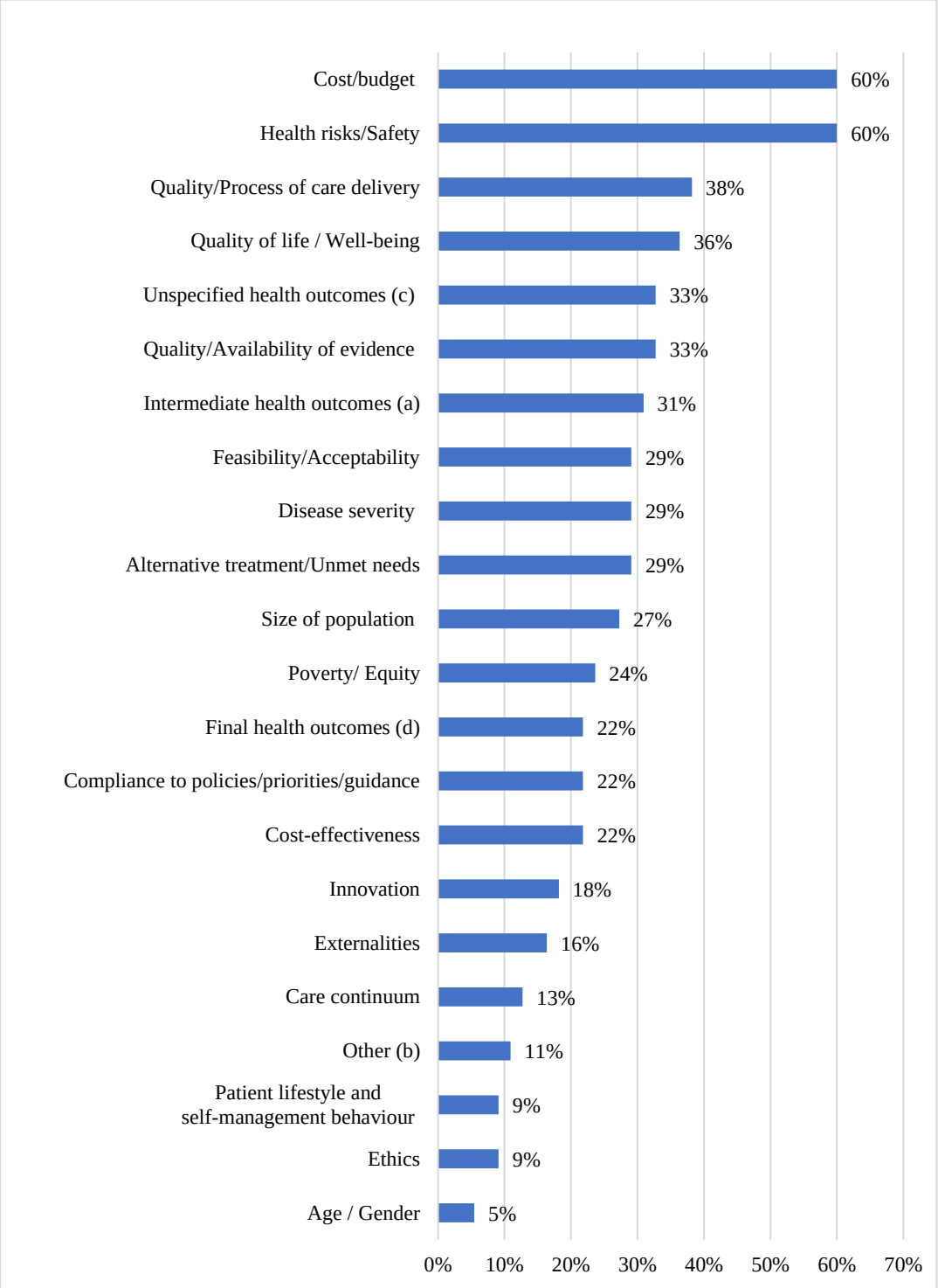
As the decision-support framework that this thesis develops aims to inform priority-setting decisions in England, **Figure 3.3.b** summaries the criteria used among studies conducted in high-income countries, aiming at prioritising interventions in healthcare (24%; n=55). The five most frequently used criteria have been 'cost' (60%, n=33), 'health risks/safety' (60%, n=33), 'quality of care delivery' (38%; n=21), 'quality of life' (36%; n=20), and 'unspecified health outcomes' (33%; n=18). **Appendix 3.D** provides more detail on the 10 most used criteria by these type of studies.

Figure 3.3. Criteria used by decision context

a. Criteria used across decision contexts



b. Criteria used by studies informing prioritisation of interventions in high-income countries



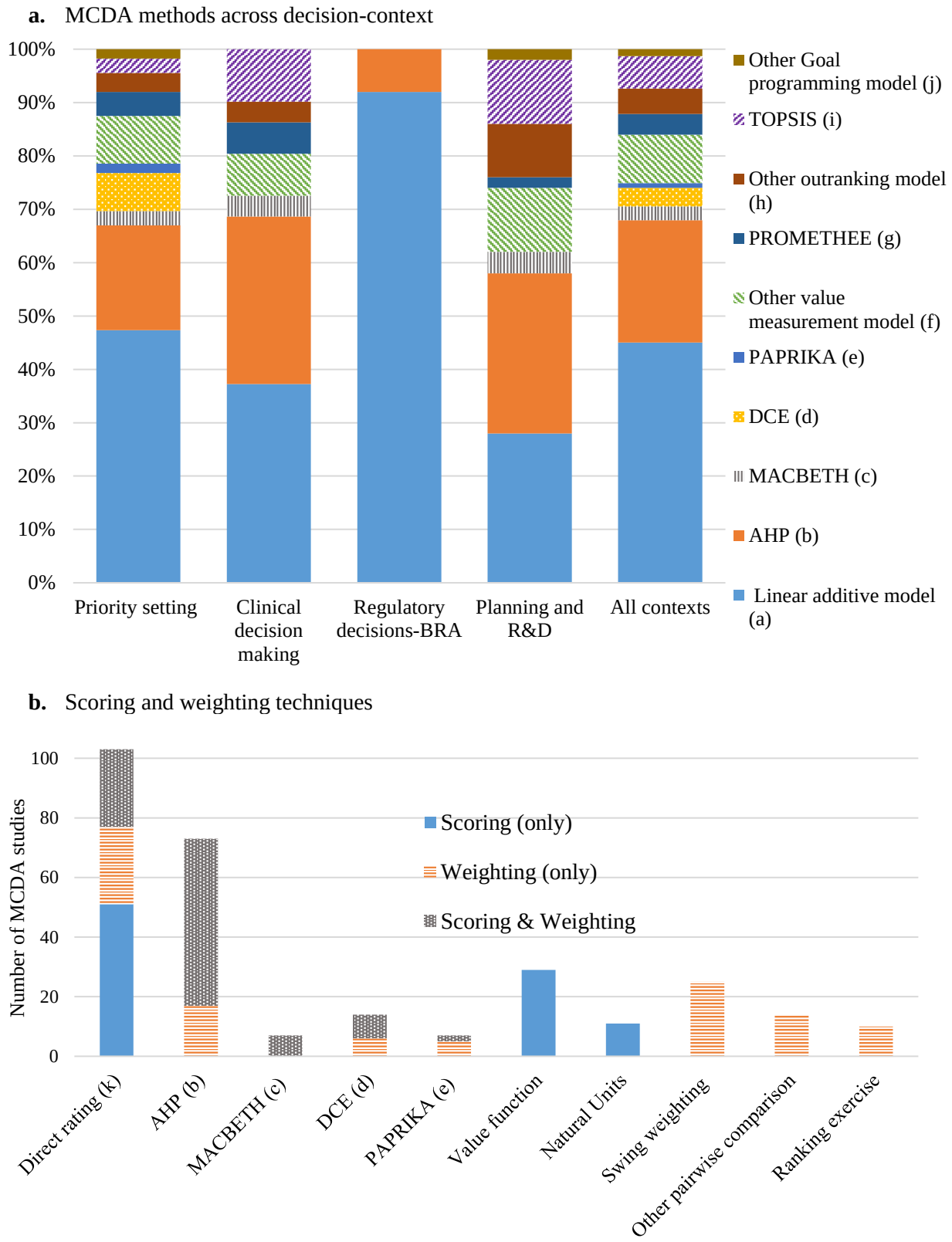
Notes: BRA indicates benefit-risk assessment; HbA1c, haemoglobin A1c; R&D, research and development. (a) Biomedical, physiological, and clinical health outcomes, for example, "reduction of HbA1c by 0.5%"; (b) Other criteria that were not classified (e.g., "Total distance to the nearest hospital": "The type of medical insurance and its place of belonging"; or criteria called as "Others" in the studies directly). (c) Unspecified health outcomes, such as "Improvement of efficacy/ effectiveness"; (d) Final health outcomes, such as mortality (for example, death from pulmonary tuberculosis) or life expectancy gains.

3.3.4. MCDA approaches

As presented in **Figure 3.4** and **Table 3.4**, the vast majority of studies have adopted a value-measured approach (192) (86%, n=189), with linear additive models (e.g. multi-attribute value theory (MAVT), weighted sum) and analytic hierarchy process (AHP) (193) as the most common approaches adopted. Only 21 studies (10%) have used outranking models (47), such as PROMETHEE (194), and 15 (7%) studies adopted a goal programming approach (47), with TOPSIS (195) being the most popular one. A similar pattern is observed across studies informing prioritisation decisions and treatment or screening decisions. Almost all studies that inform regulatory decisions via BRA have adopted the linear additive model (91%, n=21). These approaches were frequently applied using Microsoft Excel (29%, n=37) and Expert Choice (15%, n=19), although just 58% of the studies (n=127) have reported the software used. In the case of studies informing priority-setting decisions in high-income countries (24%; n=55), 51 studies have adopted value-measurement models (93%) and only four have used outranking models (7%).

More than 70% (n=156) of the studies retrieved have provided some justification for the MCDA approach applied, but few have justified the scoring and weighting technique adopted. Some of the reasons offered by these studies include the broad application of the method, the theoretical soundness, and the flexibility and relative ease to use.

Figure 3.4. MCDA methods and techniques applied



Notes: Value measurement models include: (a) Multi-attribute utility theory (MAUT), multi-attribute value theory (MAVT), weighted sum, and Simple Multi-Attribute Rating Technique methods (192). (b) AHP: Analytic

hierarchy process (193); (c). MACBETH: Measuring attractiveness through a categorical-based evaluation technique (196); (d) DCE: Discrete choice experiment (197); (e) PAPRIKA: Potentially all pairwise rankings of all possible alternatives (198) and (f) other value-measurement models (e.g. Socio-Technical Allocation of Resources). Outranking models include: (g) PROMETHEE: Preference ranking organization method for enrichment evaluations (194); (h) other outranking models (e.g. Elimination and choice expressing the reality). Goal programming models include: (i) TOPSIS: Technique for order preference by similarity to ideal solution (195); and (j) other goal programming models (e.g. VIKOR). (k) Direct rating includes scales and point allocation techniques.

Table 3.4. MCDA methods and techniques applied by decision context

	Priority-setting		**Clinical decision making**		**Regulatory decisions - BRA**		**Planning and R&D decisions**		**All contexts**	
	No	**%**	**No**	**%**	**No**	**%**	**No**	**%**	**No**	**%**
a. Type of MCDA employed										
Value measurement models	94	88%	38	79%	25	100%	37	84%	189	86%
Outranking models	10	9%	5	10%	0	0%	5	11%	21	10%
Reference-level models / Goal programming	4	4%	5	10%	0	0%	7	16%	15	7%
Other	0	0%	3	6%	0	0%	0	0%	3	1%
b. Methods to score interventions performance										
Scales	41	38%	11	23%	3	12%	11	25%	65	30%
AHP	21	20%	17	35%	2	8%	17	39%	56	26%
Points allocation	6	6%	1	2%	1	4%	4	9%	12	5%
Not reported	7	7%	5	10%	0	0%	5	11%	17	8%
DCE / conjoint analysis	8	7%	0	0%	0	0%	0	0%	8	4%
Value function	8	7%	6	13%	15	60%	1	2%	29	13%
Other	5	5%	5	10%	3	12%	2	5%	16	7%
Natural Units	5	5%	3	6%	1	4%	2	5%	11	5%
MACBETH	2	2%	2	4%	0	0%	3	7%	7	3%
PAPRIKA	2	2%	0	0%	0	0%	0	0%	2	1%
				0%		0%		0%		
c. Methods to elicit criteria weights										
AHP	31	29%	18	38%	4	16%	21	48%	73	33%

Scales	16	15%	10	21%	2	8%	5	11%	32	15%
DCE / conjoint analysis	12	11%	0	0%	2	8%	0	0%	14	6%
Not reported	6	6%	4	8%	2	8%	0	0%	11	5%
Swing weighting	6	6%	3	6%	14	56%	3	7%	25	11%
Points allocation	13	12%	2	4%	1	4%	4	9%	20	9%
PAPRIKA	5	5%	0	0%	0	0%	2	5%	7	3%
None	7	7%	0	0%	0	0%	0	0%	7	3%
MACBETH	2	2%	2	4%	0	0%	3	7%	7	3%
Other	7	7%	4	8%	0	0%	2	5%	13	6%
Ranking exercise	3	3%	5	10%	0	0%	2	5%	10	5%
Other pairwise comparison	4	4%	4	8%	0	0%	5	11%	14	6%

Notes: Value measurement models include: Multi-attribute utility theory (MAUT); multi-attribute value theory (MAVT); weighted sum, and Simple Multi-Attribute Rating Technique (SMART) methods; Analytic hierarchy process (AHP); Measuring attractiveness through a categorical-based evaluation technique (MACBETH); Discrete choice experiment (DCE); Potentially all pairwise rankings of all possible alternatives (PAPRIKA) and other value-measurement models (e.g. Socio-Technical Allocation of Resources). Outranking models include: Preference ranking organization method for enrichment evaluations (PROMETHEE); other outranking models (e.g. Elimination and choice expressing the reality). Goal programming models include: Technique for order preference by similarity to ideal solution (TOPSIS); and other goal programming models (e.g., VIKOR). Direct rating includes scales and point allocation techniques. Scales and point allocation are considered direct rating methods.

3.3.5. Performance scoring

Direct rating methods seem to be the preferred ones for scoring alternative's performance, with 30% (n=65) of the studies using direct scales (e.g. Likert scale), and 5% (n=12) using a point allocation system (see **Figure 3.4**). AHP (26%, n=56) and value function (13%, n=29) are the second and the third most frequently used methods, respectively. According to **Table 3.4,** scales are the preferred technique (38%, n=41) among prioritisation-focused studies, while in clinical decision-making and planning and R&D the majority of studies have used AHP (35%, n=17). Value function seems to be preferred when conducting BRA (60%, n=15). Among studies informing priority-setting decisions in high-income countries (24%; n=55), the preferred scoring approach has been direct rating methods (47%; n=26) followed by AHP (22%; n=12).

Regarding the sources to measure the performance of alternatives, 60% (n=131) of the studies have used evidence from the literature. Approximately 32% (n=69) of the MCDA studies have relied on experts' opinions, 16% (n=35) have used trial data, and 14% (n=30) have used routinely collected data. About 20% (n=41) of the studies have solely used research evidence, and 5% (n=12) have relied only on expert opinions as a source. Almost half of the studies (n=104) have used more than one source when measuring alternatives' performance, and over a fifth (n=50) have failed to report information on this MCDA stage. Although there are no considerable differences across decision-contexts, in the case of priority-setting studies conducted in high-income countries, almost 70% have relied on evidence from the literature (n=38) and 45% have resorted to expert's opinion (n=25). Only seven out of 55 have used routinely collected data (13%).

In terms of the stakeholders that participate in the scoring of alternatives (see **Figure 3.5**), most of the studies have involved experts (41%, n=90), healthcare providers (40%, n=88), representatives from health authorities (23%, n=50), patients or carers (15%, n=32). Very similar shares are observed in the case of studies conducted to inform the prioritisation of interventions in high-income countries (24%; n=55). Few studies involved the private sector (7%, n=15) or the general public (6%, n=14). In 98 studies (45%) more than one actor has participated in this MCDA step, and on average 69 individuals (median=12.5; SD=346.7, min=1, max=3,914) have been involved at this stage[14]. Roughly 40% (n=87) of the studies, however, failed to report the number of stakeholders involved in the scoring of alternatives.

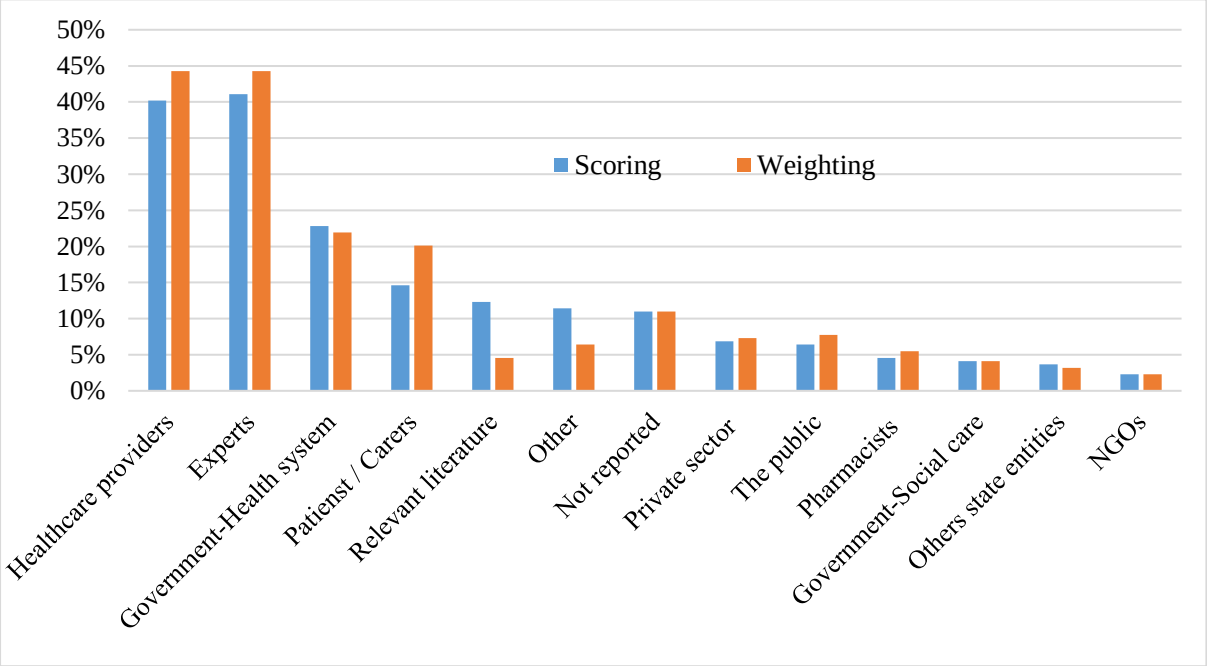
[14] Priority-setting MCDA studies have reported on average 47 individuals (median=16; SD=93.4, min=3, max=650), in clinical-decision making on average 38 individuals have taken part (median=10; SD=105.8, min=2, max=560), and studies that aim to inform regulatory decisions have involved fewer individuals, with an average of 10 (median=9; SD=5.1, min=6, max=20).

3.3.6. Criteria weighting

To elicit stakeholders' preferences between criteria, 33% (n=73) of the studies have used AHP, with 76% (n=56) of them using this technique also at the scoring stage (see **Figure 3.4.b**). Direct rating was also applied (24%, n=52), followed by swing-weighting (11%, n=25), DCE (6%, n=14) (199) and other pairwise comparison methods (6%, n=14). A similar pattern is observed across decision-contexts, except for studies informing 'regulatory decisions' in which swing-weighting has been used by most MCDA studies (56%, n=14), with relatively few using AHP (16%, n=4) (see **Table 3.4**). Among MCDA studies conducted in high-income countries aiming at informing priority-setting decisions (24%; n=55), AHP (24%; N=13), scales (16%; N=9) and DCE (9%; n=5) have been the preferred weighting methods.

Although virtually the same type of stakeholders were involved in the weighting and scoring steps (see **Figure 3.5**), a relatively larger number of individuals was on average involved in the weighting of the criteria (mean=118, median= 12.5, SD= 493.8, min=1, max=4,288).

Figure 3.5. Stakeholders involved in the scoring and weighting



3.3.7. Uncertainty analyses

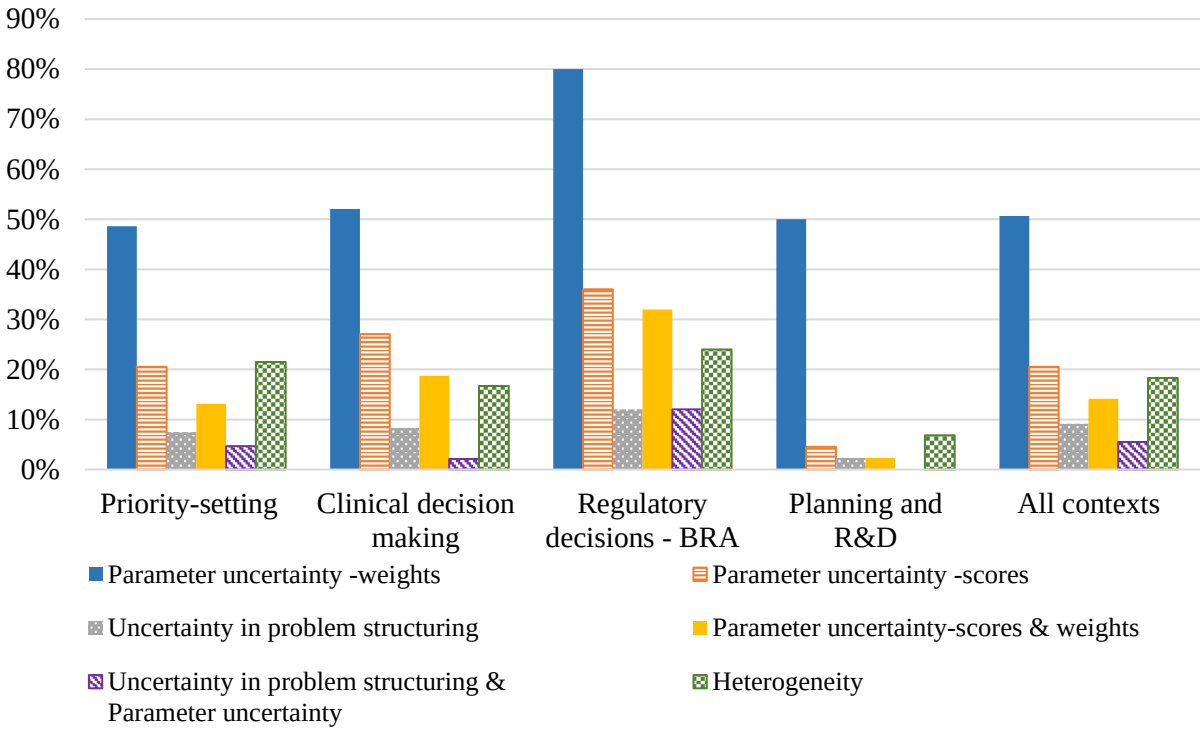
More than 65% (n=145) of the MCDA studies in healthcare have addressed one or more types of uncertainty (188). As illustrated in **Figure 3.6**, half of the MCDA studies (n=111) have addressed uncertainty in the weighting of the criteria, a fifth in the scoring step (n=45), and nearly 15% (n=31) have accounted for both types of parameter uncertainty. Heterogeneity in criteria weights or performance scores (i.e. variability attributed to a person's characteristics) was assessed in 18% of the MCDA studies (n=40). Dealing with structural uncertainty (e.g., MCDA method used, criteria included) is less common, with only 9% (n=20) of the studies reporting it. Across decision-contexts, the patterns are roughly similar. Studies informing planning and R&D decisions are however the exception, with less than 2% of these studies reporting structural uncertainty, and only 5% reporting parameter uncertainty on the performance scores.

The majority of studies that assessed uncertainty have used scenario analysis (29%, n=64) or a one-way deterministic analysis (23%, n=51). Fuzzy logic and probabilistic sensitivity analysis are less common, with only 11% (n=24) and 8% (n=18) of the studies using these methods, respectively.

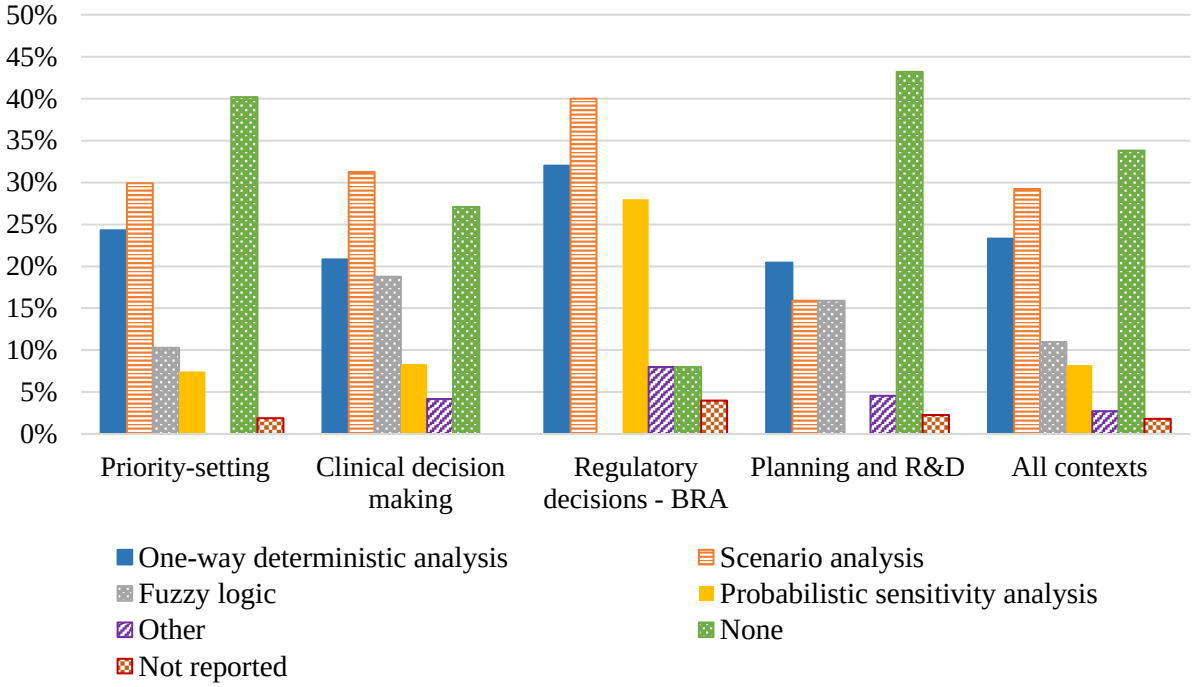
In the case of studies informing priority-setting decisions in high-income countries (24%; n=55), almost half (n=27) have addressed uncertainty in the weighting of the criteria and one-fifth have at the scoring stage (n=1). Only two studies have addressed both, uncertainty in problem structuring and parameter uncertainty. Among this group of studies, scenario analysis (31%; n=17) and one-way deterministic analysis (29%; n=16) have been the preferred methods to assess uncertainty.

Figure 3.6. Type of uncertainty and methods used to deal with it across decision context

a. Type of uncertainty



b. Methods used to deal with uncertainty



3.4. Discussion and Conclusion

The application of MCDA continues growing in healthcare, with most of the studies being conducted in high-income countries to inform priority-setting and regulatory decisions at national level. The fact that no MCDA has been conducted to inform regulatory decisions in low-middle income countries, despite the increasing use of MCDA in other decision-contexts, possibly reflects the embryonic stage of HTA in these countries (200).

As identified by Khan et al (2021) (42), it is just after 2010 that a rise in MCDA studies in healthcare took off. This upward trend might have been driven by the rise and consolidation of HTA agencies across Europe (201), an increasing concern from national and local governments to justify investment and authorisation decisions (202, 203), and the use of MCDA in emerging decision-contexts (e.g., pharmaceutical R&D). MCDA studies informing regulatory decisions emerged between 2009 and 2012, which coincides with the publication of the first report from the European Medicines Agency in 2010, acknowledging MCDA as one of the quantitative approaches to conduct BRA for medicinal products (204). In the United States, and later on across Latin America and the Caribbean, the industry might have played a role in promoting the use of MCDA when conducting BRA of medicines (205-207).

As expected, considerable differences were found across decision-contexts on the criteria used, with some noticeable preferences for specific criteria sets, particularly among studies that have informed regulatory decisions. There is, however, significant within-context variation in the type of criteria used, which does not facilitate study comparisons and can potentially affect the consistency of MCDA studies (162). Defining a set of top-level criteria clusters for each one of the different decision-contexts could contribute to the comparability of MCDA studies and help decision-makers consider all relevant aspects when agreeing on the criteria. This could reduce subjectivity and variation in criteria sets within decision-contexts and improve the MCDA's credibility in informing healthcare decisions. An example

of this is the advanced value tree with five key value domains proposed by Angelis and Kanavos for evaluating new medicines in HTA (208).

With regards to the number of criteria used, more than a third of the studies have used more than five criteria and in some cases more than 10 criteria. However, this appears to be not in line with good practice and MCDA guidelines that recommend the inclusion of only few criteria (48) This increases the risk of violating the non-overlap and preferential independence requirements of linear additive models undermining the usefulness of the most used MCDA method (209, 210). A large number of criteria also increases data requirements, making the application of MCDA more complex and compromising the quality of the results. While flexibility in the criteria selection step might be desirable in some settings (48), a uniform set of criteria by decision-context that considers all axiomatic properties (211) could serve as a starting point for future applications of MCDA in healthcare. This could potentially improve the consistency and credibility of the resulting recommendations (149, 162, 187).

Results of this chapter highlight a preference of MCDA users in applying the additive model for calculating total scores (i.e., MCDA approach) and hybrid techniques (i.e., scales for scoring and AHP for weights elicitation) in the scoring and weighting steps. Nevertheless, not all studies made a clear distinction between the MCDA approach used, and the scoring and weighting techniques applied, and very few justified the selection of the techniques used. Most of the studies have also favoured scoring and weighting techniques that are relatively easier to apply (e.g., direct rating approaches and AHP) but less theoretically sound (196). This might partly reflect the little guidance on the choice and application of MCDA methods. As highlighted previously, the lack of standardization of the terms and classifications of MCDA methods, may partially explain the flaws in reporting MCDA studies (42, 212). Khan et al (2021) identified at least four different classifications of MCDA models or approaches (42). Different nomenclatures and categorizations have also been used to summarise the scoring and weighting techniques applied. For instance, the review conducted by Marsh et al., (2014) (39) reported scales, AHP, natural units, and DCE as weighting techniques, while Oliveira and colleagues (2019)

reported point systems, direct rating, MACBETH, AHP, and selecting functions (212). This lack of consensus affects the comparability, credibility and policy-usefulness of MCDA studies in healthcare (162, 212).

Although various types of uncertainty can arise when conducting MCDA in healthcare (188, 213-215), and all guidelines recommend addressing them (47, 48), uncertainty analysis is not yet universally and consistently applied across MCDA studies. This is in contrast to traditional economic evaluations, where uncertainty is typically addressed in a systematic manner (3, 216). However, there is an increasing trend in the incorporation of uncertainty analysis in MCDA studies, indicating a growing response from MCDA designers and analysts to the calls for addressing uncertainty (47, 48, 188). In terms of the approaches adopted to deal with uncertainty, there seems to be a preference for one-way deterministic approach or scenario analysis, which is consistent with findings from Broekhuizen and colleagues (2015) (188). Structural uncertainty is barely explored, which might be due to the lack of methodological approaches to systematically deal with it (188), or the cost of re-running analyses under different MCDA specifications (e.g., another set of alternatives and/or criteria).

The possibility of involving all relevant parties in the decision-making process is one of the advantages that MCDA offers (203, 217), and this review found that different type of stakeholders have participated in the scoring and weighting steps. More than 50% of the studies, however, were not explicit about the stakeholders involved in each one of the MCDA stages. Additionally, several studies have highlighted limitations in regard to stakeholders' involvement. This includes the small number of participants (218-220) and the exclusion of relevant actors, such as patient representatives or caregivers (222-224). As big groups of participants might not be possible, some studies have suggested involving representatives of the relevant stakeholders when conducting an MCDA (225). Other studies also emphasise the need to include all relevant stakeholders from the beginning to ensure the relevance and usefulness of the MCDA (226, 227).

All flaws in the application and reporting across all MCDA steps described above compromise the transparency and legitimacy of the MCDA-based recommendations (162). The lack of standardization of MCDA methods and their classification jeopardizes the comparability, credibility and policy-usefulness of MCDA studies in healthcare (162, 212). This reinforces the need to clearly report the methods applied in each MCDA step. Even though some reports have been published to guide the application of MCDA in healthcare (47, 48), the lack of a detailed methodological and reporting guidance might partly explain the questionable methodological quality of MCDA studies in healthcare (212). This guide could include: (i) a universal classification of the MCDA methods; (ii) a guidance on the application of each method, highlighting the particularities for each one of the different decision-contexts in healthcare; (iii) a common set of criteria or macro domains by decision-context; (iv) a checklist with standards for reporting, similar to the Consolidated Health Economic Evaluation Reporting Standards statement (228) used in traditional economic evaluations. The development of this methodological and reporting guide should be based on empirical work and the experience accumulated by scholars and decision-makers in healthcare. Such a guidance could increase transparency and clarity in the application of MCDA, and ultimately improve the credibility and systematic application of MCDA in different decision-contexts in healthcare.

The results of this updated systematic literature review will inform the development of the MCDA component of the decision-support framework that this thesis develops for the assessment of new models of care. In particular, the list of the most frequently used criteria in healthcare priority-setting will be used to define the set of assessment criteria that should guide the evaluation of new models of care in England. In this context, two important considerations arise. First, while this updated systematic review provides valuable insights into defining MCDA assessment criteria, a comprehensive review of the value-based healthcare literature could have further enriched this research. Second, it is important to recognise that undertaking MCDA at the local level requires significant resources and support. This includes access to data, expertise in MCDA methods and data analysis, stakeholder engagement, clarity on the integration of MCDA results into decision-making processes, as well as a commitment to

evaluation and learning (47, 48). Many of these needs, particularly those related to data access, mirror the obstacles currently faced by local commissioners in the country, as discussed in **Chapter 2**. Nevertheless, the increasing adoption of a standardized MCDA framework in England, such as the one proposed in this thesis, has the potential to drive the collection of relevant data within NHS services for MCDA analyses. This could help to ensure that local commissioners have the necessary data inputs readily available to inform the local commissioning of new models of care.

Strengths & limitations

This is the first review that includes grey literature and summarises key methodological information by decision-context. Such information could guide researchers in designing MCDA studies by supporting the selection of appropriate methods to their decision-context. Additionally, this review serves as the initial effort in consolidating the criteria utilized in healthcare-related MCDA studies. Findings could be used as a basis to start developing a uniform set of top-level criteria by decision-context. In the context of integrated care, the criteria employed to inform priority-setting decisions in high-income countries will be used to identify the monitoring and assessment criteria to support the local commissioning of new models of care in England (see **Chapter 4**). In light of the numerous MCDA methods available (217) and the criticism of lack of transparency (162), findings from this review could also be used as a basis to start developing a detailed methodological and reporting guidance by decision-context.

This literature review is not exempt from limitations. First, the review may have overlooked non-English MCDA studies, particularly grey literature, that have informed healthcare decisions (39). Second, all searches were conducted by one reviewer, which might have affected the identification of relevant studies. Third, the inclusion criteria adopted were less restrictive compared with the ones in Marsh et al (2014) (39), posing a risk of omitting relevant scientific studies prior to 2013. To mitigate this, Marsh et al.'s studies were supplemented with two other reviews that began as early as in the 90s

and shared a similar scope (40, 41). The time-trend of MCDA studies published in recent years aligns with findings from Khan et al. (2021) (42) suggesting that the observed sharp increase from 2011 onwards is not an artefact of the chosen updating strategy (i.e., starting academic searches in 2013).

Conclusion

In light of the decision-framework that this thesis develops, finding from this systematic literature review will inform the MCDA component for assessing new models of care at local level in England (see **Chapter 4**). The results obtained from this review could also guide future MCDA applications in healthcare. Nevertheless, to support the usefulness and applicability of MCDA in healthcare, it is crucial to standardize the methods and reporting. The lack of quality and transparency are an obstacle to the adoption of MCDA widely in healthcare. A guide on how to conduct and report MCDA, that acknowledges the particularities of the different decision contexts and methods, needs to be developed.

Chapter 4. A decision-support framework to monitor and evaluate new models of care in England

4.1. Introduction

Integrated Care Systems (ICSs) in England are responsible for the commissioning of healthcare interventions at a local level. The commissioning process involves an assessment of local population needs, health service planning, priority setting, and the monitoring of the provided healthcare interventions (51). As explained in **Chapter 2**, the commissioning is an ongoing process with regular reviews and updates in line with changing needs and emerging evidence.

When commissioning healthcare interventions, including new models of care, local commissioners need to respond to the objectives of the ICSs: to improve population health and health outcomes; to tackle inequalities in outcomes, experience and access; to enhance productivity and increase value for money; and to help the NHS support wider social and economic development (7, 12). These objectives, which go beyond gains in longevity and quality of life, are consistent with the value-based healthcare approach to which health systems around the world are being transformed in response to increasing healthcare demands, rising costs of care, and tighter fiscal policies (20, 21, 54)[15].

In order to achieve these objectives, ICSs need to adopt a robust systematic commissioning process of healthcare interventions. As indicated in **Chapter 2**, local commissioners are interested in data-driven commissioning, which requires the definition of a clear and explicit set of criteria to inform local investment decisions. Furthermore, NHS England has emphasized the importance of evidence-based

[15] Although there is no broadly agreed definition of value-based healthcare, it could be described as "the equitable, sustainable and transparent use of the available resources to achieve better outcomes and experiences for every person" (21).

commissioning, which involves making decisions based on reliable evidence while recognising public preferences, to deliver effective and efficient healthcare interventions (7, 24). As noted in **Chapter 2** and in other studies, local commissioners lack sufficient capacity to use evidence for informed decision-making, and there is a lack of a structured prioritisation framework (28, 29, 153, 229). The evaluation of healthcare interventions is not systematic, fails to consider public preferences, and often presents a lack of methodological rigour.

In response, this chapter aims to develop the structure of a decision-support framework for the systematic monitoring and evaluation of healthcare interventions, including new models of care. Such a framework should consider the objectives pursued by ICSs and incorporate both evidence and public preferences for attributes or elements of value-based healthcare, to support the monitoring and evaluation stages of the local healthcare commissioning cycle. The chapter starts by outlining the principles that underpin the decision-support framework. Next, the assessment and the monitoring components of the framework are defined. Subsequently, the chapter describes how both components operate within the commissioning cycle. Lastly, the chapter closes with a discussion and conclusion remarks.

4.2. The underpinnings of the decision-support framework

To ensure that the decision-support framework is relevant and effectively supports the local commissioning of new models of care, it needs to be tailored to meet commissioners' needs and respond to the specificities of their local decision context. To start, it is vital that the framework recognises the particularities of the monitoring and evaluating healthcare interventions, including new models of care. Both stages are integral parts of the local commissioning cycle and are designed to address specific questions. Monitoring tracks the implementation of an intervention and ensures that it is proceeding as planned, whereas evaluation seeks to provide a comprehensive understanding of the intervention's

effect or implications by analysing the intervention's outcomes or impacts. The timing, information requirements, and methodological approaches for monitoring and evaluation differ (230).

A second consideration is that the framework should respond to the challenge of evaluating new models of care. As indicated in previous chapters, these models have been regarded as complex interventions and their outcomes extend beyond quality-adjusted life years (QALYs) (21, 23, 150). The value-based healthcare approach that ICSs adopt suggest that the drivers of local investment decisions should not be and are not limited to improving healthcare related outcomes. Thus, the evaluation component of the framework should adopt an analytical method that allow for the incorporation of a comprehensive set of assessment criteria, to capture the multifaceted impacts of these models and align them with the fundamental principles of value-based healthcare.

A third consideration regarding the framework's relevance relates to the need to integrate public preferences into the decision-making process. Commissioners are expected to act on behalf of the public, as they are the taxpayer and the ones who bear the costs of healthcare decisions (6, 231)[16]. It is therefore desirable to define the elements of value-based healthcare, which are the outcomes of interest, in the light of the objectives pursued by ICSs, and to assess their relative importance from the perspective of the public. Consequently, the framework should adopt an economic evaluation approach that considers the multiple elements of value-based healthcare and the preferences of the public, drawing on the voices of relevant stakeholders.

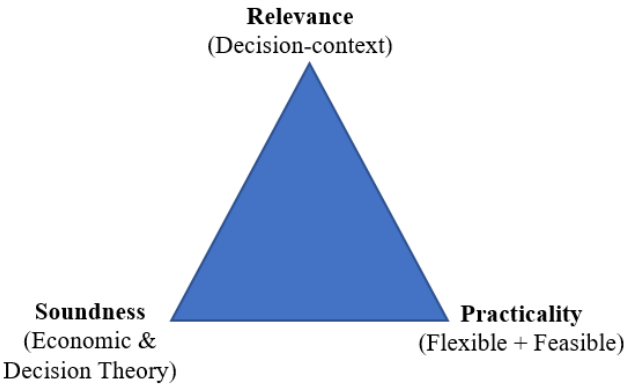
[16] Public and patient involvement (PPI) and engagement is mandated by many health research funders, including the UK's National Institute for Health Research (NIHR). Involving the public in decisions about public resources promotes democracy, and it could be argued that incorporating public preferences into decision-making can lead to more intelligent choices and improved social welfare outcomes.

A framework that is theoretically sound is desirable for informing the commissioning process of new models of care in a robust and reliable manner. The framework should adhere to the theoretical foundations of decision theory and economic theory while recognising the need to consider the multiple value-based elements that should guide local investment decisions (165). Decision theory provides a structured approach to making decisions under different conditions, including multiple objectives (161, 232). Economic theory provides insights into how scarce resources are allocated and how decision-makers choose options of allocation (3). A framework grounded in the established principles and methods of economic and decision theory enhances the credibility and effectiveness of the decision support framework.

Finally, the decision-support framework can only serve its purpose if it can be used effectively and systematically by local commissioners. To ensure the practicality of the framework, it must be based on information that is objective and accessible to commissioners, and flexible enough to be used to monitor and evaluate different types of healthcare interventions, including new models of care. Therefore, the assessment and monitoring criteria proposed by the decision-support framework should be flexible enough to be operationalised for different conditions or types of interventions, making use of routinely collected and readily available data.

Figure 4.1 summarises the three principles on which the decision-support framework is developed, for the local commissioning of new care models in England. The following sections are structured according to the components of the decision-support framework.

Figure 4.1. Principles of the decision-support framework



4.3. The evaluation component of the framework

The evaluation component of the decision-support framework relies on multi-criteria decision analysis (MCDA), a comprehensive and theory-based approach for conducting a comparative evaluation of competing investment alternatives considering multiple assessment criteria (47, 48, 192). In MCDA, healthcare interventions are assessed for their performance on all outcomes of interest, weighted according to the preferences of the public, and ranked based on an overall value. MCDA has been widely applied in the context of priority-setting in healthcare (see **Chapter 3**), and the Medical Research Council in the United Kingdom has suggested it as a methodology suitable for economic evaluations of complex interventions (38, 150).

To develop the MCDA for the decision-support framework, a value-based Multi-Attribute Value Theory (MAVT) approach is used. This approach, grounded in decision theory (233)[17], involves

[17] MAVT adheres to the fundamental axioms of value theory: transitivity and comparability (or completeness). Transitivity ensures that if option A is preferred to option B, and option B is preferred to option C, then option A is also preferred to option C. This helps maintain consistency in decision-making. Comparability, demands that, given any two options A and B, a decision-maker should be able to indicate which option is preferred or whether she or he is indifferent between the two options. It hence facilitates the direct comparison of different options based on their assigned values, enabling informed and rational choices. These axioms, integral to economic utility theory, provide the foundational logic for the existence of a real-valued function that can represent an individual's or a group's preferences (233).

identifying, structuring and quantifying various attributes or criteria relevant to a decision-making context and then weighting these criteria based on their relative importance (211, 234). MAVT is a simplification of Multi-Attribute Utility Theory (MAUT)[18] (233). While MAUT makes use of utility functions that account for decision-makers' attitudes toward risk, using the concept of lotteries, under MAVT a global value function is constructed for each alternative under evaluation to represent the global performance of the alternatives on the assessment criteria. This simplification makes MAVT more accessible and easier to implement in real-world scenarios (ibid). Although no MCDA model is without critics, models based on MAVT are the most widely accepted, and its simplicity and structured approach makes it particularly pertinent for the local evaluation of healthcare interventions, including new models of care (48, 211).

As indicated in **Chapter 3**, the development of the MCDA can be summarized in eight steps (48), beginning with the identification of the decision problem and the healthcare interventions to be compared (step 1). Once the decision goal has been set, the assessment criteria are identified along with the corresponding performance indicators (steps 2 and 3). Afterwards, performance indicators are converted onto a common scale or score (step 4) and weighting techniques are applied to elicit people' preferences between criteria (step 5). Performance values and criteria weights are aggregated in a MCDA total score (step 6), sensitivity analyses are conducted (step 7) and results are reported and examined (step 8). The subsequent subsections provide a detailed explanation of each MCDA step.

[18] MAUT is an extension of the von Neumann-Morgenstern expected utility theory (211, 235). In their work, the authors established a set of fundamental axioms of rational choice, and then mathematically demonstrated that the only way a person could behave consistently with the full set of axioms is by choosing the option which possessed the maximum subjective expected utility value (ibid). Keeney and Raiffa's work in 1976 built upon the foundational concepts established by von Neumann and Morgenstern, by introducing a method to derive a single valuation index, U, to express the decision maker's overall valuation of an alternative (e.g., healthcare intervention) in terms of the value of its performance on multiple and often conflicting assessment criteria (i.e., multi-attribute utilities). With their work, the authors demonstrated that an additive model can be a straightforward approximation to the subjective expected utility model, preserving the axiomatic foundations of expected utility theory (161, 236).

4.3.1. Decision-problem and alternatives

Local commissioners are concerned with purchasing a range of services that span across the entire care continuum, with a focus on integrating care. This includes primary care services, community health services, hospital services, mental health services, public health initiatives, urgent and emergency care, and specialized services (7). Nevertheless, the budget to respond to all health population needs is limited and, consequently, commissioners must determine which healthcare interventions should be prioritized.

Local commissioners have several healthcare service alternatives available for purchase. These include single interventions or simple services, like diagnostic tests and surgical procedures, which are often described as 'care as usual' or 'standard practice'. Additionally, commissioners can purchase new models of care, also known as integrated care programmes. These models consist of diverse interacting elements, involving multiple teams or health professionals that target groups, settings or levels rather than, or in addition to individual patients (23, 150). As explained in **Chapter 2**, these new models promote integration across the healthcare continuum (spanning from preventive care to palliative or end-of-life care[19]) or across multiple health conditions, and advocate for a more personalised and coordinated provision of care.

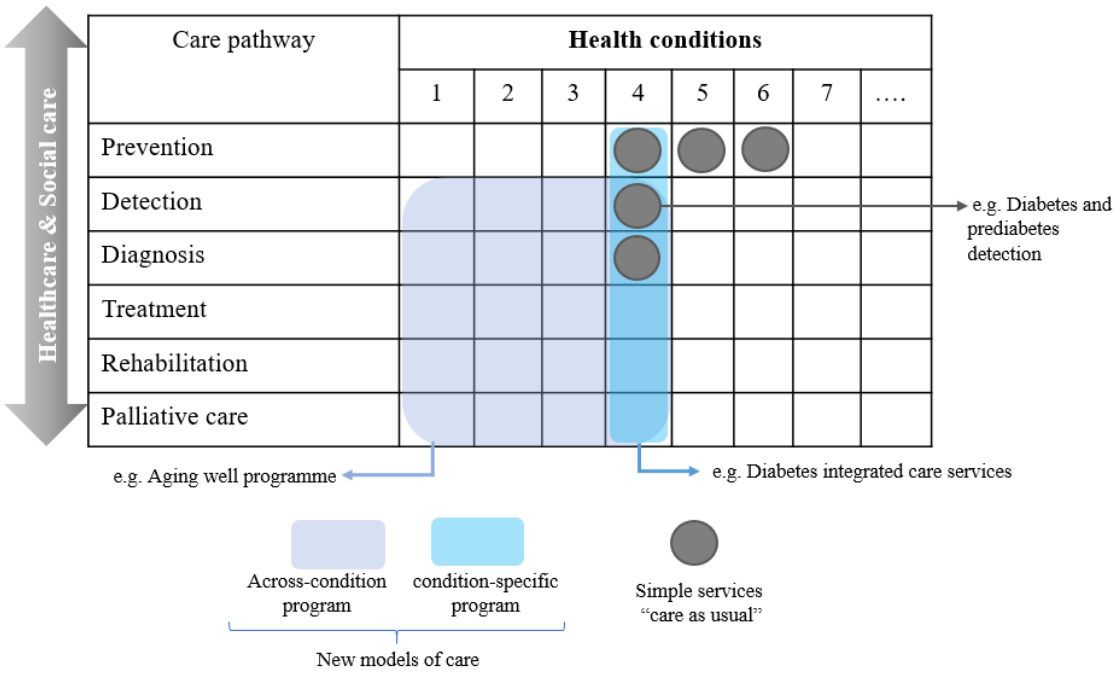
During the interviews conducted in 2021 (see **Chapter 2**), local stakeholders identified two types of models of care that in practice tend to overlap. On the one hand, there are "condition-specific programmes" with elements of integration, such as the 'Integrated diabetes care programme' (237) or 'Early Intervention in Psychosis (EIP) services' (238), which target patients with a particular condition. On the other hand, there exist "across-condition programmes" which target multiple conditions and might intersect with condition-specific programmes. Some of these programmes include the 'Aging

[19] The care continuum or healthcare pathway includes prevention (preventive care), detection, diagnosis, treatment, and palliative or end-of-life care.

well programme', which looks at individuals with multiple long-term conditions, or the 'Urgent community response programme', in which health and social care work together (19). Both programmes incorporate vertical and horizontal integration elements (e.g., integration between general practitioner (GP) practices in primary care networks), although vertical integration seemed to be more common in the context of condition-specific programmes, where primary and secondary care are collaborating closely together (229).

It is therefore possible to differentiate between three types of alternatives, as described in **Figure 4.2**: single or simple services, condition-specific programmes and across-condition programme.

Figure 4.2. The alternatives: Type of healthcare interventions to commission



The mutually exclusive alternatives from which local commissioners have to choose depend on the investment decision to be made, and the local healthcare context. An integrated care board of an ICS may be presented with a new care model, either condition-specific or an across-condition programme, which has been piloted within one of the place-based partnerships. This new model is presented to

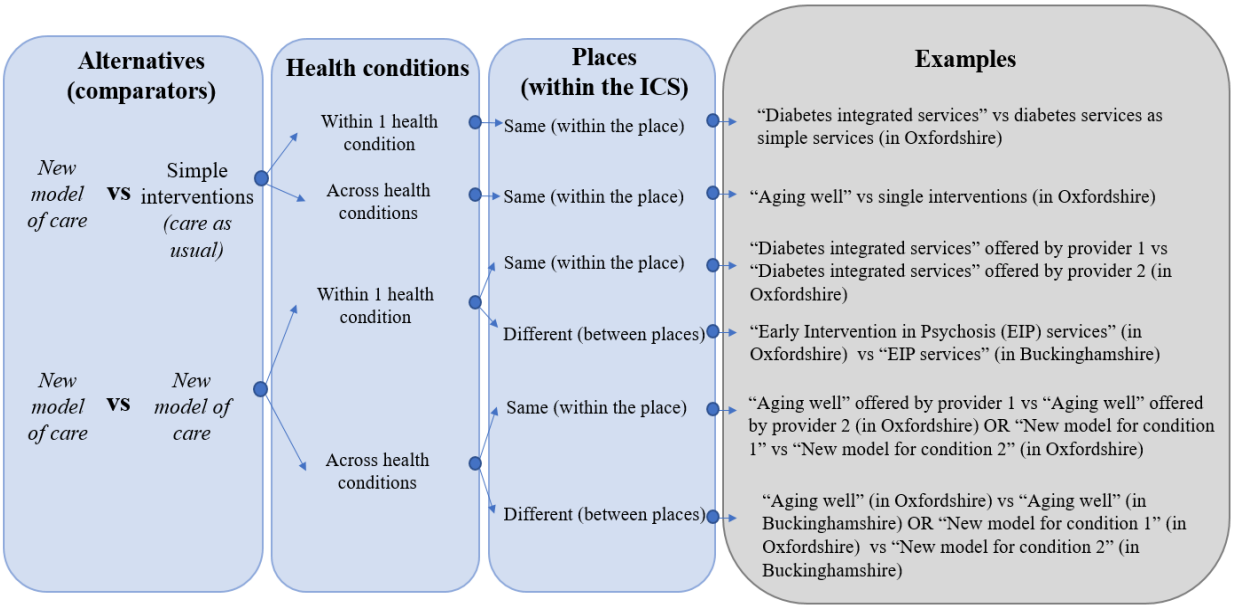
replace the current 'care as usual' offered in the ICS's catchment area[20]. Under this scenario, local commissioners would be interested in assessing the performance of the new care model compared to the current standard approach within the same place. Alternatively, commissioners might be interested in comparing two distinct new models of care that have been tested by partnerships situated in different areas to determine which model to expand across the ICS. Such an evaluation would entail assessing two new models operating in different locations by different provider collaboratives, within the ICS. The comparison of two new models of care could also take place within the same place, as local commissioners might be interested in assessing the performance of two competing services.

Comparing two different condition-specific programmes, such as diabetes integrated care services and EIP services, might not be commonplace. This is because ICS budgets have traditionally been aligned with specific funding streams set by the NHS, covering areas like maternity, community services, and mental health (see **Chapter 2**). However, with the introduction of the Health and Care Act 2022, ICSs now possess greater flexibility. As new models of care continue to emerge, ICS boards may find themselves evaluating competing care models - proposed as business cases by various place-based partnerships - aimed at tackling different conditions. This can also be the case of specialised services to support people with a range of rare and complex conditions, for which local commissioners have now greater budget allocation flexibility.

Figure 4.3 offers an overview of potential comparators that may undergo economic evaluation, along with a few examples.

[20] As explained in **Chapter 2**, ICSs operate within "places" (typically corresponding to local authority areas, covering populations of around 250,000 to 500,000 people) and further within "neighbourhoods" (serving smaller populations of around 30,000 to 50,000 people) (8).

Figure 4.3. Potential comparators for an economic evaluation



4.3.2. Assessment criteria

From a logic model approach, commonly used in healthcare policy contexts, evaluating an intervention requires rigorous and scientific examination of data regarding expected outcomes and its impact, while taking into consideration attribution and counterfactual scenarios (230). The evaluation, therefore, concentrates not only on the medium term results (i.e., outcome indicators) but also on the lasting impact and ultimate goals of the health intervention (i.e., impact indicators) (239). As noted in **Section 4.2**, the decision-support framework must account for the multiple elements of value-based healthcare that ICSs aim to achieve. Therefore, following the logical model approach, these elements of value - which represent the ultimate objectives of ICSs - should to be translated into outcome or impact indicators, ensuring that the framework thoroughly captures and evaluates the full spectrum of desired outcomes.

To identify the elements of value, henceforth referred to as 'assessment criteria', for evaluating new care models, a stepwise approach was undertaken. First, a comprehensive list of potential assessment criteria was compiled based on two sets of data. The first set comprises results from 26 semi-structured interviews, conducted in 2021, asking local decision-makers to indicate ideal criteria for guiding the

local investment decision process. A summary of these interviews is presented in **Chapter 2** (229). The second set is based on results of a systematic literature review of empirical studies that have applied MCDA to prioritize healthcare interventions in high-income countries. Information about the methodology and results obtained is summarised in **Chapter 3** (see **Figure 3b**) (240). The full list included 28 potential assessment criteria (seven from the semi-structured interviews and 21 from the systematic literature review).

Second, criteria were assessed against theoretical properties of multi-composite value metrics founded in decision theory (192, 211, 232). These properties included:

(a) *Completeness*: To start, all potentially relevant criteria for local stakeholders were considered, in addition to all the different criteria that have been used by MCDA studies aiming to inform priority-setting decisions in high-income countries.

(b) *Non-redundancy*: The focus was placed on criteria that could be translated into outcome or impact metrics, as those are the one of interest when evaluating new care models. Accordingly, criteria that were considered unimportant or irrelevant to the evaluation of new models of care at local level in England were excluded;

(c) *Non-overlap*: Criteria that may overlap with previously selected assessment criteria were excluded, aiming to avoid double counting; and

(d) *Operationality:* Assessment criteria that can be objectively assessed using data available to local decision-makers were selected.

The resulted assessment criteria were six:

(i) *Final or intermediate health outcomes*: This includes outcomes such as mortality or life expectancy. When final health outcomes are not possible, intermediate health outcomes such as biomedical indicators could be considered.

(ii) *Quality of life and well-being considerations*: This covers elements of a patient's functioning and well-being (e.g., mobility, self-care, pain/discomfort) that the intervention is anticipated to affect[21].

(iii) *Patient experience*: This criterion is about how the patient, the family and/or the carers feel about the process of receiving care.

(iv) *Equity*: This criterion aims to capture the gap in terms of health and wellbeing outcomes, or in terms of patient experience (across geographical areas, between men and women, of special groups).

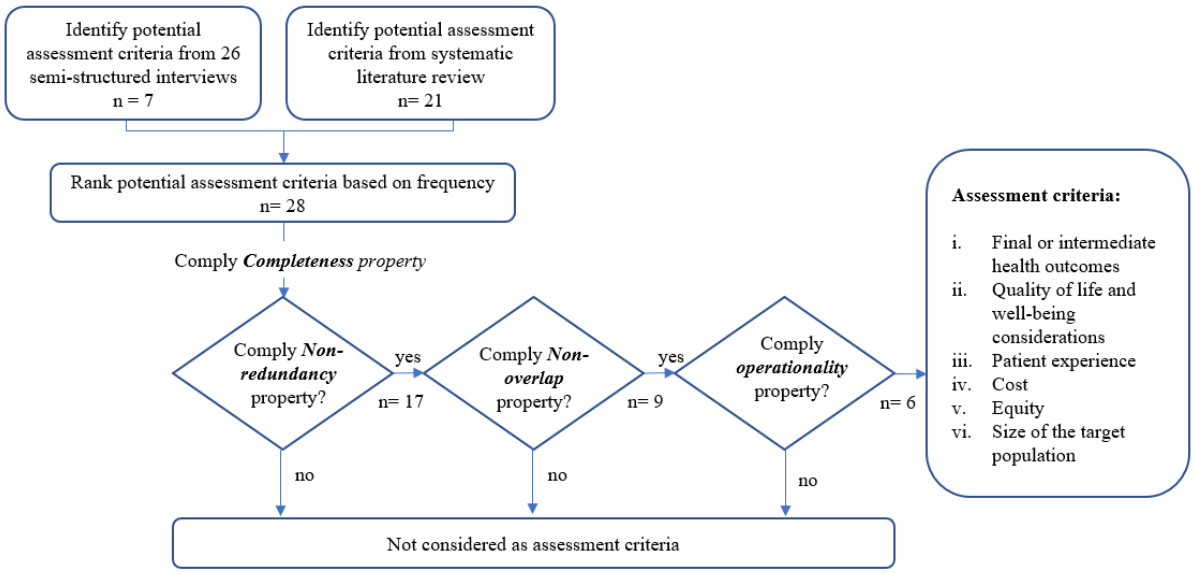
(v) *Size of the target population*: This relates to the number of beneficiaries covered by a healthcare intervention. This criterion is relevant when comparing interventions that target different conditions.

(vi) *Cost*: This includes total direct medical cost of the intervention (i.e. healthcare consumption, including all levels of healthcare and social care) per patient.

Figure 4.4 summarises the selection process of the assessment criteria. The list of all potential assessment criteria and an explanation of the stepwise identification process is summarised in **Appendix 4.A** (**Table 4.A.1.** and **Table 4.A.3**).

[21] It could be argued that all wellbeing dimensions are not fully captured by the definition that the criterion "Quality of life and well-being considerations" adopts, and in some cases, particularly in the context of integrated care, wellbeing may be a more appropriate assessment criterion than quality of life. The reason for not including wellbeing was the difficulty in operationalising this criterion using routinely collected data (see **Appendix 4.A**). However, as part of the further development of the proposed decision support framework, it is important to explore how the impact of new models of care on people's wellbeing can be taken into account.

Figure 4.4. Identification of assessment criteria



4.3.3. Measuring performance & scoring alternatives

To comply with the practicality principle of the decision-support framework, the performance of the alternatives on each of the criteria is based on routinely collected data, available to local commissioners (215). The operationalisation of the criteria (i.e., definition of performance indicators) depends on the nature of the medical condition or the interventions under assessment, as well as on the availability of data. For example, for evaluating the performance of EIP services on the "Quality of life and well-being considerations" criterion, Health of the Nation Outcome Scales (HoNOS) data-sets could be used, as HoNOS is a clinical instrument used to assess the health status and social functioning of people with severe mental illness (241, 242).

To estimate the impact of the alternatives under evaluation on the performance indicators of interest (e.g., HoNOS scores) and address the question of attribution, observational or quasi-experimental methods are used. Typically, the best evidence in economic evaluation comes from experimental approaches, such as randomised control trials (243)[22]. Although randomised trials are commonly considered the gold-standard for evidence, they can be expensive, time-consuming and might have limited external validity (150, 244). This conflicts the relevance and practicality principles of the decision-support framework, since local commissioners prioritise their own timelines and prefer context-specific evidence. Observational or quasi-experimental approaches, based on routinely collected data, can address potential endogeneity and estimate the causal impact of simple services or new model of care, over the indicators of interest (23, 150, 244).

Quasi-experimental methods, which include propensity score matching (PSM), Difference-in-Differences, regression discontinuity, among others, are statistical techniques used to establish cause-and-effect relationships between variables, without relying on random assignment (244). In the updated framework for developing and evaluating complex interventions, the UK Medical Research Council advocates the use of these methods, as they can be effectively used retrospectively with routinely collected data, and generate evidence both quickly and inexpensively. In addition, these methods have the potential to generate high external validity evidence while reducing internal validity threats, and ascertain impacts on non-health related and long-term outcomes (38, 150). The use of routinely collected data and quasi-experimental approaches offers significant advantages over other performance scoring methods. For instance, unlike the Analytical Hierarchy Process (AHP) and direct rating (see **Chapter 3**), which largely depend on subjective judgments and expert opinions, these methods provide a more empirical, data-driven basis for decision-making. **Chapter 6** presents a case study of a service evaluation of two condition-specific programmes in the context of psychosis, using quasi-experimental

[22] By randomly assigning two groups, researchers can guarantee that both groups are comparable in every aspect except for their assignment to either the treatment or control groups. Therefore, researchers can be confident that any variation in outcomes can be attributed to the treatment's impact.

methods to allow for the comparability of two EIP services in Buckinghamshire, Oxfordshire, Berkshire West ICS.

After estimating the performance of the alternatives on the assessment criteria, scores are obtained by converting performance indicators onto a common scale, considering the different units used for each criterion (215). To obtain scores on a 0 (no value) to 1 (maximum) value scale for all criteria, the relative standardization method is used (215). It has been applied by several previous MCDA studies that have also used data to estimate alternatives' performance (36, 37, 245). Although most MCDA studies in healthcare have relied on direct rating methods, as shown in **Chapter 3**, these methods can be prone to biases if decision-makers are influenced by recent experiences or other factors. Direct rating can struggle with different types of data, especially if they have very different scales or units (192). In addition, the use of categorical scales may result in a loss of information (246). Relative standardisation can take different types of data (ordinal, interval, ratio) and convert them to a common scale. The formula for relative standardisation is presented below (1).

$$S_{a,k} = \frac{x_{a,k}}{(x_{a,k}^2 + x_{b,k}^2)^{1/2}} \qquad S_{b,k} = \frac{x_{b,k}}{(x_{a,k}^2 + x_{b,k}^2)^{1/2}} \qquad (1)$$

$S_{a,k}$ ($S_{b,k}$) represents the standardisation of the performance score *x* of alternative *a* (*b*) on criterion *k* relative to the combined performance score of both alternatives, *a* and *b*, on criterion *k*. For outcomes where higher non-standardised score (i.e., on the natural scale) indicates a worse performance (e.g., lower costs are desirable), the scores $x_{a,k}$ and $x_{b,k}$ are replaced by $\frac{1}{X_{ak}}$ and $\frac{1}{X_{bk}}$, respectively. The resulted scores are called standardised performance scores (215).

As indicated by Groothuis-Oudshoorn and colleagues, the transformation of performance on the natural scale to a standardised score can be done using a local scaling approach, under which the best and the

worst unstandardized performance scores of the alternatives are used as the upper and lower bounds of the value function. The formulae for local standardisation is $S_{a,k} = \frac{x_{a,k} - x_{min}}{x_{max} - x_{min}}$, where x_{min} represents the minimum observed unstandardized performance score and x_{max} represents the maximum observed unstandardized performance score. Alternatively, the transformation can be done "globally", as described in this section, where the bounds are based on estimates of the worst and best possible outcomes, regardless of the alternatives' performance (215). For the decision-support framework that this thesis develops, the global approach is preferred over the local approach as it results more relevant when comparing performance of two interventions based on observed values, and it also makes it less sensitive to extreme values.

4.3.4. Weighting the assessment criteria

Criteria weights aim to capture priorities or preferences of the public for the different criteria. The weights represent "trade-offs" between criteria and are used to combine the scores on individual criterion into a measure of "total value" (48). To account for society preferences into the local commissioning of new models of care, it is desirable that the relative weights of the assessment criteria identified reflect preferences from the public in the country. As indicated earlier, local commissioners are expected to act on behalf of their communities when making allocation decisions. It is also argued that members of the public should voice their opinions in the priority-setting process, in return for paying taxes (231, 247). From a communitarian perspective, individuals are not only the recipients of rights, but members of a community with the duty to set the basis of the social choice rule (248). Making explicit societal preferences when setting priorities can also be justified based on democratic principles (6). Considering these ideas, which are also embedded in the underpinnings of CUA, prioritizing models of care that better satisfy societal preferences could arguably improve societal well-being and the legitimacy of the decision-making process in itself (249-252).

A discrete choice experiment (DCE) is used to elicit people' preferences between criteria. A DCE is a choice-based stated-preference method that is used to quantify public preferences for healthcare priority-setting (253-255). It involves running surveys in which participants choose between hypothetical alternatives defined by various attributes (i.e. assessment criteria). Through multiple choice sets, researchers infer preferences and quantify the trade-offs individuals make between criteria. Even though DCEs have not been the most used weighting method in MCDAs in healthcare (see **Chapter 3**), it was chosen in the development of this framework because it is theoretically sound given its grounding in consumer theory (256), random utility theory (257) and experimental design theory (258). DCEs are relatively easy to administer, do not require large samples, and participant's voices are equally taken into account as responses are usually equally weighted in aggregation (259). The method has also shown to produce reasonable predictions of health-related behaviours (260). Between 2013 and 2017, at least 47 choice experiments were conducted to investigate trade-offs between health outcomes and patient or consumer experience factors, and 27 to develop priority setting frameworks (255).

Chapter 5 of this thesis elaborates on the DCE conducted to elicit assessment criteria weights for the evaluation component of the decision-support framework. It details the experimental design and the models used to elicit people's preferences in England.

4.3.5. The MCDA total score

As indicated at the beginning of this section, Multi-Attribute Value Theory (MAVT) is used for the MCDA-based evaluation component of the decision-support framework. Under MAVT, standardised performance scores are combined using the assessment criteria weights to derive a single value score for each alternative, allowing for a comprehensive comparison (161). Under the additive representation,

which is the most common approach for evaluating alternatives (**Chapter 2**), standardised performance scores ($S_{a,k}(x_{a,k})$) are weighted according to people's preferences (w_k)[23]:

$$V(X)_a = \sum_{k=1}^{m} w_k \, S_{a,k}(x_{a,k}) \qquad (2)$$

Where $V(X)_a$ represents the multi-attribute value function, or MCDA total score for alternative *a*, based on the performance of the alternative on m assessment criteria. Since the evaluation component of the framework is based on six assessment criteria, *m* equates to six in this formula.

There are two considerations when using this approach. First, when computing MCDA total scores using the MAVT-based additive model, mutual independence of preferences is a crucial concern. This means that the weight attached to one criterion should not depend on the performance on other criteria (211, 236). Additional research is required to establish the most suitable course of action with regards to cases of preferential dependence. Some researchers would suggest restructuring the criteria set, assuring compliance with the preference independence principle, or adopting a different aggregation function (210). The former option could be problematic since relevant criteria may be excluded from the MCDA, violating the *completeness* criteria property. The second option would involve considering a multiplicative, multi-linear or Scale Loss Score (SLoS) model[24], but one needs to weigh the benefits of using these alternative models against their increased complexity (263). The main strength of the

[23] Additive functions are also frequently used in instruments designed to measure health-related quality of life, such as the EuroQol five-dimensions questionnaire (261).

[24] Under a multiplicative aggregation model the overall MCDA score for an alternative is determined by multiplying the standardised performance scores of each criterion rather than summing them, with each criterion's value being raised to the power of its respective weight before multiplication. The multi-linear model combines the linear and product model. These models are less commonly used in practice, as establishing their functional form and estimating the necessary parameters for these models is generally more complex compared to additive models (48). The SLoS model was developed by Saint-Hilary et al. to assess and compare the benefit-risk balance of treatments (262). The SLoS overall score of an alternative is based on the inverse of its performance scores across different criteria. For each criterion, the inverse of the performance score is taken, and this value is then raised to the power of its respective weight. These adjusted values are then summed across all criteria to obtain the overall SLoS-based MCDA score. A lower SLoS indicates a preferable alternative. While SLoS models may be more robust to the correlation in the criteria, their complexity may have contributed to their lack of widespread utilisation (263).

linear model is its intuitive interpretation[25], making it highly appealing to decision-makers (48). In light of this debate, the preference independence principle of the assessment criteria is statistically tested when estimating the criteria weights in **Chapter 5**[26].

The second consideration concerns whether to include cost as a criterion in the MCDA, reflecting forgone value or opportunity cost. Advocates for including cost in the MCDA contend that reducing healthcare cost growth is a key objective of new care models, and consequently, cost is an integral part of the decision-making process. In addition, by including cost in the MCDA, its relative contribution to the decision is made more explicit, and it is also possible to elicit people's willingness to pay for gains on other criteria (36, 122, 149, 264). Conversely, some authors contend that although including costs in the MCDA makes the trade-offs explicit, cost is not a value element and MCDA's potential lies in reflecting a more comprehensive benefit measure. To these authors, it is crucial to establish the opportunity costs of each additional benefit unit, which cannot be achieved by incorporating cost into the MCDA (210, 265, 266). In a budget-constrained context, stakeholders' preferences are not the appropriate basis for evaluating the efficiency of health interventions and the benefits lost from not funding alternative options are more relevant than stakeholders' willingness to pay (ibid). Consequently, commissioners need to assess whether the benefits of a new, costlier intervention justify its expense compared to the benefits that could have been produced by allocating resources elsewhere.

[25] A poor performance on criterion k can be compensated by a good performance on criterion j, and vice-versa.

[26] Arguably, the potential violation of the preference independence principle should have been formally considered during the criteria selection phase. However, this principle not only concerns the criteria themselves but also the preferences (relative weights) that stakeholders assign to them. Assessing preference independence entails understanding how preferences for one criterion are influenced by the performance of others, which can be challenging to comprehensively assess based on literature reviews. Therefore, it was decided to address the preference independence principle in **Chapter 5**, which estimates the relative criteria weights. Nevertheless, future research in the field of MCDA should explore methods to formally assess compliance with the preference independence principle during the criteria selection process.

In principle, the resulted composite benefit score could be used to calculate a modified version of the incremental cost-effectiveness ratio. The ratio of the 'extra cost per extra unit of MCDA-based benefit' could be derived by dividing the incremental cost (i.e., the difference in total costs of the two alternatives being compared) by the difference in the MCDA-based benefit scores. However, for such evaluation the modified incremental cost-effectiveness ratio must be compared with a newly defined threshold, as the conventional threshold used in cost-effectiveness analysis does not encompass all elements of value (37, 408). This threshold would not be unique, as the performance scores used to operationalize the criteria will potentially change depending on the nature of the medical condition or the interventions under assessment. Therefore, when the set of outcome measures changes, new thresholds must be determined. Additionally, the threshold may vary among ICSs due to their varying levels of economic development, which affects their ability to pay for healthcare interventions.

In recognition of this unresolved debate, for each alternative under assessment, the decision-support framework presents two value scores to the local commissioners: the MCDA total score including the cost criterion and the cost-per-value ratio (CVR), which indicates the cost of the alternative per unit of its aggregate value. The CVR of an intervention is calculated as the ratio between the unstandardized total cost of the alternative and the MCDA score excluding the cost criterion. Both scores, the MCDA total score and the CVR, are used to rank interventions. The comparison of results with these two metrics provides a more nuanced understanding of the healthcare interventions under assessment. By comparing both results, local commissioners can better understand how costs influence the overall value of interventions and assess the balance between economic and intrinsic value considerations. To be able to obtain both metrics, in **Chapter 5**, the DCE uses a split sample design to obtain assessment criteria weights for both scenarios: when the cost criterion is included and when it is excluded.

4.3.6. Sensitivity analyses

After computing the MCDA total score for each alternative under assessment, it is important to consider the potential impact of uncertainty on the MCDA results. As highlighted in **Chapter 2** and in good practice guidelines, MCDAs are susceptible to uncertainty and analysing and reporting it is essential for good practice (48).

In all economic evaluations, following Briggs et al typology, there are two forms of uncertainty: parameter and structural uncertainty (267)[27]. The former refers to the variability in the estimation criteria weights and performance scores as a result of the sampling (i.e. estimation error). The second, structural uncertainty, relates to the assumptions imposed in the decision model (e.g., aggregation model used, criteria included, selection of comparator).

To address parameter uncertainty, probabilistic sensitivity analyses (PSA) are incorporated as intrinsic part of the framework, using Monte-Carlo simulations with 1,000 replications (188, 215). As criteria weights are estimated from empirical data (**Chapter 5**), standard errors of the point estimates are used to represent the extent of the parameter uncertainty in weights. A similar approach is adopted to address parameter uncertainty in performance scores as estimates are based on clinical data. The resulting MCDA scores distributions are graphically summarised, together with the standard deviation and the corresponding 95% confidence intervals (CIs). In addition, to summarise the sensitivity of the results to uncertainty, the probability of preferring one of the interventions (*a*) over its comparator (*b*) is

[27] Different authors have suggested multiple classifications of uncertainty. Broekhuizen et al, for instance, differentiates between stochastic uncertainty (random variability in outcomes between identical patients); parameter uncertainty (uncertainty in estimation of the parameter); heterogeneity (variability between patients that can be attributed to characteristics of those patients), and; structural uncertainty (assumptions inherent in the decision model) (188). Briggs et al. distinguish between variability (the differences that occur between patients by chance) and heterogeneity differences that occur between patients that can be explained) from decision uncertainty (267).

computed. This is done by calculating the proportion of MCDA iterations in which the intervention *a* had a higher MCDA total score than intervention *b*. Furthermore, parameter uncertainty in the MCDA is graphically presented with the conditional multi-attribute acceptability curve (CMAC), proposed by Rutten-van Mölken and colleagues (37, 149). The CMAC shows, across a variety of budget levels, the probability of a healthcare intervention being the most preferred option (i.e., having the highest MCDA score) while the budget impact remains below a budget-threshold. The CMAC extends the concept of acceptability curves, commonly used in CEA, to a multi-criteria context. The vertical axis of the CMAC displays the likelihood of a healthcare intervention being accepted as the most preferred alternative, whilst the horizontal axis shows varying thresholds of maximum budget for allocating to either of the interventions under evaluation to treat a given population size. The CMAC can support budget impact analysis as it visually demonstrates how the preference for healthcare interventions changes across a range of available budgets (ibid).

The impact of structural uncertainty could be explored by running the analysis using different weighting methods or different performance indicators. However, this can be costly and time-consuming, which partially explains why most MCDA studies in healthcare overlook structural uncertainty, as found in **Chapter 2**. The choice of criteria and their relationships is another source of structural uncertainty. In the proposed decision-support framework, this type of structural uncertainty is addressed by reporting two overall measures: the MCDA total score including the cost criterion and the CVR, which excludes the cost criterion. A different aggregation model, such as the SLoS model, could also be considered, particularly if the preference independence principle, required in additive models, does not hold. In response to this, as indicated in **Section 4.3.5**, the preference independence principle is statistically tested in **Chapter 5** when conducting the DCE.

4.4. The monitoring component of the framework

The monitoring involves regularly tracking and reporting on intervention implementation (230). From a logic model approach, a set of process and output indicators is used to provide decision-makers with accurate and timely information to review progress and performance (230, 239, 268).

Monitoring criteria should help local commissioners to keep track on progress and performance of the interventions in place. To this purpose, the monitoring of interventions, including new models of care, should be based on a set of core indicators and targets that provide timely and accurate information to decision-makers[28]. As identified during the semi-structured interviews conducted, performance-monitoring data is commonly consolidated as key performance indicators (KPIs) and displayed on dashboards (see **Chapter 2**). From the logic chain perspective, these indicators primarily emphasise the process or outputs arising from the planning, resources, and inputs used to implement a healthcare intervention (239, 269).

To identify monitoring criteria, a comparable approach to the one used for selecting assessment criteria was adopted. Hence, the 28 potential indicators were reviewed to determine which are applicable to the decision-making context (i.e., monitoring the implementation of new care models) and could be developed as output or process indicators (see **Tables 4.A.2 and 4.A.3**. in **Appendix 4.A**). The resulted monitoring criteria were four. A graphical representation of assessment and monitoring criteria, from the logic chain approach, is presented in **Appendix 4.B**.

[28] As the monitoring aims to inform decisions in the short-mid-term, information has to be available with relatively high periodicity (i.e., monthly or quarterly).

i. *Patient lifestyle and self-management behaviour (and intermediate health outcomes)*: KPIs that track improvements in the choices and actions of patients related to their daily living and how they manage their health conditions. In the case of some health conditions, it is also desirable to monitor improvements in health status of patients on a regular basis.

ii. *Quality of care or features of the process of care delivery*: KPIs relating to timely access, coordination, and the effective delivery of services to improve the likelihood of desired health outcomes.

iii. *Compliance with national standards*: KPIs on the compliance with standards established by NHS England, as detailed in the NHS Long Term Plan or in NHS guidance for various health conditions or healthcare models.

iv. *Equity in quality of care*: KPIs that summarise the gap in terms of quality of care or features of the process of care delivery (across geographical areas, between men and women, of special groups).

The NHS in England has made significant progress on developing clinical audits and registries to compare the performance of health and care services against standards and national trends (126). Clinical audits and registries have been developed for multiple health conditions. Examples of these are the National Audit of Cardiac Rehabilitation, the National Bowel Cancer Audit, the National Clinical Audit of Psychosis, or the National Pulmonary Hypertension Audit (Ibid). In addition to this, NHS England has published guidelines for the implementation of new models of care, such as the "Implementing the early intervention in psychosis access and waiting time standard" (270) or the "Urgent and emergency care improvement guide to specialty support to the UEC pathway/internal professional standards" (271). In line with all these instruments, several ICSs have also developed dashboards to monitor the performance of the services that are being provided to their populations, including new care models (see **Chapter 2**). An example of it is the diabetes dashboard developed in

collaboration between South Central and West Commissioning Support Unit, Oxford University Hospitals NHS Foundation Trust and local commissioners from BOB-ICS (272). In addition to these sources, NHS England has also developed tools to identify disparities within and among different regions in England. An example of them is the Atlas of Variation, produced in collaboration with the Office for Health Improvements & Disparities, NHS England and RightCare (273). Most of the national audit reports also include a section that highlights health inequalities, with particular attention paid to gender and ethnic groups (126).

For the monitoring component of the framework, it is proposed to classify KPIs according to the four monitoring criteria. This categorisation encourages commissioners to balance and prioritize KPIs across different areas, facilitating the selection of the more appropriate ones. Given the numerous KPIs available (see **Chapter 2**), this approach would ensure that monitoring efforts are aligned with key areas of healthcare intervention performance. This could enable more comprehensive monitoring of the performance of new care models, as well as comparisons across diverse health conditions.

To identify poorly performing interventions, a 'streetlight rule' could be implemented to indicate deviations from adjusted benchmarks aligned to NHS England standards or national averages. In addition, time trends could be added to the dashboard as they could reveal patterns and trajectories of KPIs, helping to identify whether the performance of the interventions are improving, declining, or remaining stable. An example of the suggested dashboard is provided for EIP services in **Appendix 4.C.** Nonetheless, the KPIs for each criterion should be established in collaboration with clinicians and in accordance with NHS England guidelines.

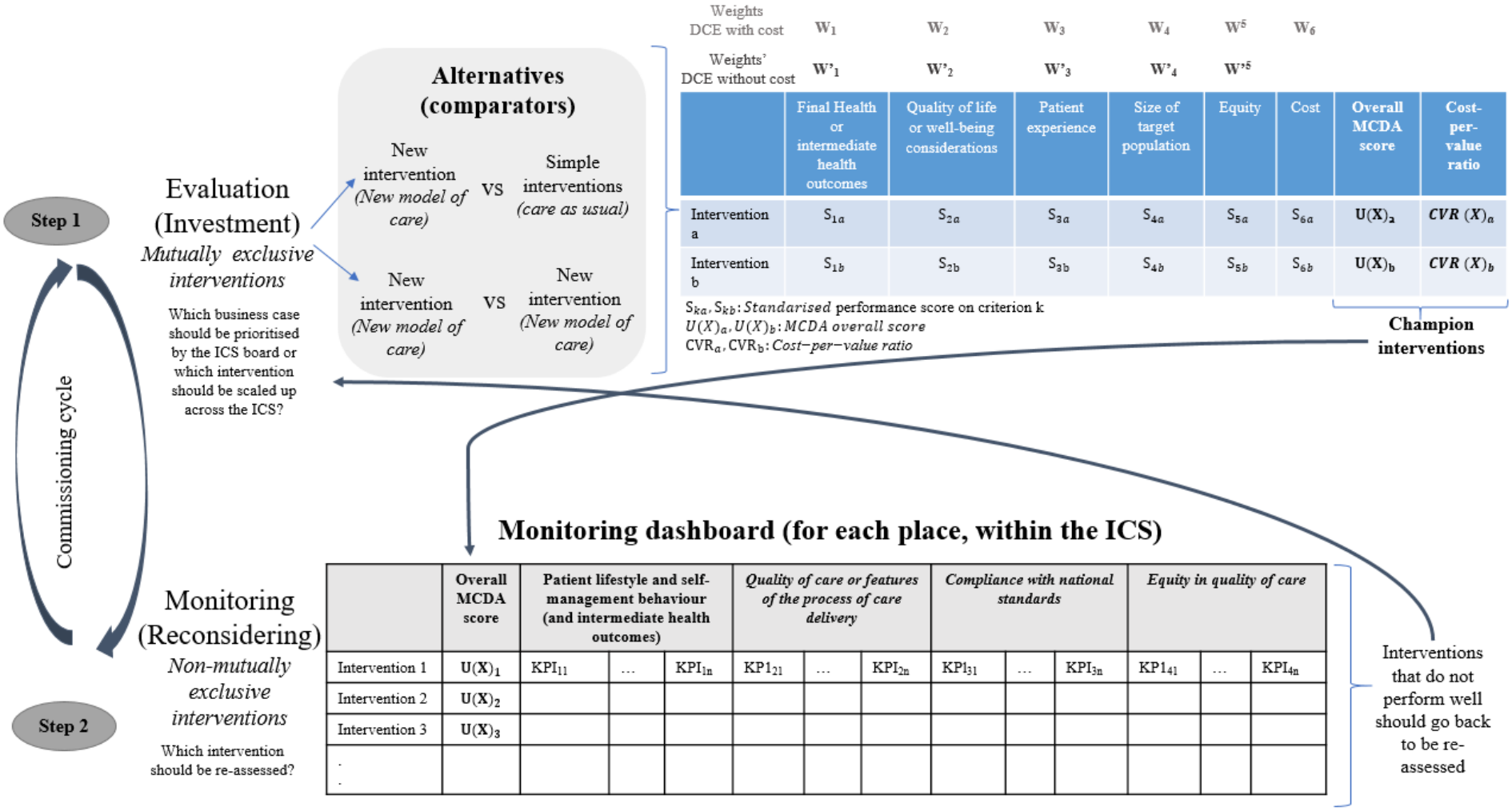
4.5. The framework as part of the commissioning cycle

As part of the local commissioning cycle, there is an interconnectedness between the monitoring and assessment of healthcare interventions, including new models of care. Ideally, all interventions in place would be assessed, using the MCDA model explained in **Section 4.3** (Step 1)**.** Each intervention is typically compared against a standard or baseline, and interventions are then ranked according to their MCDA overall score, in a composite league table format[29]. Ongoing monitoring takes place to track the progress of the intervention against the predefined monitoring criteria and their corresponding KPIs (Step 2). This allows for timely adjustments to be made if any deviations from the implementation plan are identified. Interventions that have been in place for over one, two or more years, with KPIs exhibiting low performance, are considered for reassessment. For the evaluation of these interventions, a comparator is defined and intervention are assessed to determine their prioritisation by the ICS board or to decide whether they should be implemented more extensively across the ICS (Step 1). Insights and lessons learned from the evaluation are used to inform the next cycle of commissioning. Based on the evaluation results and feedback, decisions are made about the continuation, modification, or cessation of the intervention. Interventions with the highest MCDA scores are considered 'champion interventions', and added to the monitoring dashboard. The cycle is then repeated, with interventions that yield low performance on their KPIs being considered for a new assessment. This cycle ensures that healthcare interventions are continually assessed and improved upon, ensuring they are effective and provide value for money.

[29] In priority-setting context, the aim of the league table is to rank healthcare interventions to determine which should be included or excluded from a health benefits package or public reimbursement. League tables are typically founded on incremental cost-effectiveness ratios and are not an essential component of the MCDA methods. However, composite league tables have been used in some MCDA studies to health priority-setting decisions, by ordering interventions according to value for money (e.g., cost-per-value ratio) and comparing them with the budget available to fund them. See for example Airoldi et al. (2011) (276) and Defechereux et al (2021) (345).

Figure 4.5. below summarises the features of the decision-support framework that this thesis develops, embedded in the monitoring and evaluation cycle.

Figure 4.5. A decision-support framework for the local commissioning of new models of care



4.6. Discussion & Conclusions

This chapter developed the structure of the decision-support framework proposed for local commissioning of new models of care in England. Performance scores are derived from routinely collected data and quasi-experimental methods, while criteria weights are based on preferences elicited from the general public in England using decision theory. The framework's monitoring and assessment criteria adhere to the logic model approach by capturing both output indicators and longer-term outcomes and impacts. This ensures a comprehensive evaluation of the healthcare interventions in place.

When developing the evaluation component of the framework, six value elements of healthcare were identified based on the interviews, literature and decision-theory. Health-related outcomes and quality of life reflect the primary goal of any health intervention and align with the long-standing interest in QALYs. Equity and size of target population align with a population health perspective and concerns for the wider implications of healthcare. Patient experience or satisfaction appeals to the value attributed to care processes, particularly relevant to the elderly and frail population with complex health needs. Cost is not an element of value but essential decision factor when budget is limited. This set of elements or assessment criteria is similar to those suggested by other studies in the context of integrated care and expands beyond health-related outcomes, which aligns with the holistic approach taken by ICSs in England (8, 245).

Although there is some merit in arguing that MCDA should consider the costs required to achieve service transformation for establishing different models of care, the cost criterion was limited to the total direct medical cost of the intervention. This was done to ensure comparability in performance between the two alternatives under evaluation. Nevertheless, as part of a comprehensive feasibility

analysis, it is essential for local commissioners to consider the investment required to achieve service transformation within the decision-making process.

Due to the highly diverse nature of new care models, the suggested assessment and monitoring criteria have been broadly defined. This allows decision-makers to decide on the most suitable KPIs to operationalise each assessment criteria. While this flexibility enables a tailored approach to different health conditions, it also introduces the potential for biases or inconsistencies in the choice of performance indicators and KPIs. Therefore, it is crucial to justify the selection of these indicators, considering the specific health condition or type of health intervention under assessment. For guidance, clinicians and relevant research can be consulted, along with standardised outcome measures, such as the ones proposed by the International Consortium for Healthcare Outcomes Measurement. The later cover clinical outcomes, patient-reported outcome measures (PROMS), and patient-reported experience measures (PREMS) for a range of health conditions (274).

The evaluation component of the framework is grounded on MCDA, enabling the systematic integration of all value elements of healthcare and societal preferences for these values in the priority-setting process. Nevertheless, the use of MCDA in healthcare has attracted criticism due to its complexity, susceptibility to subjective bias, and difficulties in defining and weighing criteria consistently across diverse healthcare decisions (162). **Chapter 3** draws attention to the lack of quality and transparency in many of its applications (165, 210, 240). To address these critiques, the MCDA of this framework has followed best practices and builds on the theoretical foundations of decision theory (47, 48). It is however important to note that MCDA does not intend to substitute the expertise or judgment of local commissioners. Rather, it provides detailed information on a comprehensive set of elements of value in healthcare, assisting commissioners in understanding value trade-offs while allowing the incorporation of preferences from the general public (162, 165). In fact, the examination of MCDA findings by

decision makers, including exploration of how uncertainty affects MCDA results, is part of the ISPOR MCDA Good Practice Guidelines checklist (47, 48).

The proposed MCDA of the decision-support framework accounts for the preferences of the general public in England through the weighting of criteria. However, this approach may be perceived as a limitation as one of the major strengths of an MCDA is its ability to incorporate diverse stakeholder perspectives. Comparing MCDA results based on different viewpoints could offer valuable insights into how various groups perceive the importance of different outcomes (149). While this approach could enrich the decision-making processes, when it comes to make final decisions, one has to decide which perspective should carry greater weight. This thesis argues that societal preferences should take precedence over those of other stakeholders, such as patients, healthcare professionals, or politicians. Local commissioners are expected to take into account public preferences, as they act on behalf of their communities. The public also forms the taxpayer base and are the ones that bear the cost of healthcare decisions (as current or potential patients or careers).

The decision-support framework proposed places significant emphasis on the evaluation component but falls short in elaborating on the monitoring component. This focus responds to the decision context outlined in **Chapter 2**, where local decision-makers highlighted a pressing need for rigorous evaluation of new models of care, which contrasts with the local and national development of dashboards or tools for monitoring the performance of interventions in healthcare. However, it is essential to acknowledge that the majority of priority-setting frameworks tend to overlook the relevance of monitoring. Given its critical role in the commissioning cycle, this aspect warrants further research.

In order to ensure the implementation of the decision-support framework it is important to demonstrate feasibility. **Appendix 4.C** presents an example for the application of the monitoring component and

Chapter 6 illustrates the potential applicability of the assessment component using routinely collected data. To prove the usefulness of the framework appropriate consultation with local commissioners and other relevant stakeholder is essential. This thesis has developed a decision-support framework that is relevant, sound and practical. However, the relevance of the framework, including the pertinence of the monitoring and assessment criteria have not been validated with local commissioners, which is a significant limitation.

This is not the first framework proposed for evaluating complex interventions such as new models of care. Campbell et al (2000) proposed a framework for evaluating complex interventions, but its reliance on randomized control trials - which can be expensive and time-consuming – limits its practicality (275). Airoldi and colleagues addressed some of the limitations of CEA and developed the Socio Technical Allocation of Resources framework to inform resource allocation decisions at local level in England (276). The framework is theoretically sound and builds on cost, population health benefits, health inequalities and feasibility criteria. However, it does not account for aspects such as patient experience, relevant for the current local commissioning context. In addition, the operationalisation of health benefits and feasibility criteria rely on direct rating and expert judgements (ibid). Another framework based on MCDA and designed to support local decision-makers is the Prioritisation Framework, developed by the UK Health Security Agency and Office for Health Improvement and Disparities (formerly knowns as Public Health England). It was designed to help local authority public health teams decide how to spend public health budgets. Despite the comprehensive nature of the framework, the focus is on public health interventions and its implementation requires a significant investment of time and commitment from public health teams, as the operationalisation of the criteria is based on several workshops (43, 44). The SELFIE MCDA-based evaluation framework[30] for integrated care programmes structures relevant concepts and elements of integrated care for multi-

[30] Sustainable intEgrated chronic care modeLs for multimorbidity: delivery, FInancing, and performance.

morbidity. Despite of its comprehensiveness and relevance, not all decision criteria can be systematically operationalised due to data constraints (149). More recently, Collins et al (2023) developed a combined multi-disciplinary priority setting and resource allocation framework. It outlines key stages in the priority setting process and incorporates findings from the literature and discussions with local stakeholders in Scotland. However, the evaluation criteria and the scoring method are not defined (277).

This is not an exhaustive review of all available frameworks to inform the allocation of health resources, nor does it intend to diminish their usefulness in healthcare. To date, the decision-support framework developed in this thesis appears to be the first to integrate the monitoring and evaluation stages inherent in the commissioning cycle. It defines a comprehensive yet applicable set of value-based elements and allows for the explicit recognition of public preferences in the local priority-setting process.

Chapter 5. Using stated preferences choice experiments to elicit MCDA criteria weights in the context of integrated care in England

5.1. Introduction

In the context of a fixed healthcare local budget, local commissioners are expected to prioritise new models of care based on objective evidence, favouring the most cost-effective models. In determining the cost-effectiveness of a care model, local commissioners are interested to account for multiple outcomes or value attributes and preferences from the general population. Although gains in 'years of life' and 'health-related quality of life', often measured in terms of the quality-adjusted life-year (QALY), remain of paramount importance, they are not the only outcomes that decision-makers and society values when making health care decisions (187, 278, 279). As described in **Chapters 1, 2 and 4**, in the context of integrated care in England, new care models are expected to improve population health and quality of care, tackle inequalities in outcomes, experience and access, and enhance productivity and value for money (90, 120, 229, 280).

Discrete choice experiments (DCE), a choice-based stated-preference method, are increasingly used to quantify public preferences for healthcare priority-setting, playing in important role in real-world decision-making (254, 255, 281, 282). In these experiments, participants are presented with two or more hypothetical alternatives, described by a number of attributes, and are asked to state their choice (283, 284). By presenting participants with several choice sets, it is possible for researchers to elicit their preferences, without directly asking them to state their preferences, and to quantify the trade-offs that people are prepared to make between different aspects of healthcare (285). DCEs are relatively easy to administer, do not require large samples and participant's voices are equally taken into account as responses are usually equally weighted in aggregation (6, 259, 286). DCEs are also theoretically sound as they are grounded on consumer theory (256), random utility theory (257) and experimental design

theory (258). The method has also shown to produce reasonable predictions of health-related behaviours (260). Between 2013 and 2017, at least 47 choice experiments were conducted to investigate trade-offs between health outcomes and patient or consumer experience factors, and 27 to develop priority setting frameworks (255).

In England, some choice experiments have explored people's preferences to guide resource allocation decisions in healthcare. Erdem and Thompson (2014), for instance, explored public's preferences for health service innovations in the UK (287). Shah et al (2015) conducted a DCE with members of the general public in England and Wales to examine society's preferences regarding the prioritisation of life-extending end-of-life treatments (288). Similarly, Rowen et al (2015) examined societal preferences for burden of illness alongside end of life (289). Hampson et al (2019) explored society's preferences across curative and non-curative therapies and large and small health gains (290). Morell and colleagues (2021) surveyed members of the UK public to study the relative importance of attributes of illness that influence the value placed on alleviating that illness (291). The relative importance of different aspects of mental healthcare quality to different stakeholders, including members of the general population, was estimated by Rowen et al (2022) (292).

In the context of integrated care, Rutten-van Mölken et al (2020) measured relative preferences for outcomes of integrated care of patients with multi-morbidity from eight European countries, including the United Kingdom (293). The experiment, however, did not include some relevant attributes for the English context, such as health equity, which has been at the centre of the NHS England agenda, particularly at local level (102, 119, 294). To date, no study has explored the factors that local healthcare commissioners and members of the public in England would give precedence to inform decisions in the context of integrated care.

This chapter aims to fill this gap by using DCEs to determine the relative importance of attributes of new models of care that, based on preferences from the general public, should guide local investment decisions in England. These attributes, identified in **Chapter 4** as the assessment criteria or elements of value, are integral to the decision-support framework developed in this thesis. Within this framework, the assessment criteria are used in a multi-criteria decision analysis (MCDA) to evaluate new models of care in England and support the local commissioning of integrated care. In order to account for society values into the decision-making process, it is desirable that the relative weights of the assessment criteria identified reflect preferences from the general public in the country. Therefore, the relative importance scores derived from the DCEs in this chapter will serve as weights for the assessment criteria, and subsequently used to calculate the MCDA total scores (see **Chapter 4**).

The chapter addresses the two considerations described in **Chapter 4 (Section 4.3.5)** regarding the use of a Multi-Attribute Value Theory (MAVT) approach for the MCDA-based evaluation component of the decision-support framework. First, as a MAVT model requires the outcomes of interest to be preferentially independent (48, 161, 210)[31], compliance to this property is explored by allowing for interactions between the attributes in the experimental design. The second consideration delves into the debate over including cost as a criterion in the MCDA. Proponents of including cost argue that managing healthcare cost growth is crucial in new care models, making it an essential aspect of local decision-making (36, 122, 149, 264). However, opponents argue that cost is not a value element, and the strength of MCDA lies in capturing a broader measure of benefits (210, 265, 266). To navigate this debate, for each alternative under evaluation, the framework presents two value scores to local commissioners: the MCDA total score with the cost criterion and the cost-per-value ratio (CVR), reflecting the cost per unit of aggregate value. To be able to obtain both metrics, the DCEs conducted in this chapter use a split sample design to obtain assessment criteria weights for both scenarios: when

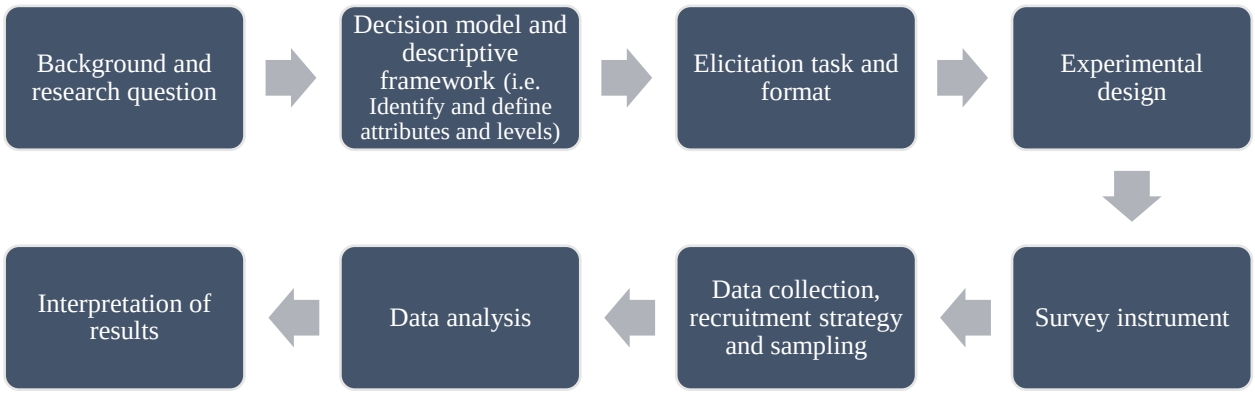
[31] For all the outcomes X: x_i where i=1,..,n where *i=1,..,n,* outcome x_i should be preferentially independent of the other outcomes, such that weight attached to one criterion should not depend on the performance on other criteria.

the cost criterion is included and when it is excluded. This allows local commissioners to make informed decisions based on both metrics.

In addition to estimating assessment criteria weights for the decision-support framework, this chapter conducts further analyses to provide a more comprehensive understanding of people's preferences. The extent to which preferences vary by individual respondent characteristics is explored, which is known in the DCE literature as observable heterogeneity in preferences (295, 296). In addition, to further inform local commissioners, the chapter provides estimates of people's willingness to pay (WTP) for improvements in health outcomes (e.g., years of life), quality of life, quality of care, and equity, as well as the size of the target population.

To conduct this stated-preference study, the framework suggested by Rivero-Arias et al (2022) (297) was followed (see **Figure 5.1**). The study also followed the best practice guidance for the development of this type of studies (59, 284, 285, 298), including the ISPOR Good Research Practices for the application of conjoint analysis in health (258, 299).

Figure 5.1. A framework for stated-preference studies



Source: Adapted from Rivero-Arias et al (2022) (297)

5.2. Decision model and descriptive framework

5.2.1. Identification and definition of attributes and levels

As indicated in **Chapter 4**, six assessment criteria were identified as the elements of value to assess new models of care, using an MCDA-based framework. For the definition of each criterion (henceforth, 'attribute') and their corresponding levels, previous stated-preference studies in healthcare priority-setting were reviewed to identify all possible options. In addition, definitions from policy documents and surveys conducted by NHS England were also analysed. Based on both sources, a preliminary definition for each attribute and selection of levels was developed, taking into account formative research on best practice guidance (59, 258, 284, 285, 297-299).

The attribute definitions and corresponding levels were presented to the Patient and Public Involvement (PPI) group from the National Institute for Health and Care Research (NIHR) Oxford and Thames

Valley Applied Research Collaboration (OTV ARC) in June 2022. Each one of the attributes was discussed and the wording was refined based on PPI's feedback. Additionally, simple and easy-to-understand icons were used to avoid potential choice inconsistencies (301). The selection of levels considered the ability to calculate relevant marginal rates of substitution (MRS), the possibility to test the functional form of quantitative attributes[32], as well as potential participant cognitive burden (i.e., levels are limited to three per attribute) (284, 299).

The attribute names, definitions and levels used in the descriptive system of this experiment are presented **Table 5.1**. **Appendix 5.A** summarises the rationale behind the definition of each one of the attributes, and **Appendix 5.B** includes an in-detail explanation of the levels selection, with the pictorials. When defining the attributes, special attention was place to t

All attributes were expected to have a positive relation with utility except for the 'Additional budget needed'. The combination of the two first attributes, i.e. additional years of life and quality of life improvements, allowed the estimation of QALY measure, which is the most commonly used measure of effectiveness in healthcare economic evaluations. As the relative importance of the QALY compared with other value elements is unclear, a QALY gain variable using the collected experimental data was created (see **Section 5.2.7**).

Table 5.1. Attributes & levels

MCDA assessment criteria	Attribute name	Definition	Levels
Final or intermediate health outcomes (a)	**Additional years of life (**YoL)	The number of additional years of life that each patient is expected to have with the care programme.	▪ 0.5 year of extra life ▪ 1 year of extra life ▪ 3 years of extra life

[32] Three levels were assigned to each attribute to allow for the identification of nonlinear preferences.

Quality of life and well-being considerations (b)	**Quality of life improvements** (QoL)	Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression, due to the care programme.	▪ 20 points improvement ▪ 40 points improvement ▪ 60 points improvement
Patient experience (c)	**Patient experience** (Exp)	How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, etc…	▪ Poor ▪ Fair ▪ Good
Equity	**Target population** (Equ)	Proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, people from low-income households.	▪ 25% disadvantaged ▪ 50% disadvantaged ▪ 75% disadvantaged
Size of the target population	**Size of target population** ($Size$)	The number of people who benefit directly from the care programme per 100,000 citizens.	▪ 50 / per 100,000 citizens ▪ 5,000 / per 100,000 citizens ▪ 10,000 / per 100,000 citizens
Cost	**Additional budget needed** ($AddTax$)	Additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme.	▪ £20 additional tax / per year ▪ £40 additional tax / per year ▪ £60 additional tax / per year

(a) The decision not to establish a baseline for the 'Years of life' attribute (ie., potential patient age) was made in alignment with the thesis goal of developing a decision-support framework applicable to various new care models, irrespective of the age or condition of the target population (e.g., individuals with multi-morbidity, children, the elderly). Although some DCEs have set clear baselines for longevity attributes, literature suggests that preferences may differ based on the baseline, with the general public often favouring gains in years of life for younger individuals over the elderly (209, 253, 309). While one option could have involved setting an age for patients benefiting from care programs, varying based on respondents' age, this approach was not considered during the experiment's design phase. Moreover, while the absence of a baseline for this attribute could be seen as a limitation (as it could be difficult to interpret a change without a clear baseline), PPI contributors validated the definition and format in which levels for this attribute were presented in the experiment.

(b) Improvements in quality of life are described by using the EQ-VAS scale numbered from 0 to 100, in which 100 means the best imaginable health state and 0 means the worst imaginable health state. While labelled as 'Quality of life improvements,' this attribute encompasses not only enhancements in physical functioning (such as mobility, self-care, usual activities, pain, and discomfort) but also factors related to psychological well-being (anxiety and depression). However, it could be argued that other aspects of well-being, such as social relationships or enjoyment of life, are not fully captured by this definition, and consequently the EQ-VAS might not be the optimal instrument for framing this attribute comprehensively. Despite this, discussions with PPIs revealed that the proposed definition was deemed easy to understand. Furthermore, it enabled the treatment of this attribute as numerical, which was essential for the research's goal of estimating people's preferences for QALY gains.

(c) The concept of patient experience encompasses various dimensions throughout the care delivery journey, reflecting its broad and multifaceted nature (359). To enhance respondent understanding, discussions with PPI contributors explored the possibility of breaking down this attribute into two or three components, such as "Continuity of care" and "Person-centeredness" (239). However, PPIs indicated a preference for not braking down the criterion and validated the definition of 'experience' that the experiment adopted, which is aligned with the one used in NHS surveys33. In addition, it was determined that maintaining patient experience as a singular concept was essential in order to comply with the practicality principle of the decision-support framework. Attributes like "Continuity of care" and "Person-centeredness" are difficult to operationalise based on routinely collected data.

5.2.2. Elicitation task and format

An unlabelled paired comparison elicitation format was used in this experiment. Each paired comparison included two hypothetical healthcare interventions (henceforth, 'care programmes') and the profiles for each alternative, describing the attributes with their specific levels. For each choice task, participants were asked to choose between two 'Care programmes', with no opt-out option[34]. Participants were asked to imagine themselves as local commissioners of health and social care that have to decide which care programme to prioritize for funding, given a limited budget[35]. To increase participant's attention and reduce task complexity, colour coding was used (302). The choice task was discussed in a second meeting with the NIHR OTV-ARC PPI group, and tested at the Oxford Biomedical Research Centre Open Day in July 2022, with members of the public. For the latter, copies with different potential choice tasks were printed, and members of the public were asked to complete the choice task and give their opinion (see **Appendix 5.C**).

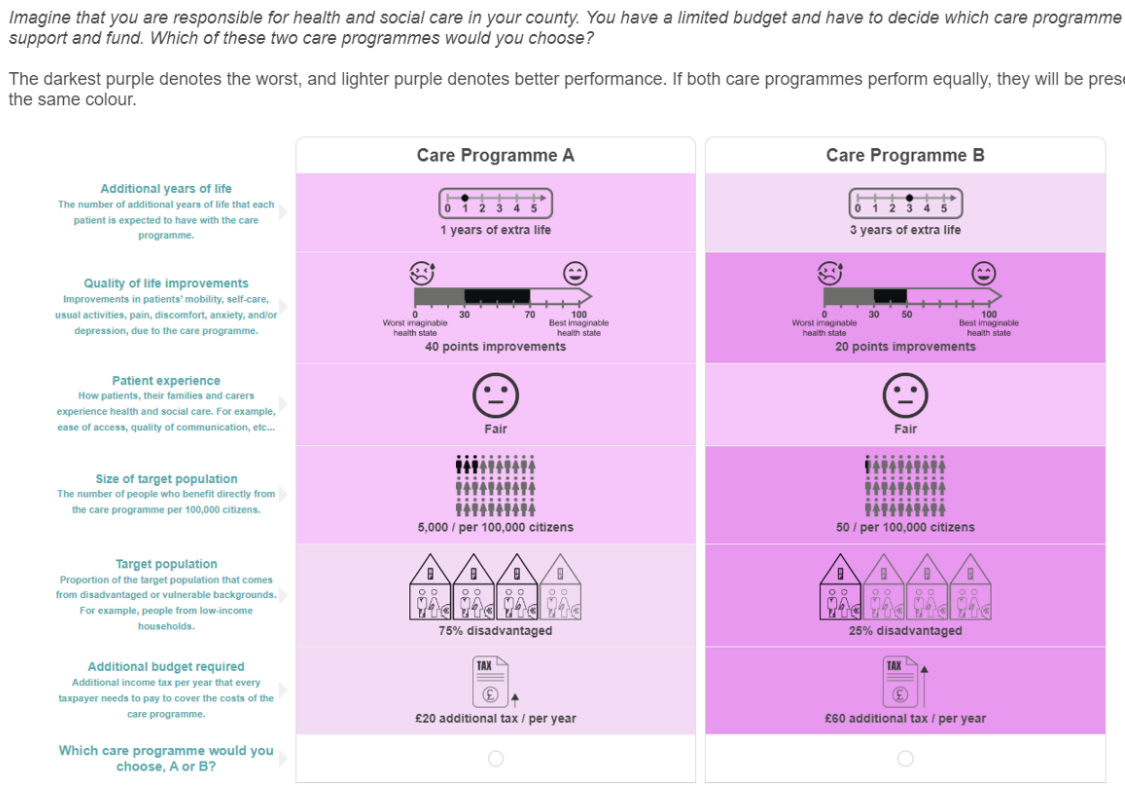
[33] The GP patient survey, run by Ipsos MORI on behalf of the CareQuality Commission and NHS England, includes the question "Overall, how would you describe your experience of your GP practice?" (488, 489).

[34] In the real world, local healthcare commissioners do not have the option to opt-out completely as they are required to respond to the needs of the communities that they serve. By not including the opt-out option, the DCE reflects the actual decision-making context more accurately, ensuring that respondents make choices within the constraints they would encounter in real life situations.

[35] This is the most conventional format used in choice experiments in healthcare, and has been extensively used by previous stated choice experiments to develop priority-setting frameworks (255, 284, 309).

An example of a choice task is presented in **Figure 5.2**. To explore the impact of including cost in a multi-composite value metric on the relative importance of other elements of value, participants were randomly assigned across two DCE versions: one excluding the '*Additional budget needed*' attribute (i.e., DCE-NoAddTax) and the other including it (i.e., DCE-AddTax). Both DCEs were identical in all other respects.

Figure 5.2. Example of choice task



5.2.3. Experimental design

An experimental design contains the attribute levels of all alternatives, of all (unique) choice tasks. It is normally represented by design matrix X, where each row consists of a choice set or choice task, in which each column represents an attribute in an alternative (284).

The number of possible combinations of attributes and levels (i.e., full factorial design) is given by *LA*M*, where *L* is the number of levels, *A* is the number of attributes and *M* is the number of alternatives (283, 284). With six attributes of 3 levels each, a full-factorial would have 531,441 choice tasks (i.e., 3^{6x2}). With five attributes, and 3 levels each, a full-factorial would have 59,049 choice tasks (i.e., 3^{5x2}). As it is virtually impossible to present respondents with the full range of possible choice tasks, two fractional factorial designs were created: The 'NoAddTax' Design included the first five attributes from **Table 5.1**, each one with 3 levels, while the 'AddTax' Design included the attribute "*Additional budget required*" as well.

To create a set of choice tasks that yield as much statistical information as possible for estimating unbiased and precise preference parameters, two Bayesian D-efficient factorial designs were generated in Ngene software (www.choice-metrics.com) (303). D-efficient designs result in data that generate parameter estimates with as small as possible standard errors, improving the likelihood that the parameters of interest can be estimated (258, 304, 305). For this reason, they became the most popular approach to generate stated choice tasks (255). The Bayesian D-efficient design is increasingly considered a state-of-the-art approach as it allows for the use of prior parameter random distributions, and result in less loss of information when the true parameter values deviate from the informative local priors (306-308).

The two heterogeneous experimental designs[36] were created using results from a pilot study with 40 respondents from the general public (see **Appendix 5.D** for more details). Each design included 28 rows divided into two blocks, to which participants were randomly allocated, resulting to fourteen choice tasks in each block. A similar number of choice task has been used in previous similar studies

[36] A heterogeneous design is constructed by first explicitly creating design matrix X and then splitting it into two or more parts known as blocks. These designs are generally assumed to be a better choice as they provide more information (284).

(255, 291, 309, 310). To reduce task complexity and improve choice consistency, attribute level overlap was used, with two attributes at the same level in each choice task (302). To this purpose, partial profile designs were used by creating candidate sets in Stata (311)[37].

Although in the pilot study interaction terms were not included, they were reconsidered as a way to assess the preference independence principle of the attributes (48, 161) (see **Chapter 4**). To allow for the possibility of estimating interactions of interest, a model averaging approach was adopted (312). It involves calculating the asymptotic variance covariance matrix for two or more models that might be estimating post-data collection, and minimising a weighted average measure of efficiency across these models[38]. For each one of the designs two models were averaged: model one is a Bayesian D-Efficient that uses priors from the second pilot (M1), and model two (M2) is a D-efficient design with fixed priors close to zero, indicating the expected direction of the coefficients, and with interactions between '*Improvements in quality of life*' and '*Patient experience*', and between '*Additional budget required*' and the first four attributes, separately (i.e., '*Additional years of life*', '*Improvements in quality of life*' '*Patient experience*' and '*Size of target population*') (307). The interaction between '*Improving quality of life*' and '*Patient experience*' is potentially relevant, as it could be argued that improvements in patient experience could ultimately lead to improvements in health-related quality of life. Similarly, more expensive care programmes could potentially be associated with better health and non-health value outcomes. Interactions between '*Additional budget required*' were only added for the 'AddTax' Design. When using the model averaging approach, the efficiency measure for model M1 was multiplied by 1.5 in the Ngene syntax (303, 312).

[37] When constructing the candidate set, choice sets with strong dominant alternatives were rejected. Only choice sets with attribute levels overlap on two attributes were used. See **Appendix 5.E.** for more in detail information on the construction of the candidate set.

[38] The model averaging approach was proposed by Rose and colleagues (2009) to address the issue of how to generate efficient stated choice experiments under uncertainty over the final model that will be estimated on data collected using the design. "*The model averaging process provides a degree of robustness if the analyst is unsure as to exactly what model will be estimated post data collection*" (312).

Following the syntax used in the pilot study, in the 'AddTax' Design, all attributes but "*Patient experience*" were treated as numerical. To assure more than one choice sets with overlap on the different attributes, in the 'NoAddTax' Design "*Patient experience*", "*Quality of life improvements*" & "*Additional years of life*" were treated as categorical, and the other two attributes as numerical. Attribute level constraints were also imposed in all attributes to guarantee a degree of attribute level balance[39]. Both designs had a low D-error, and a relatively even distribution of movements between levels for each attribute. A summary of the resulted designs is presented in

[39] Ngene does not use often the middle levels of numerical attributes since this is less efficient (making larger trade-offs between extreme levels provides more information) (303).

Table 5.2. **Appendix 5.E** summarises more in detail the experimental designs created.

Table 5.2. Characteristics of experimental designs

Design	Interaction terms added to M2	Degrees of freedom	Overlaps						M1	M2
			YoL	QoL	Exp	Size	Equ	AddTax		
AddTax Design ●M1: Bayesian D-efficient.All as continuous except for "Exp" which is treated as categorical. Priors from pilot study. ●M2: D-Efficient design. Priors close to zero (0.0000001) Interactions: - Exp & QoL - AddTax & Size - AddTax & YoL - AddTax & QoL - AddTax & Exp	●Exp.dummy[1]*QoL Prior: 0.000001 ●Exp.Dummy[2]*QoL Prior: 0.000001 ●Exp.dummy[1]* AddTax Prior:0 ●Exp.dummy[2]* AddTax Prior:0 ●YoL*AddTax Prior:0 ●Qol* AddTax Prior:0 ●Size* AddTax Prior:0 "QoL" as continuous and "Exp" as categorical in the interactions.	14	11	8	6	10	12	9	D-error: 0.0014 A-error: 0.2767	D-error: 0.0002 A-error: 1.3777
NoAddTax Design ●M1: Bayesian D-efficient."YoL", "QoL" and "Exp" are treated as categorical. "Size" and "Equ" are treated as numerical Priors from pilot study. ●M2: D-Efficient design. Priors close to zero (0.0000001) Interactions: Exp & QoL	●Exp.dummy[1]*QoL.dummy[2] Prior: 0.000001 ●Exp.dummy[2]*Qol.dummy[1] Prior: 0.000001 "QoL" as continuous and "Exp" as categorical in the interactions.	17	7	13	14	12	10	.	D-error: 0.0219 A-error: 0.7523	D-error: 0.0392 A-error: 0.7532

Note: To be able to estimate *K* parameters (*K*=14 for Design 1 and *K*=11 for Design 2), it must hold that $|S| * (|I| - 1) \geq K$, where $|S|$ denotes the number of choice tasks (i.e. rows) and $|I|$ denotes the number of alternatives (i.e., Care programme A & Care programme B). The difference between the actual number of choice tasks in the design and the minimum required design size is referred to as the degrees of freedom (284).

5.2.4. Survey instrument

The survey was administered online and consisted of eight sections. The first one corresponded to the landing page, which welcomed participants and inform them on the length of the survey. Section 2 included five screening questions and Section 3 presented information about the survey, including the need for decisions on which care programme should be funded, the privacy notice and the informed consent. A description of the attributes and corresponding levels was provided in Section 4. For each attribute, participants were asked to answer warm-up tasks to give their view on the attribute for a hypothetical situation. This was followed by two practice questions, with the second one being a dominant test used to disqualify those that fail it (Section 5), and the 14 choice tasks (Section 6). At the end of the choice tasks, participants were asked two debriefing questions (Section 7), followed by Section 8 which contained questions about the respondent's educational level, working sector, and health (measured by the EuroQoL 5D-5L (261) and a multiple choice question about their morbidities), and health problems that respondents or respondents' relatives might have experienced[40]. The survey instrument used can be found in **Appendix 5.F**.

5.2.5. Data collection, recruitment strategy and sampling

The DCE survey was completed by adult members from the general public in England. The online version of the survey was created by Dynata, a market research company. The same company was used to collect the data. National census balanced quotas and a target recruitment strategy were used to ensure that the sample was representative of the general population in England in terms of regions, gender, age and ethnicity[41]. Based on this quota sampling strategy, Dynata targeted invitations to panel members

[40] For the question on the highest educational or school qualification, the International Standard Classification of Education (ISCED) was used as a reference (221). It was used in the Annual Population Survey - Labour Force Survey (415). For the question on the working sector, the standard industrial classification of economic activities was used, which is also used in the Labour Force Survey. The question on the health problems is also included in the DCE conducted by Rutten-van Mölken et al (2020), covering eight European countries, including the UK (293). This question was however expanded by asking participants whether they have experienced any of the health problems listed in their families or as carers as it can potentially affect people's preferences for outcomes of care.

[41] For the quota sampling interlocked nationally representative quotas on age, gender and region were used. On ethnicity, non-interlocked quotas were applied (348-349).

and took actions to ensure that respondents included in the final sample met all quality checks. **Appendix 5.G** summarises the census balanced quotas used.

To estimate the minimum sample size requirements, the parametric approach proposed by de Bekker-Grob et al. (286) was followed, using data from the pilot study. For the design that includes the monetary attribute ("AddTax" Design), a minimum sample size of 166 respondents was needed, with a statistical power of 0.8 and assuming a confidence level of 95%. For the "NoAddTax" Design, in which the monetary attribute was excluded, the minimum required sample size was 120 respondents. Based on these results and typical sample sizes in the health literature (258, 297), the study aimed to collect 200 responses for each design (i.e., 400 in total). As explained before, each design was split into two blocks, for each participant to respond to 14 choice tasks. In addition, to account for potential left-right bias, 'mirror blocks' were created by reversing the order in which profiles for 'Care programme A' and 'Care programme B' were displayed (313, 314). In the original block, the profiles for 'Care programme A' were on the left and the profiles for 'Care Programme B' were on the right side. In the mirror block, the experimental design is the same, but profiles of Care programme A were on the right. The 400 sample of respondents was therefore randomized across the following eight blocks (see **Appendix 5.E.**):

- Block 1 – 1st original block of Design 1: 50 respondents
- Block 2 – 2nd original block of Design 1: 50 respondents
- Block 3 – Mirror block of Block 1: 50 respondents
- Block 4 – Mirror block of Block 2: 50 respondents
- Block 5 – 1st original block of Design 2: 50 respondents
- Block 6 – 2nd original block of Design 2: 50 respondents
- Block 7 – Mirror block of Block 5: 50 respondents
- Block 8 – Mirror block of Block 6: 50 respondents

The data of the main survey was collected between the 1st of November and the 1st of December, 2022.

5.2.6. Data quality

When conducting choice experiments it is crucial to ensure the validity and reliability of results, and with this, add to the quality of responses (299, 315). Reliability of choice experiments refers to the degree of reproducibility of the results, while validity resorts to the extent to which the method truly measures participants' preferences (315-317). The literature distinguishes between external and internal validity. The second one focuses on the core assumptions of the choice experiment while external validity refers to the generalizability of the results to other circumstances (315).

Multiple strategies have been suggested to test reliability of choice experiments. The test-retest approach, for instance, compares the consistency of participants' responses for the same choice tasks by using the same survey tool at two different points in time (315, 317). It therefore requires administering the survey twice, which can be very costly. Another approach consist of exposing participants to extensive information about the experiment, and one could compare participants understanding between those with relatively low exposure to information and those with greater exposure (317, 318). A version of a consistency test can also be applied by asking participants to respond to different versions of the same choice experiment and compare the extent to which responses vary between both versions. This can be done by introducing small changes to the survey version (in the background scenario, in the framing of the question or in the attributes or levels used), although it requires running the experiment twice, which again increases the costs (315, 317, 319-321). Other option, in the context of blocked experimental designs in which different respondents see different combination of choice tasks, involves adding one or more fixed-choice tasks to the design. Responses to this fixed choice task can then be compared across survey blocks by using a simple t-test (315, 322). Despite all suggested tests, due to the variation on participants' responses, the measurement of reliability will never be perfect in any situation (315).

The hypothetical nature of the choice scenarios used in choice experiments makes difficult the validity of stated-preferences (315). Nevertheless, there is a call for assessing the validity of choice experiments in healthcare and guidelines in the field consider it a good practice (259, 299, 315). External validity is generally examined by comparing estimated preferences from the choice experiment with the choices people make in the real world (315). For internal validity, studies have typically explored whether respondents' choices do not deviate from a priori expectations (face or theoretical validity) (315). The non-satiation (dominance) test, which consists of adding a choice task in which one of the alternatives has no worse levels for any of the attributes and better attributes for at least one of them is one of the most commonly (315, 323). Adding a repeated choice task (i.e., in an experiment with 8 choice tasks, adding a ninth task that equals the second or the third choice task), aiming at testing for stability of preferences, has also been widely used to test internal validity (317, 324, 325). Internal validity also refers to the extent to which the attributes and levels of the choice experiment take account of all things deemed important in the constructs' domain (known as content validity) (315). To assess this, studies have mostly used debriefing questions at the end of the experiment, or qualitative techniques to assess respondent behaviour or perceptions (317).

Despite the wide applicability of some of these tests, there are no gold-standard tests to determine the reliability or true validity of participants' responses (325). These tests also come with limitations as most of them require adaptation to the study design. Some tests, such as the version consistency test, require administering the survey twice, which increases the cost of the study. Other tests require adding one or more choice tasks to the study, making the survey longer. It is also crucial to take into account good research practices on stated-preference studies as data quality is not guaranteed by focusing only on reliability and validity tests (255, 281, 299). Aspects such as respondent's understanding or having a diverse sample are crucial to ensure the quality of the collected data (315).

In light of this, for the experiment conducted in this chapter, various strategies were employed to enhance data quality. To ensure content validity, the attributes, the definitions used, the levels and the choice task format were extensively discussed in the meetings with the PPI group (255, 281, 326, 327). Attribute definitions were refined based on feedback from these meetings and the test conducted at the BRC Open Day (299). Prior to responding to the choice tasks, participants were provided with a detailed narrative of the experiment, including practice questions for each attribute to enhance their understanding. Two practice choice tasks were also added to the survey to familiarise participants with the experiment. Before starting to respond to the choice tasks, respondents were encouraged to provide honest answers and explain the relevance of their participation in this study (328, 329).

In terms of the framing of the choice task, graphics for attribute levels were used as they may encourage respondents to be more attentive to the attribute levels and the choice task as a whole (330-333). To reduce task complexity and improve choice consistency, level overlap and colour coding were used, following suggestions from Jonker and colleagues (2018) (302). For level overlap partial profile designs were applied to ensure that each choice task had overlap on two of the attributes. The attribute levels were colour-coded using shades of purple, with the darkest shade representing the worst level and lighter shades representing better levels (302). Colour blindness simulator (www.color-blindness.com) was used to check the different potential visualisations, based on all possible colour vision deficiencies.

To ensure a sample representative of the general population of England in terms of age, gender, ethnicity, and region, stratified quota sampling was used. Additionally, the order of tasks was randomized to control for learning curves and fatigue (258). Three quality checks were applied during the data collection process: (i) Respondents who completed the survey in less than one third of the median time (i.e., speeders) were disqualified (323). (ii) Those who chose the same alternative across all 14 choice tasks (i.e., straight liners) were also disqualified. (iii) Moreover, those who failed the dominant test presented as the second practice choice task in the survey were disqualified.

To assess internal validity, a root likelihood (RLH) test was used, even though repeated and dominant choice tasks are more commonly employed (317, 325). A recent study by Jonker et al. evaluated the sensitivity and specificity of repeated and dominant choice tasks, finding them to be unreliable screening tests that diminish statistical power (325). The authors compared the performance of the internal validity tests with that of the RLH test, a commonly used likelihood-based statistical validity test, and found that the performance of the RLH test was considerably better than that of the dominant and repeated choice tasks. The RLH test is a likelihood-based statistical validity test that reflects the fit of the choice model on the respondents' choices (325, 334). It is defined as the geometric mean of the respondents' likelihood, with the RLH being higher for respondents whose utility model fits their observed choices well and lower for those whose utility model fits their choices poorly (325, 335).

Relatively few studies in healthcare have examined external validity due to the limited opportunities to observe people' preferences through real-life healthcare choices (315). Since participants of the experiment are people from the general public in England, and there is therefore not data on real-life choices, it was not possible to test for external validity.

5.2.7. Data analysis

Characteristics of the respondents were summarised with descriptive statistics. To model participant's choices for each one of the designs, a random utility model was assumed. In this model, the two alternative care programs, A and B, are characterized by a utility function comprising both a deterministic and a random component (257). For each respondent n, the utility function U_{ni} of alternative i is a random variable based on attributes which influence individual's behaviour (i.e., deterministic or systematic component), V_{ni}, and a stochastic disturbance term ε_{ni} (i.e., random component). The later measures the deviation from the modelled utility for alternative *i* and respondent

n, and is assumed to be identically and independently distributed, leading to the multinomial logit model (MNL):

$$\text{DCE-NoAddTax: } U_{ni}^{NoAddTax} = V_{ni}^{NoAddTax} + \varepsilon_{ni}^{NoAddTax} \; where \; i = A, B \quad (1.a)$$

$$\text{DCE-AddTax: } U_{ni}^{AddTax} = V_{ni}^{AddTax} + \varepsilon_{ni}^{AddTax} \; where \; i = A, B \quad (1.b)$$

The deterministic part of the utility, V_{ni}, is typically assumed to have an additive structure defined by the *K* attributes of the alternatives and the corresponding estimated parameters β, as follows:

$$V_{ni}^{\text{NoAddTax}} = \sum_{k=1}^{5} \beta_{nX_k}^{NoAddTax} X_{kni} = \; \delta_{10} + \beta_{11} YoL + \; \beta_{12} QoL + \; \beta_{13} Exp + \beta_{14} Size + \; \beta_{15} Eq \quad (2.a)$$

$$V_{ni}^{\text{AddTax}} = \sum_{k=1}^{6} \beta_{nX_k}^{\text{AddTax}} X_{kni} = \; \delta_{20} + \beta_{21} YoL + \; \beta_{22} QoL + \; \beta_{23} Exp + \beta_{24} Size + \beta_{25} Equ + \; \beta_{26} AddTax \quad (2.b)$$

With δ_{10} and δ_{20} defining the intercepts or alternative-specific constant indicating the propensity of participants to choose Care programme A over Care programme B, and expected to be not significant in the model because of the mirror blocking used in the experimental design. The estimated parameters $\beta_{11:15} \; and \; \beta_{21:26}$ capture the marginal sensitivity to changes in the attributes (see **Table 5.1**).

The data collected, with the choices made by each individual across the different scenarios (*s*), $Y_{ns} = i$, was linked to the associated utilities by assuming that individuals follow a rational behaviour and therefore choose alternative *i* if it delivers the highest utility in comparison with the utility associated with the other alternative *j* (259, 283). Thus, the probability of choosing alternative *i* was modelled, with the probability of choosing *i* increasing with deterministic component of the utility V_{ni}:

$$P_{ni}^{\text{NoAddTax}} = P \; (V_{ni}^{\text{NoAddTax}} + \; \varepsilon_{ni}^{\text{NoAddTax}} > V_{nj}^{\text{NoAddTax}} + \; \varepsilon_{nj}^{\text{NoAddTax}}) \quad (3.a)$$

$$P_{ni}^{\text{AddTax}} = P \; (V_{ni}^{\text{AddTax}} + \; \varepsilon_{ni}^{\text{AddTax}} > V_{nj}^{\text{AddTax}} + \; \varepsilon_{nj}^{\text{AddTax}}) \quad (3.b)$$

Where $i = A, B \ \ and \ j = A, B$. After normalising the variance of the error terms (257), the probability of choosing *i* takes the following MNL form:

$$P_{nis}^{NoAddTax} = \frac{e^{\mu V_{nis}^{NoAddTax}}}{\sum_{j=1}^{J} e^{\mu V_{njs}^{NoAddTax}}} \quad (4.\text{a}) \quad ; \quad P_{nis}^{AddTax} = \frac{e^{\mu V_{nis}^{AddTax}}}{\sum_{j=1}^{J} e^{\mu V_{njs}^{AddTax}}} \quad (4.\text{b})$$

Where μ is the scale parameter (inverse of the standard deviation of the disturbance), which cannot be separately identified in the standard MNL model and by convention is set to unity (259). The denominator is the sum across all alternatives (i.e., Care programme A and Care programme B), so that $0 \leq P_{nis} \leq 1 \ \forall j$ and $\sum_{j=1}^{J} P_{njs} = 1$. To assess the functional form of the attributes, the MNL model was estimated using dummy coding for all attributes (dummy-variable-coded conditional logit model). The results were visualised by plotting the estimated coefficients (i.e., part-worth utilities) against the attribute levels (298).

After estimating a MNL model for each design, observed heterogeneity was explored. To do so, interactions with respondent's characteristics were included based on prior empirical evidence and their impact on the model's fit (296, 313). Interaction terms attributes were also used between to explore whether attributes are preferentially independent. As in the experimental design stage, interactions between 'QoL' and '*Exp*', and between '*AddTax*' and the first four attributes (i.e. '*YoL*', '*QoL*', '*Exp*' and '*Size*') were tested (307).

To account for random variation across respondents and relax the preference homogeneity assumption, mixed multinomial logit (MMNL) models were estimated, with all parameters set as random and normally distributed (295, 336). The MMNL model has been the most popular model used in health-related DCEs over the last years (255). In these models, continuous statistical distributions were used to capture unobserved heterogeneity, by assuming that sensitivity, β_{nX_k}, vary across individuals, such

that $\beta_{nX_k} = \overline{\beta_{X_k}} + \Delta\beta_{nX_k}$. Hence, the probability of choosing *i* is a function of the attribute levels that characterize the alternatives (i.e., Care programme A and Care programme B) and a random error term that adjusts for individual-specific variations in preferences (298):

$$P_{nis}^{NoAddTax} = \frac{e^{V_{nis}^{NoAddTax}(\widetilde{\beta_n}, X_k)}}{\sum_{j=1}^{J} e^{V_{njs}^{NoAddTax}(\widetilde{\beta_n}, X_k)}} \quad (5.a) ; \qquad P_{nis}^{AddTax} = \frac{e^{\mu V_{nis}^{AddTax}(\widetilde{\beta_n}, X_k)}}{\sum_{j=1}^{J} e^{\mu V_{njs}^{AddTax}(\widetilde{\beta_n}, X_k)}} \quad (5.b)$$

Where $\widetilde{\beta_n} = f(\beta, \sigma | v_n)$, with β and σ being the parameters to be estimated on the basis of systematic variations in preferences across respondents, given the variable v_n characterizing individual-specific heterogeneity. The distribution of $\widetilde{\beta_n}$ across respondents is determined by the continuous function $f(\cdot)$, and it is commonly assumed to be normally distributed with mean β and standard deviation σ (298). Hence, the MMNL models estimate both a mean effect and a standard deviation of that effect across respondents. 1,000 Modified Latin Hypercube Sampling (MLHS) draws per individual (337) were used for the simulation of the choice probabilities[42]. Although latent class models are also used to examine heterogeneity, the MMNL was considered more appropriate. Both models are used to allow for random heterogeneity, with each one of them having its imitations and virtues (338)[43].

To determine the relative "*weight*" of the attributes (i.e., MCDA assessment criteria weights), relative importance (RI) scores were calculated following Gonzalez (2019) (339) and Genie et al (2021) (310), using the equation below:

$$RI = \frac{Max(pwu_k) - Min(pwu_k)}{\sum_k (\max(pwu_k) - Min(pwu_k))} * 100 \quad (6)$$

[42] The studies that have used mixed logit models to account for preference heterogeneity used on average 1354 draws (median 1000, minimum 50, maximum 10,000), with 53 studies (out of 65) assuming that parameters followed the normal distribution (255).

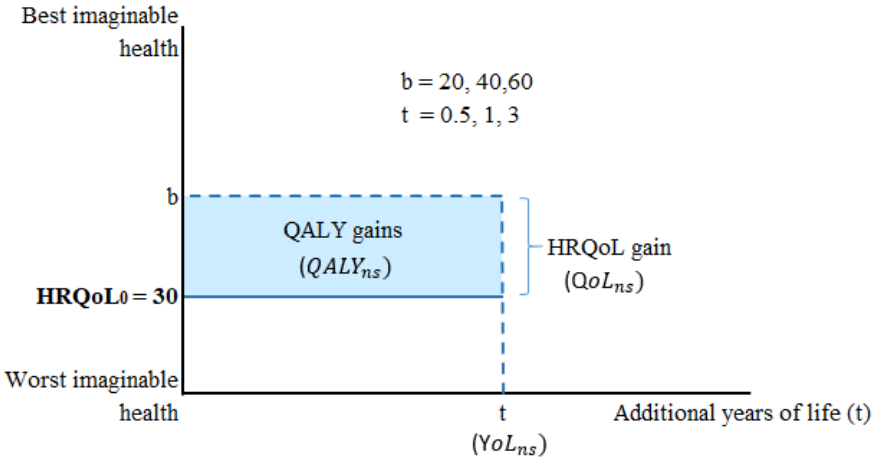
[43] The latent class model is also a Mixed logit model, but assumes discrete random heterogeneity. It therefore aims to identify subgroups of respondents with similar preferences. Since the latent class model approximates the underlying continuous distribution with a discrete one, it is considered to be less flexible than the MMNL model. Nevertheless, latent class models are semi-parametric models and as such free the analyst from unwarranted distributional assumptions (388).

Were pwu_k corresponds to the coefficients for the attribute *k*. To obtain a 95% confidence interval around the RI scores, a bootstrapping procedure with 1000 replications was used (310)[44].

To estimate the relative importance of 'QALY gain' compared with the other elements of value, the survival attribute (i.e., 'YoL') was combined with the health-related quality of life attribute (i.e., 'QoL'). Traditionally, to calculate QALYs, the time in a specific health state is multiplied by a score representing the value of that specific health state (with 0 representing death and 1 full health) (216). The value of health states is estimated using valuation methods such as the EQ-5D visual analogue scale (VAS), which was used in this experiment's descriptive system to describe the three levels of the 'QoL' attribute (261, 342). Since 'QoL' and 'YoL' had three levels each, the 'QALY gains' variable consisted of seven values ranging from 0.1 (i.e., 0.5 additional years of life and 20 points improvement in 'QoL') to 1.8 (i.e., 3 additional years of life and 60 points improvement in 'QoL'). An illustration of the combination of the years of life and quality of life attributes to generate QALYs is presented in **Figure 5.3**.

[44] Other approaches to investigate the relative impact of attributes includes calculating how much each attribute contributes to the overall log-likelihood of the choice model (i.e. partial log-likelihood). This approach, however, requires having an orthogonal data set to run independent estimation of all main effects (259). Other approach involves including in the estimation all main effects and all two-way attribute interactions for attributes X_1 to X_K, and then calculating the marginal rates of substitutions (MRS). By doing this, the MRS is not simply the ratio of two attribute parameter estimates since each attribute is entering the utility function both, linearly and multiplicatively (ibid). The present experimental design of this DCE, however, did not account for all two-way attribute interactions. In addition, with this approach is not possible to measure the relative weight for the monetary attribute 'AddTax', as it is used as the common base. For the same reason, the approach that uses the Hicksian compensating variation was not considered as it also requires WTP or accept compensation for changes in a given attribute (264).

Figure 5.3. Illustration of generating QALY gains from two attributes



To obtain people's WTP for gains in 'YoL', 'QoL', 'Exp', 'Size' and 'Equ', MMNL models in a WTP space were estimated (343). The WTP is the MRS between the value attributes, with the monetary attribute (i.e., 'AddTax') being in the trade-off so that the MRS can be expressed in monetary terms (£). In this context, WTP estimates indicate the amount of additional income taxes individuals are willing to pay in order to get additional gains in the value attribute of interest (313). As an example, the equation below indicates how the WTP for gains in longevity (i.e., YoL) (344):

$$MRS_{AddTax,YoL} = WTP_{YoL} = \frac{\beta_{YoL}}{\beta_{AddTax}} \qquad (7)$$

Where β_{YoL} and β_{AddTax} are the coefficients for 'YoL' and 'AddTax' in the choice model, respectively. MRS with 'YoL' and 'QoL' in the denominator were also used to compare results between the two DCE sub-samples (i.e., DCE-NoAddTax vs DCE-AddTax).

A significance level of 5% was used to determine statistically significant coefficients and standard deviations. Measures of goodness-of-fit were analysed to compare models. A RLH test was conducted to check on the fit of the choice model on the respondents' choices (325). All data analyses for DCE were performed in Apollo software (346-347).

5.3. Results

5.3.1. Respondents' characteristics

The survey was completed by 402 respondents via an online panel (201 in each one of the DCEs). The median completion time in DCE sub-sample with five attributes (i.e., DCE-NoAddTax) was 10.2 minutes and in the sub-sample with six attributes (i.e., DCE-AddTax) was 11.4 minutes. Around 20% of the respondents found the survey difficult or very difficult (NoAddTax: 18.4%; AddTax:19.9%). **Table 5.3** summarises the socio-demographic characteristics of the respondents. The total sample was representative of the English population in terms of gender, age, region and ethnicity. The two DCE sub-samples were fairly similar, with people from the 'NoAddTax' DCE reporting few more co-morbidities and a slightly lower EQ-5D-5L index or VAS-based quality of life score. None of these differences, however, was statistically significant (p-value< 0.05).

Table 5.3. Socio-demographic characteristics of the respondents

	DCE - NoAddTax attribute n=201	**DCE- AddTax attribute n=201**	**Total sample**	**National statistics(a)**
Gender (n, %)				
Male	95 (47.3%)	95 (47.3%)	190 (47.3%)	49.0%
Female	105 (52.2%)	105 (52.2%)	210 (52.2%)	51.0%
Other	1 (0.5%)	1 (0.5%)	2 (0.5%)	
Age (n, %)				
18-24 years	23 (11.4%)	18 (9.0%)	41 (10.2%)	10.6%
25-34 years	31 (15.4%)	35 (17.4%)	66 (16.4%)	17.1%
35-44 years	34 (16.9%)	33 (16.4%)	67 (16.7%)	16.2%
45-54 years	40 (19.9%)	35 (17.4%)	75 (18.7%)	16.9%
55-64 years	33 (16.4%)	30 (14.9%)	63 (15.7%)	15.7%
65+ years	40 (19.9%)	50 (24.9%)	90 (22.4%)	23.5%
(Mean ; sd)	(47.41 ; 16.08)	(48.38 ; 16.31)		
(Median ; IQR)	(49 ; 34-60)	(49 ; 34-64)		
Region (n, %)				
South East	31 (15.4%)	34 (16.9%)	65 (16.2%)	16.3%
South West	21 (10.5%)	21 (10.5%)	42 (10.4%)	10.2%
London	30 (14.9%)	33 (16.4%)	63 (15.7%)	15.6%
East of England	27 (13.4%)	18 (9.0%)	45 (11.2%)	11.1%

West Midlands	20 (10.0%)	23 (11.4%)	43 (10.7%)	10.5%
East Midlands	13 (6.5%)	20 (10.0%)	33 (8.2%)	8.7%
Yorkshire & the Humber	21 (10.5%)	19 (9.5%)	40 (10.0%)	9.8%
North West	27 (13.4%)	25 (12.4%)	52 (12.9%)	13.0%
North East	11 (5.5%)	8 (4.0%)	19 (4.7%)	4.8%
Ethnic group (n, %)				
White	171 (85.1%)	170 (84.6%)	341 (84.8%)	84.2%
Mixed / multiple ethnic groups	3 (1.5%)	3 (1.5%)	6 (1.5%)	1.9%
Asian or Asian British	15 (7.5%)	19 (9.5%)	34 (8.5%)	8.3%
Black, Black British, Caribbean or African	10 (4.5%)	7 (3.5%)	17 (4.2%)	3.7%
Other ethnic group	2 (1.0%)	2 (1.0%)	4 (1.0%)	1.9%
Education (n, %)				
Primary education	1 (0.5%)	1 (0.5%)		
Secondary education	56 (27.9%)	83 (41.3%)		
Short cycle tertiary education	39 (19.4%)	16 (8.0%)		
Bachelor's or equivalent level	69 (34.3%)	67 (33.3%)		
Postgraduate education	35 (17.4%)	32 (15.9%)		
I prefer not to answer	1 (0.5%)	2 (1.0%)		
Health problems (n, %)				
At least one	146 (72.6%)	129 (64.2%)		
Two or more	83 (41.3%)	63 (31.3%)		
In the family	114 (56.7%)	93 (46.3%)		
Carer for a family member	36 (17.9%)	33 (16.2%)		
Utility (EQ-5D-5L) (b)				
(Mean ; sd)	(0.85 ; 0.20)	(0.87 ; 0.19)		
(Median ; IQR)	(0.92 ; 0.82-1)	(0.92 ; 0.83-1)		
Utility (EuroQol VAS) (b)				
(Mean, sd)	(71.4 ; 20.0)	(73.9 ; 19.0)		
(Median, IQR)	(75 ; 60-87)	(79 ; 69-86)		

Notes: (a) For the quota sampling interlocked nationally representative quotas on age, gender and region were used (348). Non-interlocked quotas were applied on ethnicity (349); (b) The EQ-5D-5L utilities were computed using tariffs from the general public in England (350). For both questions, the EQ-5D-5L and the Quality of life –VAS, the official instruments were used (261).

5.3.2. Multinomial logit models (MNL) & observed heterogeneity

Table 5.4 summarises results of the multinomial logit joint estimation models with and without interaction terms between attribute levels and socio-demographics. Observed heterogeneity was independently explored for each one of the DCEs (i.e., "ADDTax" and "NoAddTax") and with a joint estimation, pulling together data from the two DCE sub-samples (345). More than 60 MNL specifications were assessed. Based on dummy-coded multinomial logit models and part-worth utilities (see **Appendix 5.H**), 'Size', 'Exp' and 'Equ' were modelled as a categorical variables. All main-effects coefficients had the expected sign and are statistically significant (p-value<0.05). The model fit was improved when interaction terms and attributes were included in the model (Likelihood ratio test p-value: 0.0002535). Nevertheless, only few interaction terms were statistically significant. In particular, people working on human health and/or social work activities seemed to have a higher preference for a good patient experience than people working on other sectors of the economy in England. On the contrary, people living in the North of England (351), the region with the highest proportion of deprived households, seemed to derive a relatively lower utility from having a good patient experience.

Table 5.4. Joint estimation (No AddTax & AddTax) of multinomail logit models: Main effects & observed heterogeneity

	MNL DCE - Main effects			**MNL DCE - Observed heterogeneity**		
Variable	**Coefficient (s.e)**	**Confidence interval**	**t-ratio**	**Coefficient (s.e)**	**Confidence interval**	**t-ratio**
asc	0.031	-0.020 to 0.08	1.184	0.029	0.022 to 0.080	1.113
	(0.026)			(0.026)		
Additional years of life - YoL	0.513	0.425 to 0.602	11.393	0.425	0.248 to 0.601	4.717
	(0.045)			(0.090)		
Quality of life improvements - QoL	0.021	0.017 to 0.025	10.230	0.030	0.021 to 0.0385	6.735
	(0.002)			(0.004)		
Patient experience - Exp						
Fair	1.075	0.911 to 1.240	12.821	1.076	0.913 to 1.240	12.974
	(0.084)			(0.083)		
Good	1.311	1.126 to 1.497	13.848	1.333	1.140 to 1.527	13.520
	(0.095)			(0.099)		

Size of target population - Size						
5,000 /per 100,000 citizens	0.860	0.695 to 1.025	10.204	0.863	0.697 to 1.030	10.150
	(0.084)			(0.085)		
10,000 /per 100,000 citizens	1.111	0.903 to 1.318	10.485	1.130	0.917 to 1.343	10.392
	(0.106)			(0.109)		
Target population - Equ						
50% disadvantaged	0.313	0.203 to 0.4223	5.577	0.315	0.203 to 0.426	5.530
	(0.056)			(0.057)		
75% disadvantaged	0.404	0.271 to 0.537	5.933	0.360	0.219 to 0.502	4.988
	(0.068)			(0.072)		
Additional budget needed - AddTax	-0.025	-0.030 to -0.020	-9.880	-0.027	-0.033 to -0.021	-8.927
	(0.003)			(0.003)		
YoL * Age				0.002	.-0.002 to 0.005	0.842
				(0.002)		
YoL * Health problems family member				0.023	-0.0904 to 0.136	0.395
				(0.058)		
YoL * Number of health problems				0.004	-0.026 to 0.034	0.283
				(0.015)		
QoL * Age				0.000	-0.000 to 0.000	-2.138
				(0.000)		
QoL * Number of health problems				0.000	-0.001 to 0.001	0.178
				(0.001)		
Exp_Good * Social/Health worker				0.643	0.251 to 1.036	3.211
				(0.200)		
Exp_Good * North Region				-0.322	-0.608 to -0.035	-2.200
				(0.146)		
10,000_Size * London				-0.074	-0.391 to 0.242	-0.461
				(0.162)		
75%_Equ * North Region				0.236	-0.006 to 0.478	1.911
				(0.124)		
75%_Equ * Black				0.186	-0.384 to 0.757	0.641
				(0.291)		
AddTax * Unskilled				0.007	-0.002 to 0.015	1.562
				(0.004)		
Log likelihood (final)		-3416.49			-3399.02	
AIC		6854.99			6842.04	
BIC		6920.35			6972.78	

Note: Total sample: 402 respondents. asc, Alternative-specific constant; s.e., standard error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion. The "Age" variable indicates the age reported by the respondents in the survey (question S1 in the survey). "Health problems family member" is a binary variable equal to 1 for respondents who indicated having a family member with health problems (Q8 from the survey). "Number

of health problems" is a continuous variable indicating the numbers of health problems reported by participants, based on Q7 from the survey. The covariate "Social/Health worker" is a binary variable that equals 1 for respondents that work in the 'Human health & social work activities' sector (Q4). The "North Region" refers to respondents living in the North West or North East of England (S3). Similarly, the "London" variable equals 1 for people living in London and zero otherwise. The "Black" variable equals 1 for respondents that identified themselves as part of the ethnic group "Black, Black British, Caribbean or African" (S4). Finally the "Unskilled" variable is also dummy coded equal to 1 for semi-skilled, unskilled manual workers, state pensioners, casual and lowest grade workers and unemployed with state benefits only (S5).

5.3.3. Mixed logit models (MMNL) & unobserved heterogeneity

Results from the random parameter panel mixed logit (MMNL) models are presented in **Table 5.5.** Results from the RLH test indicate a good fit of the choice model on the respondents' choices[45]. All attributes were treated as categorical to later estimate RI scores. Improvements over the MNL specifications are highly significant (see **Appendix 5.I.**). When the 'AddTax' attribute is not included (i.e., DCE-NoAddTax), all coefficients have the expected sign and are statistically significant. The derived standard deviation of most of the random coefficients is highly significant (p-value<0.05), indicating the existence of heterogeneity across respondents around the mean parameter estimate.

[45] The RLH test reflects the fit of the choice model on the respondents' choices. As explained earlier, it can be defined as the geometric mean of the respondents' likelihood (the sth root of the likelihood (L) with s denoting the number of choice tasks per respondent, ie., 14 in this experiment) (325, 335). The probability of an RLH value being <= 1/k was computed, with k denoting the number of alternatives per choice tasks. Hence, respondents with a RLH<=0.5 were considered as bad quality respondents (the individual-level root likelihood value is defined as RLH_n=(L_n)^(1/s)). In all models, the Pr (RLH ≤ 0.5) is close to zero. In the MMNL DCE – AddTax model, the respondent with the worst average fit across observations had a RLH of 0.368. In both models, less than 10% of the sample failed the test. RI scores (i.e., criteria weights) did not change (no statistically significant difference) when these respondents were removed from the models.

Table 5.5 Parameter coefficients from the mixed logit model (MMNL)

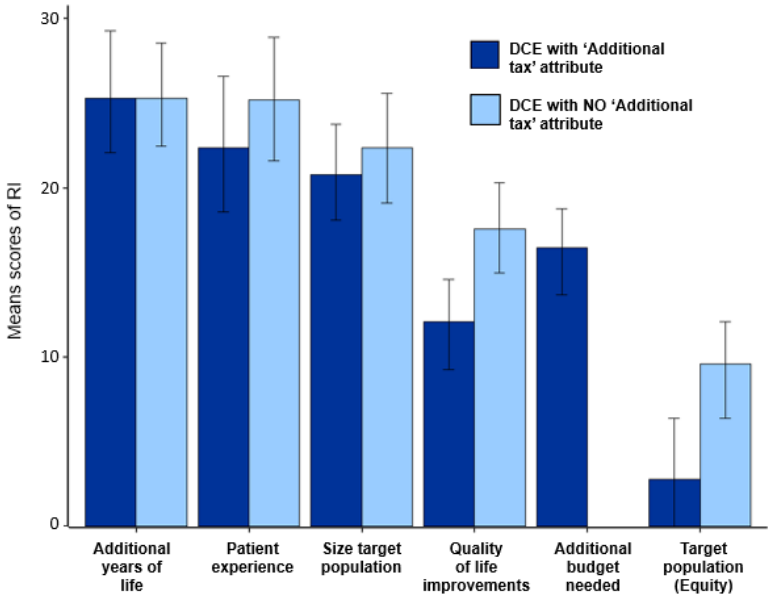
Variable	MMNL DCE - NoAddTax					MMNL DCE - AddTax				
	Coefficient (s.e)	Confidence interval	t-ratio	Standard dev. (s.e)	t-ratio	Coefficient (s.e)	Confidence interval	t-ratio	Standard dev. (s.e)	t-ratio
asc	0.094 (0.054)	-0.010 to 0.199	1.766	0.190 (0.139)	1.330	-0.003 (0.058)	-0.114 to 0.109	-0.049	0.236 (0.137)	1.754
Additional years of life-YoL										
1 year of extra life	0.711 (0.085)	0.537 to 0.885	7.990	-0.025 (0.396)	-0.106	0.731 (0.137)	0.466 to 0.995	5.411	-0.194 (0.310)	-0.678
3 years of extra life	1.891 (0.174)	1.520 to 2.263	9.979	1.100 (0.158)	6.542	2.288 (0.265)	1.791 to 2.785	9.023	1.330 (0.218)	6.058
Quality of life improvements-QoL										
40 points improvement	0.690 (0.134)	0.401 to 0.980	4.681	0.015 (0.207)	0.265	0.742 (0.160)	0.4330 to 1.051	4.706	0.008 (0.176)	0.133
60 points improvement	1.313 (0.14)	1.007 to 1.619	8.416	-0.482 (0.224)	-1.784	1.096 (0.185)	0.686 to 1.505	5.241	-0.603 (0.172)	-3.659
Patient experience-Exp										
Fair	1.610 (0.137)	1.275 to 1.945	9.419	0.615 (0.198)	2.832	1.403 (0.218)	0.998 to 1.808	6.790	-0.911 (0.161)	-5.576
Good	1.880 (0.192)	1.455 to 2.305	8.665	-1.147 (0.209)	-4.686	2.02 (0.206)	1.590 to 2.450	9.213	1.430 (0.174)	7.885
Size of target population-Size										
5,000 /per 100,000 citizens	1.068 (0.140)	0.753 to 1.384	6.637	-0.335 (0.315)	-1.243	1.204 (0.207)	0.866 to 1.542	6.982	-0.033 (0.161)	-0.932
10,000 /per 100,000 citizens	1.672 (0.151)	1.309 to 2.035	9.040	-1.159 (0.145)	-7.424	1.879 (0.232)	1.386 to 2.372	7.471	-1.207 (0.136)	-8.576
Target population-Equ										
50% disadvantaged	0.768 (0.139)	0.495 to 1.040	5.520	-0.565 (0.243)	-2.373	0.311 (0.140)	0.025 to 0.598	2.129	-0.115 (0.316)	-0.535
75% disadvantaged	0.714 (0.111)	0.468 to 0.961	5.674	-0.959 (0.127)	-6.633	0.256 (0.222)	-0.136 to 0.649	1.282	0.329 (0.265)	1.450
Additional budget needed- AddTax										
£40 additional tax / per year						-0.737 (0.144)	-1.005 to -0.468	-5.381	0.047 (0.231)	0.564
£60 additional tax / per year						-1.491 (0.214)	-1.923 to -1.058	-6.754	1.265 (0.143)	8.319

Log likelihood (final)	-1587.49	-1611.79
AIC	3218.98	3275.59
BIC	3349.71	3.43E+03

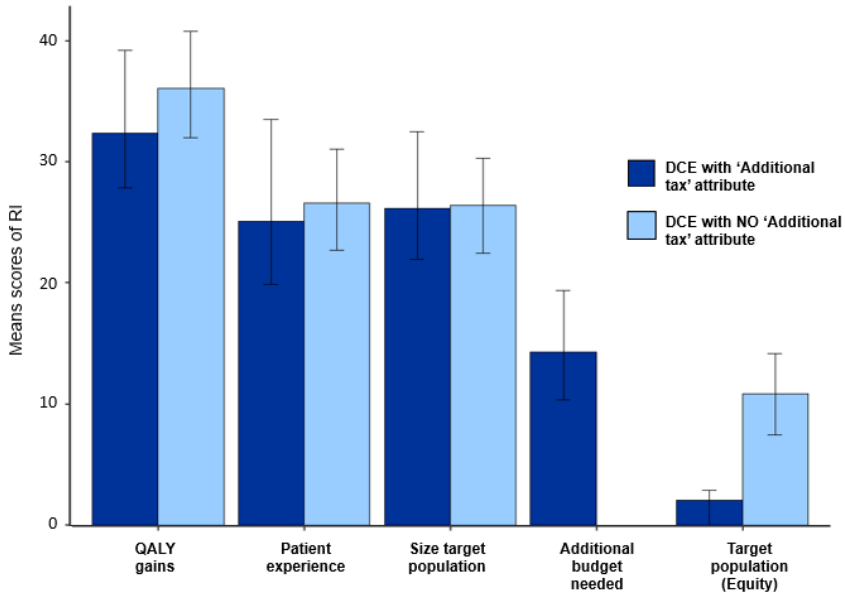
Note: ASC,Alternative-specific constant; s.d.,standard deviation; s.e., standard error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion. The RLH test reflects the fit of the choice model on the respondents' choices. The probability of an RLH value being <= 1/k was computed, with k denoting the number of alternatives per choice tasks. In all models the Pr (RLH ≤ 0.5) was close to zero.

Figure 5.4. Comparison of relative importance (RI) scores*

(a) RI scores for the six attributes



(b) RI scores with QALY gain variable derived from DCE data



*Error bars (95% confidence interval)

Relative importance scores

According to the RI means scores (**Figure 5.4**), people in England assigned the highest values to 'YoL' (25.3%; CI 95%:22.5%-28.6%), 'Exp' (25.2%; CI 95%:21.6%-28.9%) and 'Size' (22.4%; CI 95%:19.1%-25.6%), followed by 'QoL' (17.6%; CI 95%:15.0%-20.3%). The 'Equ' attribute was the least important in this experiment's descriptive system (9.6%; CI 95%:6.4%-12.1%). When the 'AddTax' attribute is present (i.e., DCE-AddTax), 'Equ' is no longer statistically significant (2.8%; CI 95%:-2.1%-6.4%), and there is an equal-sized trade-off between the importance of 'AddTax' and the importance of 'QoL' and 'Equ'. 'YoL' (25.3; CI 95%:22.1%-29.3%), 'Exp' (22.4%; CI 95%:18.6%-26.6%) and 'Size' (20.8%; CI 95%: 18.1%-23.8%) remain as the most important attributes, followed by 'AddTax' (16.5%; CI 95%:13.7%-18.8%) and 'QoL' (12.1%; CI 95%:9.3%-14.6%).

When the 'QALY gains' variable is used in the DCE – NoAddTax model, all parameter estimates have the expected signs and are statistically significant at the 95 percent level of confidence (see **Appendix 5.I**). According to the RI scores (Error! Reference source not found.**b**), 'QALY gains' (36.1%; CI 95%: 32.0%- 40.8%) and 'Exp' (26.6%; CI 95%: 22.7%- 31.1%) were valued more by people, followed by 'Size' (26.4%; CI 95%: 22.5%- 30.3%) and 'Equ' (10.9%; CI 95%: 7.5%- 14.2%). If the 'AddTax' attribute is present, 'QALY gains' (32.4%; CI 95%: 27.9%- 39.2%), Size (26.2%; CI 95%: 22.0%- 32.5%) and 'Exp' (25.1%; CI 95%: 19.9%- 33.5%) remain as the most preferred elements to the general public in England, but the 'Equ' parameter is not statistically different to zero. As before, having to pay additional income taxes for a care programme seems to contribute to patients' choice (14.3%; CI 95%: 10.4%- 19.4%), although not as much as the first three value elements.

Marginal Rates of Substitution (MRS)

Table 5.6 and

Table 5.7 sumarise results of the MRS estimated, based on mixed logit models in which 'Yol', 'QoL' and 'AddTax' are treated as continuous. Willingness to pay (WTP) estimates (i.e., 'AddTax' in the denominator) have the expected sign and are statistically significant for all value attributes but 'Equ'. According to these results, people are willing to pay an additional tax increase of £24 for one additional year of life, per person, in one year. For one point increase in the health-related quality of life VAS scale people are willing to pay an extra £0.8 in income taxes (£16.1 for 20-point increase). The general public also seem willing to pay an additional annual income tax of £45 and £44 for improving patient experience from "poor" to "fair" and for increasing the size of the target population from 50 to 5,000 per 100,000 citizens, respectively.

Results from the MRSs also indicate that people seem willing to trade gains in terms of years of life or quality of life, for improvements in patient experience or size of the target population (e.g., almost 2 years of life or 55 points improvement in quality of life for improvements in patient experience). Differences in MRS estimates expressed in terms of years of life ('YoL') between the two DCE subsamples (i.e., NoAddTax vs AddTax) are statistically significant for all value attributes except for 'Size'. When the denominator is 'QoL', differences in the MRS estimates are statistically significant for 'YoL', 'Size' and 'Equ'.

Table 5.6. MMNL DCE – Add Tax: MRS estimates

		Denominator								
		Additional years of life (YoL)			**Point improvements in quality of life (QoL)**			**Additional tax payment (AddTax) (b)**		
		ratio	**s.e**	**t-ratio**	**ratio**	**s.e**	**t-ratio**	**ratio**	**s.e**	**t-ratio**
Numerator	**Additional Years of Life (YoL)**				31.39	4.24	7.40	£ 24.12	3.36	-7.19
	Quality of life improvements (QoL)(c)	0.86	0.12	7.13				£ 0.80	0.12	-6.51
	Patient experience (Exp_fair - Exp_poor)(d)	1.90	0.23	8.24	54.76	7.43	7.37	£ 44.89	5.95	-7.54

	Size of target population(Size_5000 - Size_50)(d)	1.64	0.22	7.50	44.90	6.44	6.97	£ 43.62	6.92	-6.31
	Target population (Equ_50 - Equ_25)(d)	0.15	0.15	1.04	1.98	4.23	0.47	£ 3.75	3.76	-1.00
	Additional budget needed (AddTax)	-3.70	0.22	-16.55	-0.20	0.19	-1.05			

Notes. (a) Mixed logit models in a marginal rate of substitution space, estimated with the Berndt-Hall-HallHausman algorithm (352). For the simulation of the choice probabilities with inter-individual mixing, 1,000 MLHS draws were used (337). A negative Log-normal distribution is used for 'Additional budget needed' and a Normal distribution for the other attributes. (b) The ratios (i.e., MRS) in which the denominator is the monetary attribute (i.e., 'AddTax') are also known as the WTP. (c) This result suggests that people are willing to pay £16.1 extra tax for a 20 point improvement in QoL. The assumption of linearity was tested by running the mixed logit model in a WTP space with a squared term for the QoL attribute. This term was not statistically significant. (d) Numerator is the estimate for the middle level of each attribute (i.e., change in utility levels between the middle level and the worst (set at 0)). The MRS indicates the change in one attribute that is required to counteract a change in another and keep utility the same. (e) The s.e. indicates robust standard error.

Table 5.7. MMNL DCE – NoAdd Tax: MRS estimates

		Denominator					
		Additional years of life (YoL)			**Point improvements in quality of life (QoL)**		
		ratio	**s.e**	**t-ratio**	**ratio**	**s.e**	**t-ratio**
Numerator	**Additional Years of Life (YoL)**				23.91	2.91	8.21
	Quality of life improvements (QoL)	0.05	0.01	8.19			
	Patient experience (Exp_fair - Exp_poor)	2.31	0.26	8.75	55.27	7.12	7.76
	Size of target population (Size_5000 - Size_50)	1.67	0.23	7.17	41.25	6.36	6.48
	Target population (Equ_50 - Equ_25)	1.15	0.20	5.66	23.24	5.12	4.54

Notes. (a) Mixed logit models in a marginal rate of substitution space, estimated with the BFGS algorithm. For the simulation of the choice probabilities with inter-individual mixing, 1,000 MLHS draws were used (337). A Normal distribution was used for all attributes. (b) Numerator is the estimate for the middle level of each attribute (i.e., change in utility levels between the middle level and the worst (set at 0)). The MRS indicates the change in one attribute that is required to counteract a change in another and keep utility the same. (c) The s.e. indicates robust standard error.

Preference independence principle

To assess the preference independence principle (see **Chapter 4**), all potential interaction terms between attributes using multinomial logit models were assessed. The results are presented in **Table 5.8**. The interaction term between 'YoL' and 'Size' is negative and statistically significant (p-value <

0.05). This indicates that individuals that assign higher importance to gains in longevity, place less importance on changes in the size of target population, between the lowest (50/per 100,000 citizens) and the medium level (5.000/per 100,000 citizens). In contrast, individuals who derive higher utility from improvements in quality of life ('QoL') seem to place greater value on increases in the size of the target population (p-value < 0.05). Additionally, the estimate of the interaction between 'YoL' and 'AddTax' is positive and statistically significant (p-value < 0.05). This indicates that those that value more gains in longevity also place a higher importance on paying additional taxes to cover the cost of care programs.

Table 5.8. Estimates of interaction terms between attributes based on MNL models

Model	Interaction	Estimate	Standard error	t-ratio
MNL1	YoL & QoL	-0.002	0.001	-1.534
MNL2	YoL & Exp_fair	0.087	0.919	0.095
	YoL & Exp_good	0.174	0.329	0.529
MNL3	YoL & Size 5000	-0.173	0.080	-2.164
	YoL & Size 10000	-0.069	0.075	-0.921
MNL4	YoL & Equ 50	-0.036	0.078	-0.463
	YoL & Equ 75	0.107	0.087	1.228
MNL5	YoL & AddTax	0.003	0.002	1.800
MNL6	QoL & Exp_fair	0.003	0.016	0.182
	QoL & Exp_good	0.006	0.012	0.488
MNL7	QoL & Size 5000	0.017	0.006	2.906
	QoL & Size 10000	0.016	0.005	3.035
MNL8	QoL & Equ 50	0.007	0.005	1.371
	QoL & Equ 75	-0.005	0.005	-1.054
MNL9	QoL & AddTax	0.000	0.000	0.805
MNL10	Size & Exp_fair	0.000	0.000	0.249
	Size & Exp_good	0.000	0.000	0.554
MNL11	Equ 50 & Exp_fair	0.043	0.114	0.378
	Equ 75 & Exp_good	0.012	0.167	0.072
MNL12	AddTax & Exp_fair	-0.004	0.017	-0.217
	AddTax & Exp_good	-0.008	0.007	-1.084

MNL13	AddTax & Equ 50	0.000	0.000	-0.873
	AddTax & Equ 75	0.000	0.000	-0.748
MNL14	AddTax & Size 5000	0.001	0.004	0.359
	AddTax & Size 10000	0.003	0.004	0.767
MNL15	AddTax & Equ 50	-0.002	0.006	-0.288
	AddTax & Equ 75	-0.008	0.004	-1.729

Note: MNL models are estimated using the sub-sample DCE AddTax, with 201 respondents.

As there were statistically significant interaction terms between attributes, the same mixed logit models were run while allowing for correlation between the random parameters (295, 345). Results from the correlated mixed logit were similar to the ones reported in **Table 5.5**. No significant improvement in the model fit was found (see **Appendix 5.J.**). When calculating the RI scores, the results showed the same attribute ordering for both MMNL and Colleleated MMNL models, with only marginal differences in the RI scores (see **Table 5.9**).

Table 5.9. Comparison of Relative Importance (RI) scores across estimates

Attributes	**DCE - NoAddTax**			**DCE- AddTax**		
	MNL	**MMNL**	**Correlated MMNL**	**MNL**	**MMNL**	**Correlated MMNL**
Additional years of life (YoL)	24.6%	25.3%	25.5%	24.6%	25.3%	25.2%
Patient experience (Exp)	26.9%	25.2%	25.3%	20.6%	22.4%	23.7%
Size of target population (Size)	21.2%	22.4%	22.3%	21.6%	20.8%	21.2%
Improvements in quality of life (QoL)	18.0%	17.6%	17.9%	12.9%	12.1%	11.9%
Additional budget required (AddTax)				16.2%	16.5%	16.8%
Target population (Equ)	9.2%	9.6%	9.0%	4.2%	2.8%	1.2%

5.4. Discussion and Conclusion

This chapter has assessed public preferences for the elements of value or assessment criteria identified in **Chapter 4**. Findings from this study suggest that people attain a relatively high importance to longevity and health-related quality of life, providing therefore reassurance that the use of QALYs, as a bi-composite measure of value, in prioritization decisions in healthcare is meaningful and relevant (216, 279). However, the evidence from this chapter suggests that the value of healthcare is broader than a QALY as it extends to patient experience, and covers distributional aspects regarding the size of benefitted population and its socio-economic vulnerability. This finding is in line with several international and national trends in value-based healthcare. The UK government, in particular, has incorporated the patient experience and equity at the core of ICSs' purpose (7). Previous literature has also shown that people do value benefits of healthcare other than health benefits and are even willing to trade-off health gains for improvements in the experience with care (250, 279, 293).

Moreover, results show that patient experience was the second most important element of value after only 'Additional years of life', with people willing to pay more than £40 in additional annual income taxes for improvements in patient experience. Previous studies have shown that experience with care outcomes do matter to people, although none of these studies have used a unique 'Patient experience' attribute (293, 353-355). In a DCE study conducted in eight European countries, for instance, 'continuity of care' and 'person-centeredness' are used as measures of patient experience. According to results from the UK, respondents do not place these two attributes at the top of the preference ranking but the sum of their RI scores suggests that people value patient experience more than physical functioning (293).

According to results of the experiment, individuals seem willing to trade almost 2 years of life or 55 points improvement in quality of life for enhancements in patient experience. While some studies have shown that patients are willing to trade survival time for better patient experience, particularly at the

end of life[46], the extent of this trade-off in the context of broader healthcare decisions might seem unusually high, raising concerns about the realism of the marginal rate of substitution estimates.

The relatively high value assigned to 'Patient experience' may reflect challenges in the process of care delivery, particularly in terms of long waiting lists, continuity of care or coordination between professionals. The COVID-19 pandemic might have played a role by increasing the number of patients waiting for treatment as well as people's demand for better provider interaction (356). Another explanation may be the rising numbers of people living with multiple conditions (>1 in 4 of the adult population in England) or with a long-standing health problem (37.7% in the United Kingdom). Healthcare needs for these patients are typically more intricate, requiring more complex and coordinated healthcare than healthier patients or those living with single conditions (357, 358). Regardless of the underlying reasons, the high relative importance of patient experience highlights the need for measuring patient experience in the NHS and providing this data to decision-makers. However, patient experience is a broad and complex concept that refers to the entire care delivery journey, and as such, encompasses many dimensions (359). It is therefore crucial to agree on a set of universal patient reported experience measures (PREMs) and standardize the collection of the necessary data to support the systematic monitoring and assessment of care models. Ongoing initiatives to routinely collect data on continuity and coordination of care, responsiveness to patient concerns, the opportunity of care, and professionals-patient communication should be strengthened (360, 361).

Furthermore, 'Quality of life improvements' received a much lower RI than 'Additional years of life', adding to the voices that question the use of QALYs in certain decision-making contexts. The QALY is a composite indicator that does not take into account differences in preferences for its components or distributional issues. Hence, a new model of care that leads to a 0.8 QALY gain is seen as equal in value

[46] Bryce et al (2004), for instance, showed that patients are willing to shorten healthy life expectancy between 9.6 and 11.4 months for better end-of-life care (340). Similar results were found by Rocca et al (2012) (341).

to another model that also results in 0.8 QALY gain. Findings of this chapter suggest that care models and services that increase life expectancy at lower levels of quality of life may generate more value to the population than care models that achieve the same QALY gain but with higher improvement in quality of life and less gains in life expectancy. This is in line with studies showing that, from a societal perspective, some QALY gains due to improvements in longevity are more valuable than QALY gains associated to improvements in health-related quality of life (278, 362, 363).

With regards to equity, findings suggest that it is the least important element of value especially when the 'Additional budget required' was considered as part of the value. This is comparable to the findings of previous stated-preference studies that people display stronger preferences for efficiency or effectiveness over equity (364, 365). The relatively low value assigned to equity contrasts with the high value given to size of the target population, suggesting that the public favours an allocation of healthcare benefits across more people than to favour the most vulnerable populations. At first glance, this does not conform to John Rawls' difference justice principle, under which social and economic inequalities must be arranged so that they work to the greatest benefit of the less fortunate (366). However, this result may also reflect a society that recognises the NHS as a universal health system under which all residents are assured access to health care. As such, people place a higher value on a more equitable distribution of healthcare, and therefore on care models that offer high population coverage.

The relatively low importance attached to 'equity' contrasts with the increasing emphasis placed by NHS England on reducing unwarranted variations in care, and the requirement for ICSs to increase their efforts to address inequalities in access to care and health outcomes, particularly between people from different ethnic and socio-economic groups (12, 294). While it is possible that the way the attribute was framed in the experiment failed to capture all potential dimensions of equity, the discrepancy between public preferences and organisational priorities raises questions about the alignment of health policy with societal values. Formal debriefing of these results with PPI representatives and local

commissioners could be valuable in reconciling both perspectives and better understanding people's views.

Results also reflect the heterogeneity in preferences across the general population in England. Although not many observable socio-demographic characteristics seem to be relevant to explain such heterogeneity, people working on human health and/or social work activities seem to have higher preferences for a good patient experience. This might reflect a more comprehensive understanding of the patient experience concept among health and care workers, or their role and commitment to quality improvements. Results also suggest that people living in the north of England, the region with the highest proportion of deprived households (351), derive a smaller utility from patient experience improvements. Despite the presence of social gradients in quality of care (367, 368), it could be argued that demands for innovation and quality improvements tend to increase with income (369, 370).

Regarding the healthcare costs, results suggest that people care about the additional healthcare costs that a care programme entails. Although this attribute did not receive the highest value and its inclusion did not alter preferences ordering for the other elements of value, it reduced the importance attained to 'Quality of life improvements' and 'Target population' attributes, almost by the same proportion, and had a 'crowding-out' effect on the 'equity' attribute. Differences in the MRS estimates between the two sub-samples (i.e., NoAddTax vs AddTax) also suggest that the inclusion of the monetary attribute changes people's willingness to trade off gains in patient experience, quality of life or equity for improvements in longevity. These results might be partially explained by the cost of living crisis that the country experienced during the time the DCE data was collected (371). From the decision-making process perspective, results of this chapter suggest that a decision framework is incomplete when the preferences for the monetary attribute are ignored. Nevertheless, its inclusion in the value metric reduces the relevance of other elements of value metrics, with some more affected by others.

Both approaches might be relevant and useful to decision-makers, and are consequently incorporated in the decision-support framework, as indicated in **Chapter 4**. The relative weights derived in this study can be used to calculate the cost-per-value ratio (CVR) of new models of care, with the denominator based on multi-attribute benefit scores (i.e., using the results when costs were excluded). When comparing between two alternatives, this approach would make more explicit the efficiency of the care programmes under evaluation. On the other hand, the relative weights of the six elements of value can be used to obtain a single MCDA-based composite measure for all different alternatives, making more explicit the contribution of each element of value (279). Under this approach, decision-makers could also develop a ranking of several competing alternative care models to inform budget allocation decisions (36). Both approaches are suggested as part of decision-support framework that this thesis develops (see **Chapter 4**).

The chapter further investigated potential violations of the preference independence principle (see **Chapter 4**) by assessing all potential interaction terms between attributes in MMNL models and comparing RI scores derived from MMNL and Correlated MMNL models. Although some interaction terms were statistically significant, the preference ordering remained consistent, and differences in RI scores between the two models (MMNL and Correlated MMNL) were minimal. Consequently, the RI scores estimated from the MMNL models are considered suitable to be used as assessment criteria weights in the decision-support framework developed in this thesis for evaluating new models of care in England. Although, based on these results, it can be reasonably assumed that the criteria are preferentially independent of each other, a certain level of mutual preference dependence cannot be completely ruled out.

Strengths & limitations

One of the main strengths of this study is the decomposition of cost from value, expressed in terms of additional income taxes. This study represents the first choice experiment to employ a split-sample design where costs are expressed as an increase in income tax, providing novel insights to the field's

literature. Previous studies using split sample designs have employed a definition based on out-of-pocket expenditure or changes in insurance premiums. The results of these studies suggest that the inclusion of the 'cost' attribute does not affect the preference structure for value attributes (310, 372-375). These studies have investigated patient preferences for the diagnosis and treatment of severe knee injuries (372) and personalisation of chronic pain self-management programmes (310). However, none of the studies have examined preferences from the general public to inform local investment decisions in integrated care. In principle, the findings of this chapter suggest that the payment vehicle used to define the monetary attribute appears not to significantly influence health-related choice preferences (310).

The second strength of this study lies in the rigours methodology followed to conduct the DCE, including (a) two pilot studies; (b) involvement from PPI representatives; (c) a choice task format designed to reduce complexity and improve respondents' engagement; and (d) the use of quality checks used helped us to increase the internal validity.

This experiment has also several limitations. First, preferences of different stakeholder groups for the elements of value were not elicited. However, the sample was representative of the English general population and as such included all types of stakeholders. In addition, a former study exploring patients' and other stakeholders' preferences for outcomes of integrated care did not find statistically significant differences (293). The focus was also placed on preferences from the general population, recognizing that members of the public should have a voice in the priority-setting process in exchange for paying taxes (231). A last limitation concerns the relative weight estimated for 'QALY gains' as it was a construction from the stated preference data set. Future studies could explore the extent to which preferences for QALY gains differ when its elicitation is derived from a choice experiment that directly includes QALY gains into the descriptive system.

Given the timing of the data collection, amidst the lingering effects of the COVID-19 pandemic (e.g. high waiting lists) and a cost of living crisis, there may be concerns about the credibility of the study's findings. While it is not possible to determine the extent to which context drove people's preferences, it is likely that the relatively high importance given to 'Additional years of life' and 'Patient experience' may have been influenced by the unique circumstances that prevailed during the data collection period. The situation in England at the time may also explain the relatively low importance given to 'quality of life improvements'. These findings may raise questions about the usefulness of RI scores (i.e., assessment criteria weights) in future MCDA applications.

Conclusion

Although preferences for maximizing population-level health gains remain highly relevant, results of this chapter highlight the importance of elements of value that are not included in traditional health economic evaluations as they are not captured by conventional value metrics, such as QALY. The relatively high weight assigned to patient experience stresses the importance of collating data on this dimension and adopting economic evaluation approaches that account for the multiple elements of value-based healthcare. Researchers and policy makers should also be cautious when including cost or the monetary attribute in a multi-composite value metric as it has an uneven impact on the other criteria or elements of value.

Understanding the relative importance that the general public gives to elements of value in healthcare (i.e., assessment criteria) allows for the incorporation of societal preferences into the decision-making process. In particular, the scores derived from preferences of the general population in England can be used in the decision-support framework that this thesis develops in order to inform the commissioning of new models of care in England.

Chapter 6. The economic evaluation of new models of care in England using multi-criteria decision analysis: A case study on mental health integrated care services

6.1. Introduction

Evidence on the performance of new models of care is crucial to inform the priority (or investment) decisions that integrated care systems (ICSs) in England have to make (102, 142). A robust theory-based framework, which takes into account the public's preferences and uses routinely collected data, may facilitate the systematic and impartial evaluation of these models. In previous chapters, a decision-support framework grounded in multi-criteria decision analysis (MCDA) was developed, with a set of assessment criteria aligned with ICSs objectives[47] and relative weights that encapsulate the preferences of the general public in England (see **Chapter 4** and **Chapter 5**). To guarantee practical applicability of this decision-support framework, local commissioners should be able to use it with routinely collected data.

This chapter aims to explore the potential feasibility of the decision-support framework developed in this thesis, by evaluating the performance of early intervention in psychosis (EIP) services in Oxfordshire and Buckinghamshire, using electronic health records[48]. Psychosis is a severe mental illness (SMI) characterised by hallucinations and delusions that affect the perception of reality, causing impairment in social functioning and shorter life expectancy. In England, EIP services are specialized multi-disciplinary teams that provide comprehensive treatment to people with the first episode and early

[47] As explained in **Chapter 2**, ICSs aim to improve outcomes in population health and healthcare, tackle inequalities in outcomes, experience and access, and enhance productivity and value for money (102).
[48] The two are counties located under the Buckinghamshire, Oxfordshire, Berkshire West Integrated Care System (BOB ICS). This ICS is responsible for planning and providing joined up health and care services to people who live in the Buckinghamshire, Oxfordshire and Berkshire West area.

psychosis (376, 377). These EIP services can be considered a new model of care, as they are provided in the community by multidisciplinary teams whose goal is to enhance patient recovery and mitigate relapse risks (238, 270). These services are recommended by the National Institute for Health and Care Excellence (NICE) for individuals aged 14 to 65 with psychosis, and are considered one of the cornerstones to achieve some of the mental health objectives outlined by the NHS Long Term Plan (24, 378).

To evaluate EIP services in Oxfordshire and Buckinghamshire, electronic health record data from Oxford Health NHS Foundation Trust (OHFT) is used. To allow for the comparability of the two EIP services under evaluation, a retrospective longitudinal observational study is conducted, using quasi-experimental methods. The results of this study will highlight gaps in current datasets, and explore the suitability of the decision support-framework to support investment decision on EIP services across the BOB-ICS.

6.2. Methods

The assessment component for the decision-support framework, which relies on an MCDA, was used to evaluate EIP services. As explained in **Chapter 4**, the application of the MCDA can be summarised in eight steps: after defining the decision-problem and the alternatives to be evaluated (step 1), their performance is assessed on the selected assessment criteria (step 2). To do so, each assessment criteria is operationalised using performance indicators based on routinely collected data (step 3). After estimating the performance of the alternatives, scores are obtained by standardising performance indicators (step 4). Afterwards, criteria weights, estimated in **Chapter 5** (step 5), are combined with the performance scores to obtain a MCDA total scores (step 6). Finally, sensitivity analyses are conducted (step 7) and results are reported and examined (step 8). This section describes the application of these steps for the study on mental health.

6.2.1. Alternatives to be evaluated

EIP services provide timely and comprehensive support to individuals during the initial stages of psychosis (376). These models of care usually operate stand-alone services, with EIP teams working independently from the generic community mental health teams (CMHTs). Patients receive a time-limited treatment package for about three years, and upon its culmination, they are either discharged to their primary care physician or transferred to a generic adult CMHT (270, 377).

The EIP teams comprise care coordinators, psychiatrists, psychological therapists (or clinical psychologists), education and employment specialists, peer support workers, social workers and clinicians experienced in working with children (270). These teams are required to follow NICE guidelines, such as the "Psychosis and schizophrenia in adults: prevention and management" (379). Quality standards have also been set by NHS England, including maximum waiting times from referral to treatment, psychological therapy for patients and family members, medication management, crisis care and others (270). The first access and waiting time standards for mental health, covering EIP services, were introduced by the Five Year Forward View for Mental Health in 2015 (88, 380)[49]. Since then, standards for EIP services have been reviewed and updated, holding a significant position within the NHS England's mental health agenda (378).

Although EIP services are guided by national standards and clinical guidelines, the budget allocations, team arrangements, care delivery processes, and interactions between EIP teams and other mental and

[49] Requiring that, by April 2016, more than 50% of people experiencing a first episode of psychosis should be treated with a NICE approved care package within two weeks of referral. In line with this commitment, in 2016 the NHS England released a guidance document to support commissioners and providers with helpful resources to support implementation (238). In 2019, The NHS Long Term Plan reconfirmed the NHS's commitment to the EIP access and waiting time standard that 60% of people with a first episode of psychosis should start treatment within two weeks of referral and receive a high quality NICE recommended care package (378).

non-mental health services may differ between services. As a consequence, performance across EIP services, even within the same ICS, can vary considerably, as documented in the last EIP Audit national report (381).

In BOB-ICS, different teams provide EIP services, such as the 'Adult Mental Health Organisation (AMHO) EIS Service' in Oxfordshire and the 'B Early Intervention Service' in Buckinghamshire. In this study, it is hypothesize that local commissioners are interested in expanding one of the two EIP services to a wider geographic footprint within the ICS. To assist local commissioners on this decision, the two EIP services are evaluated using the decision-support framework that this thesis proposes.

6.2.2. Assessing performance of EIP services

The performance of two EIP services in BOB-ICS was assessed through a retrospective longitudinal observational study using electronic health record data from OHFT. OHFT offers a comprehensive range of physical, mental health, and social care services to individuals of all ages in Oxfordshire, Buckinghamshire, Swindon, Wiltshire, Bath, and North East Somerset. These services are provided in various settings, including community bases, hospitals, clinics, and directly in patients' homes.

This section starts by describing the dataset used (**Section 6.2.2.1**), followed by an explanation of the performance indicators chosen to evaluate the performance of EIP services in Buckinghamshire (Bucks EIP) and in Oxfordshire (Oxf EIP) against each of the MCDA assessment criteria (**Section 6.2.2.2**). Afterwards, **Section 6.2.2.3** describes the doubly robust estimation method used to obtain predicted mean scores for each one of the performance indicators. **Section 6.2.2.4** describes the standardisation method used to obtain standardised performance scores. These scores are subsequently input into the MCDA decision-support framework developed to evaluate the overall performance of the two EIP services.

6.2.2.1. Dataset

Access to individual patient data was possible via the UK Clinical Interactive Search (CRIS). This system provides authorised researchers with regulated, secure access to anonymised information extracted from some NHS Foundation Trust electronic clinical records systems (382).

The dataset contained data of EIP patients diagnosed with SMI who were registered with OHFT and were active from the 1st of January 2016 until August 4th, 2022, which marks the date of the most recent record. The original dataset included information on 3,142 unique patients that were in contact with six different EIP services and 4,268 care episodes, defined as a referral to an OHFT service. The start and end dates of each episode of care is documented. Although the majority of patients (N=2,283; 72.67%) had only one episode of care recorded, 859 (27.34%) patients were found in the records with more than one episode of care. The dataset also contained information on mental hospital admission (hospital wards), contact with outpatient EIP services (e.g., contact with crisis teams)[50], SMI diagnosis variable with ICD-10 codes, and socio-demographic characteristics including date of birth, date of death, gender, ethnicity and area of residence (i.e., lower layer super output area (LSOA)[51]).

After consulting with clinicians from OHFT, the sample was refined to include patients who met the following criteria:

i. Those who started their first episode of care in or after January 2016, in line with the publication of NHS guidance on EIP services (238, 270).

[50] Patients may have contact with various outpatient EIP services, such as: community-based clinics, where they receive assessment, treatment, and support for psychosis-related issues; hospital-based services, provided within hospitals, particularly for patients who require more intensive care or crisis intervention, and; home-based care services, where healthcare professionals offer support and treatment in a familiar environment.

[51] "LSOAs typically comprise between 400 and 1,200 households or between 1,000 and 3,000 persons. They were created to improve the accuracy and relevance of statistics in smaller areas, enabling more precise local planning and resource allocation (349).

ii. Those who received treatment from either the 'AMHO EIS Service' in Oxfordshire (Oxf EIP) or the 'B Early Intervention Service' in Buckinghamshire (Bucks EIP). In the other four EIP services, patients were only assessed and referred to either 'AMHO EIS Service' or 'B Early Intervention Service'.

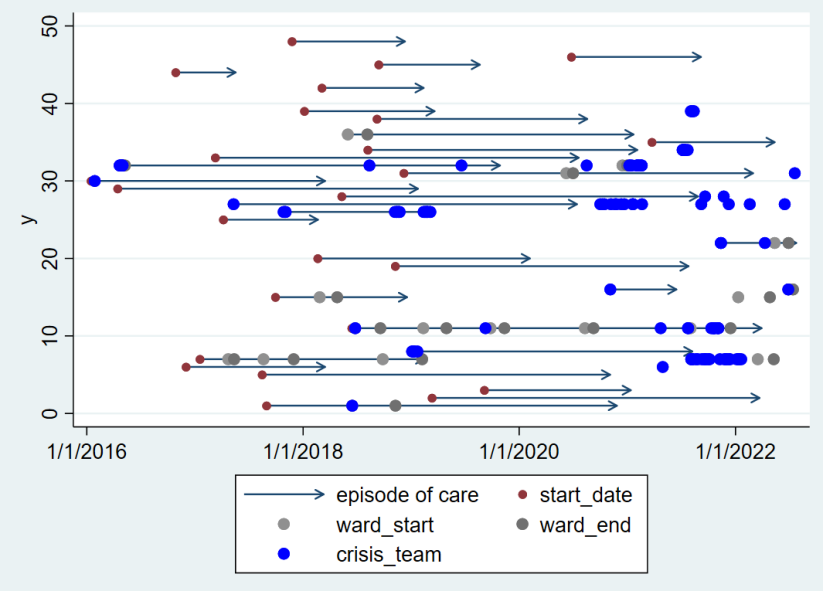
iii. Those whose episodes of care lasted more than 30 days. This is done as patients with an episode of care that last less than 30 days are likely to have only been assessed and not started treatment.

iv. Those with closed episodes, i.e. episodes that were closed either by discharge or death, in order to facilitate the construction of outcome variables.

The resulting sample comprised 1,127 patients with psychosis of which, 328 patients were treated in Buckinghamshire and 799 in Oxfordshire. In the total sample, 150 patients (13.31%) had multiple episodes of care, which were combined to create a single period of EIP episode of care[52] (see **Figure 6.A.2** from **Appendix 6.A**). Therefore, the service evaluation is based on a patient-level dataset, which includes for each patient the start and end date of the episode of care, the start and end date of each hospital ward, the dates on which the patients were in contact with EIP services, along with the socio-demographic and diagnostic variables mentioned above. **Figure 6.1** below illustrates the episodes of care for a subset of patients whose episodes last longer than 100 days and their initial 50 interactions with crisis teams.

[52] Once a patient is discharged from an EIP service, they would not normally be accepted back onto their caseload, as patients are typically discharged to their primary care physician or transferred to a generic adult community mental health team (377). Based on consultations with clinicians from OHFT, it is likely that records of more than one episode of care per patient are an artefact (e.g., patients were assessed twice, resulting in two different referral dates). Therefore, it was possible to combine these records into a single EIP episode of care.

Figure 6.1. Episodes of care for a subsample of patients with episodes longer than 100 days and first 50 contacts with crisis teams.



6.2.2.2. Operationalization of the assessment criteria

As outlined in **Chapter 4**, the operationalisation of the assessment criteria is contingent upon the nature of the medical condition or the interventions under assessment, in addition to the availability of data. To define the performance indicators for this service evaluation, a list of potential performance indicators was initially drafted, taking into account existing literature. This included studies that have previously assessed EIP services in England or similar settings (242, 377, 383-385), in addition to the National Clinical Audit of Psychosis (381) and documents from NHS England (238, 270, 386). Afterwards, the feasibility of each indicator was analysed in light of data availability. As part of this process, psychiatrists from OHFT were consulted between June and July 2023 to confirm the appropriateness of the performance indicators selected. **Appendix 6.B** summarises the rationale followed to choose each performance indicator. The sixth criterion of the decision-support framework developed, 'size of the target population', was not considered in this context as the two alternatives under evaluation (Bucks EIP and Oxf EIP) target the same condition (see **Chapter 4**).

The first criterion, 'final health outcomes', was measured as years of life during the observational period. Given that data spanned from 1 January 2016 to 4 August 2022, the maximum years of life

observed in the dataset was 6.6 years. Although the dataset only included information for six years, and patients could not be observed along all their lifespan, this variable was considered a good proxy of patient's longevity.

For the second criterion, 'quality of life and well-being', the average time-to-relapse (number of days) across all relapses for each patient was used as a proxy. In the context of psychosis, a patient was considered to have a relapse when psychotic symptoms return or get worse after a period of improvement or remission. Following scientific literature (377, 387) and clinical advice, a relapse was defined in the data when a patient was seen by a crisis team or was admitted to a mental hospital (i.e., the start of a ward stay). Hence, time to relapse refers to the time from the first remission of symptoms (i.e., date in which the care episode started) and the occurrence of a relapse (i.e., first contact with crisis teams or inpatient care). It is expected that EIP services have a positive effect on reducing the time to relapse. Reducing relapses aligns with the goals of EIP services, as outlined in the NHS England Guidelines (238, 270). The Health of the Nation Outcome Scales (HoNOS), a clinical tool used to evaluate the physical and social well-being of individuals with SMI (241), was considered as a potential indicator. However, there were several patients with no HoNOS records (%51 of the sample).

Disengagement from the service was used as a proxy for the third criterion, 'patient experience'. Following clinical advice and scientific literature, a patient was classified as having disengaged from EIP services if there has been no contact for three consecutive months following the episode's start date (377, 385). This outcome measure has been used in studies examining psychosis care pathways, and is therefore considered a good proxy of patient experience (ibid).

The fourth criterion, 'equity', focused on disparities in quality of life and well-being by calculating the average time-to-relapse ratio between the most deprived deciles (i.e., deciles 1, 2 and 3) and the richest

decile (i.e., decile 10)[53]. Health disparities are typically expressed as relative measures (i.e. ratios) obtained by comparing the groups with the lowest and highest levels of health (388, 389), and reducing unwarranted variation in health related health outcomes, including quality of life, is one of the ICSs' aims in England. To compute the ratio for each EIP service, an index of multiple deprivation (IMD) decile was assigned to each one of the patients in the dataset, in accordance with their LSOA (390). Differences among ethnic groups were also considered, but ethnicity data was not available for all patients in the dataset (23% patients had no ethnicity data).

Finally, for the 'cost' criterion, the average annual cost for each patient over the observational period was estimated. Three unit costs were considered: (i) the cost of each contact with outpatient mental health services (e.g., contact for assessment, contact with crisis teams, group contacts); (ii) the cost of a mental hospital admission (i.e., the start of a ward stay); and (iii) the cost per bed day. Unit costs were obtained from the NHS National Cost Collection for mental health services, summarised in the Unit Costs of Health and Social Care 2022 Manual (391). The cost of each outpatient contact with mental health services (unit cost: £241) took into account all contacts attended and cancelled by patients observed during the observation period. Total hospital costs for each patient were calculated based on the cost of hospital admission (unit cost: £294), the cost per bed day (unit cost: £341), and the length of stay in the hospital. To annualize the cost, the total cost for each patient was calculated and divided by the number of years of life during the observational period. **Table 6.1** summarises the assessment criteria and outcome measures used in the analyses.

Table 6.1. Operationalisation of the assessment criteria

Framework's assessment criteria	**Performance indicators**	**Calculation**
1. *Final or intermediate health outcomes*	Years of life during the observational period	For each patient, it was calculated as the difference between the date of death and the start date of the care episode. For patients that did not die during the observational period, the

[53] Decile 1 denotes the most deprived decile, while 10 represents the least deprived. Due to the low patient count in decile 1 and 2 (i.e. less than 2% of the sample), deciles 1, 2, or 3 - the three most underprivileged deciles - were used to calculate the equity indicators.

		latest date in the sample (cut-off date), 4th of August 2022, was used.
2. Quality of life and well-being considerations	Average time-to-relapse (days) across all relapses per patient	For each patient, the first relapse was calculated as the number of days between the start of the care episode and the date of the first contact with crisis teams or inpatient care (i.e. the start of a ward stay). For subsequent relapses, the number of days between the end of a contact with the crisis team or inpatient care and the start of a new contact with the same services was calculated. All recorded times to relapse during the observation period were then averaged for each patient.
3. *Patient experience*	Disengagement with the service	A patient was considered to have a disengaged from EIP services if there was no contact for a consecutive 3-month period after the start date of the episode. This variable equals 1 for patients that had no contact in 3 months after start of episode of care, and 0 otherwise.
4. Equity	Ratio of the average time-to-relapse between the most deprived deciles and the richest decile.	For each EIP service, Oxfordshire and Buckinghamshire, the median time to relapse was estimated by deprivation decile groups: that is, for patients in decile 1, 2 or 3 (deprived deciles) and for patients in the richest decile (10). The ratio between the two measures was then calculated.
5. Cost	Annual average cost across the observational period of EIP patients	Annual average cost across the observational period, including attended and not attended contacts (£, prices 2019). The following unit costs were used (391): -Any contact with EIP services: £241 -Hospital admission: £294 -Hospital bed per night: £341

6.2.2.3. Propensity Score Matching and Statistical Analyses

Propensity score matching (PSM) was used to reduce the observed cofounding and increase comparability between the two sub-samples, Bucks EIP and Oxf EIP services. PSM is a quasi-experimental technique introduced by Rosenbaum and Rubin in 1983 (392). It can be described as a semi-parametric approach that makes it possible to select a subset of each one of the EIS sub-samples, with similar covariate distributions, and with this, reduce the selection bias caused due to non-random assignment and endogenous participation decision (393).

The method involves pairing patients from Bucks EIP and Oxf EIP services with similar observable characteristics, using a function known as the propensity score. The propensity score is defined as the response probability of being treated (i.e. D=*1*, receiving EIP services in Buckinghamshire (Bucks EIP), the sub-sample of reference), conditional on a set of observable covariates (*X*):

$$P(X) = \Pr(D = 1|X) \qquad (1)$$

The method relies on two assumptions, (i) the strongly ignorable treatment assignment (SITA) and (ii) the Stable Unit Treatment Value Assumption (SUTVA) (244, 392, 393). The SITA assumption can be decomposed in two parts: (ia) the Conditional Independence Assumption (CIA), which implies that the treatment assignment (W) is independent of the potential outcomes, $Y_{oxf,}$ Y_{buck} (e.g. Additional years of life), given the covariates (X), and; (ib) the Common Support Assumption, according to which for all patients there is a positive probability of receiving treatment for all values of X[54]. In regards to the SUTVA, the treatment of one patient does not interfere with the effect of the treatment on another patient (i.e. no spill-over) and that there is no different version of treatment for different patients (i.e., consistency).

The effectiveness of the PSM strategy depends on the above assumptions. The CIA assumption would be violated if selection into treatment (i.e. D=1) is based on unobservable characteristics, which cannot be directly tested. In this study's context, while the possibility of potential unobserved confounders cannot be entirely ruled out, the likelihood of violating the CIA assumption is low. This belief is grounded in the fact that EIP services operate on a catchment area basis, and within the NHS England referral system, GPs refer patients based on specialist availability in these areas (238,270). The common support assumption is directly testable by examining the density distribution of the propensity score in both sub-samples, showing the existence of potential matches in the control group. In regards to the

[54] There is no unit in the support of the covariates such that D = 0 or 1: $0 < \Pr(D = 1|X) < 1, \ \forall X$. This assumption guarantees that all patients treated in Bucks EIP have a counterpart in Oxf EIP, and that anyone constitutes a possible "beneficiary" (394).

SUTVA, this assumption would be violated if the treatment status of a patient in Oxfordshire affects outcomes of a patient being treated in Buckinghamshire. This is unlikely due to service independence and the resources allocation mechanisms in place[55]. Finally, the consistency assumption is also likely to hold as all EIP services in the country are required to follow NICE guidelines, which includes the measuring and reporting performance against predefined standards (238, 270).

To perform the PSM, first, the propensity score is estimated for each patient by using a probit model, where treatment status is the dependent variable (i.e., D=1 for patients treated at EIP services in Buckinghamshire). The explanatory or observable covariates (X) include all patient characteristics of the dataset: age at start of the care episode, year of start of the care episode, gender, ethnicity, mental health diagnosis, and IMD decile[56]. Afterwards, the two sub-samples were matched using multiple matching algorithms, aiming at identifying the method that that produces the best balance between the treated and control groups in terms of observed covariates. Several matching methods were explored: (i) the *Mahalanobis metric matching*, under which the Mahalanobis distance is used to find control patients that are similar to treated patients, based on the controls, X, selected (395); (ii) the *nearest neighbour matching*, under which for each treated patient, the algorithm identifies one patient from the control group (i.e., Oxf EIP) that has the closest propensity score to the treated patient (393). This algorithm was used without replacement (i.e., after a control patient has been matched to a treated patient, it is removed from the pool of potential controls) and with replacement, and allowing for a

[55] Each EIP service operates independently, with its own clinical teams and management, ensuring treatment approaches are localized and patient-centered, tailored to individual needs. While one could argue that EIP services might compete for resources within the same ICS, influencing the quality of care across different areas, it is important to note that the study period spans from January 2016 to August 2022. During most of this period, healthcare budget allocation and oversight were primarily under the jurisdiction of Clinical Commissioning Groups, with ICSs officially assuming statutory responsibilities for budgets only from 1st July 2022 (see **Chapter 2**). In addition, it is important to note that ICSs have to comply with NHS England regulations, including the application of financial rules and adherence to the Mental Health Investment Standard (12). While ICSs have autonomy to set delegated budgets for place-based partnerships, they are required to make allocation decisions based on the principle of equal access for equal need. Regulatory bodies, such as the Care Quality Commission, and quality assurance processes also maintain consistent service standards across the regions.

[56] To satisfy CIA, it is important to include in the matching all variables known to be related to both treatment assignment and the outcome of interest (393).

maximum difference in propensity scores of 0.25 (i.e., caliper of 0.25) (396)[57]; (iii) *kernel matching*, which averages over multiple patients in the control group for each treated individual, with weights defined by the distance between the propensity scores of the treated and control groups (395, 397); and (iv) *the local linear matching estimator*, which is a variant of the kernel matching, under which the weighted control patients (from the kernel) are used for each of the treated patients, and then imbalances in the propensity scores among those closely matched units are adjusted by fitting a linear regression (398)[58]. In addition to these PSM techniques, stabilized inverse probability weighting (IPW) was performed, under which each patient in the treated sub-sample (i.e., Bucks EIP) and in the control sub-sample (i.e., Oxf EIP) are weighted by their inverse probability weights (399, 400)[59].

For examining the performance of the different PSM techniques implemented, three balance measures suggested by Rubin (2001) were used (401): the standardised difference of means (or standardised mean difference, SMDs), the Rubin's R, and the Rubin's B. The SMDs, also known as the "standardised bias" is the difference in means of each covariate, divided by the standard deviation in the full treated group. For a PSM to be trustworthy, the difference in the means of the propensity scores in the two groups under comparison should be small (e.g., the means must be less than half a standard deviation apart). The Rubin's R is the ratio of the variances of the propensity score in the treated and control groups, with values ranging between 0.5 and 2. Values for the Rubin's R close to 1 suggest that the sub-samples under comparison have similar variability in their propensity scores, which is desirable. Finally, the Rubin's B corresponds to the absolute difference in the means of the propensity scores for the treated and control groups, expressed as a percentage of the average standard deviation of the propensity scores

[57] A caliper of 0.25 means that, for a match to be valid, the absolute difference in propensity scores between a treated patients and a potential control match must be less than 0.25 times the standard deviation of the propensity scores.

[58] When implementing this technique, it is necessary to specify a bandwidth parameter to determine the 'local' region around a propensity score, within which the linear regression is fitted. In general, large bandwidths increase bias but reduce variance by weighting patients that are further away from the treated patient of interest (393).

[59] Inverse probability weights for patients from the treatment group ($W_{Bucks\,i}$) are computed as $W_{Bucks\,i} = \frac{S}{P(X)}$ and for the control group ($W_{Oxf\,i}$) with the following formulae, $W_{Oxf,j} = \frac{1-S}{1-P(X)}$, Where S represents the proportion of patients in the sample who are treated in Oxfordshire, and P(X) the propensity score.

in the two sub-samples. It provides a measure to gauge the overlap in the distribution of propensity scores between treated and control patients. Small values of the Rubin's B are therefore desirable (i.e., below 25%) as it suggest that distribution of the propensity score in both sub-samples are closely cantered around the same value.

To further reduce the confounding in the study and estimate the mean scores of the performance indicators, generalised linear models (GLMs) were used to model each performance indicator as a function of the treatment indicator (i.e., D=1,0, where 1 indicates being treated at Buckinghamshire EIP services) and the observed covariates, *X* (36, 399). These models are an extension of classical linear models that are suitable for non-normally distributed outcome variables. GLMs have three components (402): (i) the random component of Y, the outcome variable (i.e., performance indicators), which can follow any of the distributions from the exponential family (e.g., Binomial, Poisson, Gamma); (ii) a systematic component in which the linear combination of the covariates, *X*, produces a linear predictor, which is analogous to the deterministic part of the ordinary linear regression: $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$, where the β_0 is the intercept and $\beta_1, \beta_2, \ldots, \beta_p$ are the corresponding coefficients for each of the covariates or predictors, *x*; and (iii) a link function which specifies the relationship between the expected value of the outcome variable, Y, and the linear combination of predictors, η. The link function thus links the random component (i) with the systematic component (ii). The distribution family and link function in each regression were selected based on the nature of the outcome variable and the goodness of fit using the Akaike's information criterion (AIC) and the Bayesian information criterion (BIC), also known as the Schwarz Criterion[60] (403-405).

After testing different GLM specifications and selecting the appropriate distribution and link function, predicted mean scores were obtained for Oxf EIP and Bucks EIP services, on each one of the performance indicators. However, for the "ratio of the average time-to-relapse between the most

[60] The AIC is an out-of-sample prediction error estimator that rewards goodness of fit but penalises the more parameters included. Similarly, the BIC also rewards goodness of fit and penalises model complexity (i.e., adding more parameters), with the penalty increasing logarithmically with sample size (403).

deprived deciles and the richer decile", a GLM was not used as this indicators remained constant within each sub-sample. Consequently, the analysis focused on directly comparing the ratios between the two sub-samples. To address statistical uncertainty and variability in this comparison, bootstrapped standard errors were calculated for the ratios within the common support.

All estimations were carried out using Stata 17 (311). To summarise patients' characteristics, means, medians and standard deviations for continuous variables were used, whilst reporting frequencies for categorical variables[61].

6.2.2.4. Standardisation of performance scores

To calculate the overall MCDA score for the two EIP services under evaluation, the predicted mean scores were standardised using the relative standardization method (215), as indicated in **Chapter 4**. The following formulas were therefore applied:

$$S_{Bucks,k} = \frac{\mathrm{x}_{Bucks,k}}{\left(x^2_{Bucks,k} + x^2_{\mathrm{Oxf},k}\right)^{1/2}} \qquad S_{Oxf,i} = \frac{\mathrm{x}_{\mathrm{Oxf},k}}{\left(x^2_{Bucks,k} + x^2_{\mathrm{Oxf},k}\right)^{1/2}} \qquad (2)$$

Where $S_{Bucks,k}$ represents the standardisation of the performance score (x) of Bucks EIP services on criterion k relative to the combined performance score (x) of both alternatives, Bucks EIP and Oxf EIP services, on criteria k. As in the case of the "Disengagement with the service" and "Annual average cost" lower levels are preferable, the inverse of the unstandardized score was used, by replacing the scores in natural scales $\mathrm{x}_{Oxf,k}$ and $\mathrm{x}_{Bucks,k}$ are replaced by $\frac{1}{X_{Oxf,k}}$ and $\frac{1}{X_{Bucks,k}}$, respectively .

[61] The possibility that some contacts with EIP services or diagnoses were not reported cannot be entirely ruled out. However, there was no data to account for potential missing values in the 'contact' or 'diagnoses' variables. Clinicians confirmed that possibilities of missing 'discharge' or 'death' dates was minimal in these datasets. Furthermore, there were no missing values in age or gender, and only 3% of the sample lacked LSOA data (34 out of 1,127). These results are similar to those found in other studies using CRIS datasets (377). A category called 'not reported' was created for those without data on 'ethnicity'. This control variable was not statistically significant when estimating the p-score or in the GLM models.

6.2.3. MCDA total score

To obtain MCDA total scores for Bucks EIP and Oxf EIP services ($U(x)_{Bucks}$;$U(x)_{Oxf}$), the standardised performance scores ($S_{Bucks,k}$; $S_{Oxf,k}$) were combined with the relative weights (w_k) estimated in **Chapter 5.** These weights were derived from the discrete choice experiment, which elicited preferences from the public in England. As outlined in **Chapter 4,** MCDA total scores are determined using the Multi-Attribute Value Theory (MAVT) approach (217, 406), which follows the additive function presented below:

$$U(x)_{Bucks} = \sum_{k=1}^{5} w_k s_{Bucks,k}(x_{Bucks,k}) \quad ; \quad U(x)_{Oxf} = \sum_{k=1}^{5} w_k s_{Oxf,k}(x_{Oxf,k}) \qquad (3)$$

The MCDA overall scores were calculated both inclusive and exclusive of cost, in response to criticisms of combining cost with other benefit-related criteria. As explained in **Chapter 4**, those arguing against cost inclusion in the MCDA state that it fails to adequately address the opportunity costs of an intervention (210, 266). In addition, the MCDA overall score without cost was used to calculate the cost-per-value ratio, which gives the cost of the EIP service per unit of its aggregate value. If the cost-per-value ratio of the EIP service under evaluation is lower than its comparator, it suggests that the service offers greater value per unit cost, which also can be helpful in deciding which EIP service should be prioritized (162, 407).

6.2.4. Sensitivity analysis

To investigate how sensitive MCDA total scores are to the uncertainty surrounding the criteria weights and performance scores, as outlined in Chapter 4, a probabilistic sensitivity analysis (PSA) was conducted using Monte-Carlo simulations (188, 215). To this purpose, 1,000 replications of both the unstandardised mean scores and the importance weights. Normal distributions were assumed for all criteria weights. For the performance scores, ‘years of life during observational period’ and ‘disengagement with the service’ a normal distribution was assumed. For ‘average time-to-relapse’ and

'annual average cost' performance scores, a gamma distribution was assumed, and for the 'ratio between the poorest and the richest decile on time to relapse' a beta distribution. From these simulations, the standard deviation and the 95% confidence intervals (CIs) around the overall MCDA scores were obtained for each EIP service. To summarize the sensitivity of the results to uncertainty, the probability of favouring Oxf EIP services over Bucks EIP was computed, by calculating the proportion of MCDA iterations in which the Oxf EIP had a higher overall value score than Bucks EIP.

In addition to providing a graphical summary of the resulting MCDA scores distributions, parameter uncertainty in the MCDA is also graphically presented with the conditional multi-attribute acceptability curve (CMAC) (36, 149). As described in Chapter 4, the CMAC's vertical axis displays the likelihood of a new care model (e.g., EIP Oxf) being accepted as the preferred alternative to the comparator, whilst the horizontal axis shows varying thresholds of maximum budget for allocating to either EIP service to treat a given population size. The CMAC shows, across a variety of budget levels, the probability of the EIP service being the chosen option (i.e., having the highest MCDA score) while the budget impact remains below a budget-threshold.

6.3. Results

6.3.1. Selection of Propensity Score Matching technique

Table 6.2 summarises the performance of the different PSM techniques explored. Kernel-based matching was the best matching technique, as it resulted a high number of respondents with the least covariate imbalance (i.e., SMD > 0.25) and with matching statistics within the desired range (Rubin's B = 13% and Rubin's R=1.23). **Appendix 6.D** includes the propensity score histogram, showing a good overlap and balance of the propensity scores between the Bucks EIP and Oxf EIP patients. The same appendix includes graphs of the standardised mean difference before and after PSM, for each one of the matching techniques explored. After PSM, there were no statistically significant (P-value < 0.05) differences in baseline characteristics between the two sub-samples.

Table 6.2. Summary results of performance of PSM and Stabilized Inverse Probability Weighting

Techniques using propensity score	Covariates with SMD>0.25	Mean bias	Median bias	Maximum SMD (%bias/100)	Rubin's B (%)	Rubin's R
Unmatched	4	10.1	5,1	0.377	72.8*	0.65
Nearest Neighbour without replacement (a)	0	3.2	2.1	0.078	24.5	1.07
Nearest Neighbour with replacement (a)	0	3.1	2.2	0.097	19.4	1.05
Mahalanobis	1	6	3	0.262	48.7*	1.29
Kernel	**0**	**2.1**	**1.4**	**0.059**	**13.0**	**1.23**
Stabilized Inverse Probability Weighting (b)	1	7.1	3.9	0.261	50.8*	0.49*

SMD, Standardised mean difference
(a) with caliper:0.25. (b) Stabilized Inverse Probability Weighting is not a PSM technique, but a weighting technique that uses the inverse of the probability of receiving the treatment to weight each subject in the sample. Convergence was not achieved when using the local linear matching estimator.
* if Rubin's B > 25%, Rubin's R outside [0.5, 2].

Characteristics of EIP patients, before and after the PSM, are presented in **Table 6.3**. Before the PSM, age, gender and start year distributions of patients receiving care episodes from both EIP services were similar. In both samples, the mean age of patients at the start of their episode of care was approximately 29 years old and around 42% were female. The majority of patients initiated their episode of care prior to 2020, with a similar distribution of the sample across the calendar years of observation. There were however statistically significant differences in some of the mental health diagnoses, as patients who received treatment at Bucks EIP reported higher prevalence of schizophrenia, affective disorders, and substance use disorders. In addition, there were proportionally more patients with Asian ethnicity in Bucks EIP and proportionally more patients with White ethnicity in Oxf EIP. Moreover, Oxford EIP sub-sample had a higher percentage of individuals in the second decile, whereas in the Bucks EIP, the proportion of individuals in the fourth decile was twice that of the Oxford EIP. **Appendix 6.C** provides summary statistics of the outcome variables before PSM.

In total, 318 patients from Bucks EIP were matched with 740 patients from Oxf EIP. After the matching, there were no statistically significant differences at 95% confidence level between the two sub-samples.

All SMDs are below 0.10, suggesting that the two sub-samples were well balanced in terms of the observed characteristics.

Table 6.3. Characteristics before and after PSM

	Before PSM				After PSM			
	Bucks EIP (treatment) N=328	**Oxf EIP (control)** N=799	**p value**	**|SMD|**	**Bucks EIP (treatment)** N=318	**Oxf EIP (control)** N=740	**p value**	**|SMD|**
Age, mean (s.d.)	29.98 (11.52)	29.45 (12.05)	0.50	0.05	30.2 (11.6)	29.62 (12.08)	0.81	0.02
Female, % (n)	42.07% (138)	41.43% (331)	0.84	0.01	42.14% (134)	41.89% (310)	0.98	0.00
Year, % (n)								
2016	19.51% (64)	16.65% (133)			19.5% (62)	17.03% (126)		
2017	23.48% (77)	25.03% (200)	0.58	0.04	23.58% (75)	24.32% (180)	0.71	0.03
2018	23.48% (77)	20.9% (167)	0.34	0.06	23.27% (74)	15.54% (115)	0.84	0.02
2019	16.77% (55)	17.02% (136)	0.92	0.01	17.3% (55)	17.03% (126)	0.86	0.01
2020	10.06% (33)	12.27% (98)	0.30	0.07	9.75% (31)	12.57% (93)	0.96	0.00
2021	5.79% (19)	6.63% (53)	0.60	0.04	5.97% (19)	6.62% (49)	0.66	0.04
2022	0.91% (3)	1.5% (12)	0.44	0.05	0.63% (2)	1.49% (11)	0.86	0.01
Mental health diagnosis, % (n)								
Schizophrenia	47.56%(156)	37.17%(297)	0.00**	0.21	47.8%(152)	37.7%(279)	0.91	0.01
Affective D.	24.09%(79)	14.27%(114)	0.00**	0.25*	24.21%(77)	14.73%(109)	0.94	0.01
Organic D.	1.22%(4)	0.75%(6)	0.45	0.05	1.26%(4)	0.81%(6)	0.60	0.05
D. substance use	17.38%(57)	10.01%(80)	0.00**	0.22	16.98%(54)	9.73%(72)	0.53	0.05
Other D.	16.16%(53)	15.77%(126)	0.87	0.01	16.04%(51)	15.14%(112)	0.74	0.03
Ethnicity, % (n)								
White	49.09%(161)	61.7%(493)	-	-	49.06%(156)	61.22%(453)	-	-
Black	5.18%(17)	4.26%(34)	0.50	0.04	5.35%(17)	4.46%(33)	0.77	0.02
Mixed	4.57%(15)	3.75%(30)	0.52	0.04	4.09%(13)	3.78%(28)	0.74	0.03
Asian	15.85%(52)	5.63%(45)	0.00**	0.33*	16.04%(51)	5.81%(43)	0.96	0.01
Other	1.22%(4)	2.13%(17)	0.31	0.00	1.26%(4)	2.16%(16)	0.92	0.01
Multi deprivation index decile, % (n)								
1 (most deprived)	0.91%(3)	1%(8)		-	0.94%(3)	1.08%(8)	-	-
2	0.61%(2)	8.01%(64)	0.00**	0.38*	0.63%(2)	4.19%(31)	0.47	0.02
3	5.79%(19)	4.76%(38)	0.480	0.05	5.97%(19)	5.14%(38)	0.91	0.01
4	14.63%(48)	7.76%(62)	0.00**	0.22	15.09%(48)	8.38%(62)	0.55	0.05
5	8.23%(27)	6.51%(52)	0.311	0.07	8.49%(27)	7.03%(52)	0.50	0.06
6	9.76%(32)	11.14%(89)	0.483	0.05	10.06%(32)	12.03%(89)	0.89	0.01
7	13.41%(44)	13.27%(106)	0.966	0.00	13.84%(44)	14.32%(106)	0.88	0.01
8	13.11%(43)	11.14%(89)	0.361	0.06	13.52%(43)	12.03%(89)	0.87	0.01
9	13.41%(44)	16.52%(132)	0.183	0.09	13.84%(44)	17.84%(132)	0.92	0.01
10 (least deprived)	17.38%(57)	16.77%(134)	0.826	0.02	17.61%(56)	17.97%(133)	0.93	0.01

Age refers to age at the start of the episode of care; Year to the year the episode of care started; D indicates mental health disorders; PSM, propensity score matching; s.d., standard deviation; SMD, standardised mean differences. *P <0.01; **P <0.001; *SMD > 0.25.

6.3.2. Performance scores and MCDA total scores

Appendix 6.E provides an overview of the various GLM specifications tested, indicating the selected model for each performance indicator and summarising the regression outputs. **Table 6.4** presents the predicted mean scores for patients from Bucks EIP and Oxf EIP across the five criteria. Both EIP services displayed nearly identical longevity, with an average of approximately 4.07 years of life observed during the period and a mean difference close to zero and not statistically significant (P-value > 0.05). Bucks EIP outperformed Oxford EIP in the average time to relapse (376.84 days vs. 313.40 days), but the observed difference of 63.50 days did not reach statistical significance (P-value > 0.05). A 63.5 days difference, however, can constitute a clinically meaningful difference and should be evaluated in consultation with clinicians. In regards to disengagement rates, in both EIP services about 10% of the patients had no contact for a consecutive 3-month period after the start date of the episode of care. The mean difference between both services was close to zero and not statistically significant (P-value > 0.05). In terms of equity, measured as the ratio of average time to relapse between the poorest and the richest IMD decile, the ratio for Bucks stood at 0.50, lower than Oxford's 0.87, resulting in a mean difference of -0.37. Although the ratio is closer to one in Oxf EIP services than in Bucks EIP, suggesting less inequality in the time-to-relapse among socioeconomic groups in Oxford, the mean difference between both EIP services is not statistically significant at 95% confidence level. The annual average cost of Bucks EIP services (£6,867.48) was substantially higher than that of Oxford EIP services (£5,745.40), with a statistically significant mean difference of £1,122.08 (P-value <0.05).

Table 6.4. Unstandardized performance scores after propensity score matching

Performance indicator	Bucks EIP		Oxf EIP		Mean difference	p-value(f)
	Predicted mean scores	s.e.	Predicted mean scores	s.e.		
Years of life during observational period (a)	4.071	0.033	4.075	0.030	-0.004	0.931
Average days-to-relapse across all relapses per patient (b)	376.843	31.036	313.358	24.350	63.485	0.125
Disengagement with the service (c)	0.098	0.015	0.100	0.010	-0.0027	0.884
Ratio of the average time-to-relapse between the most deprived deciles and the richer decile (d)	0.504	0.183	0.874	0.209	-0.3689	0.184
Annual average cost £ (e)	6,867.481	460.053	5,745.397	396.087	1,122.084	0.022*

s.e. refers to standard error; 95% CI, 95% confidence interval. (a) GLM with Gaussian distribution and log link function. (b) GLM with Gamma distribution and log link function. (c) GLM with binomial distribution and logit link function. (d) No GLM model was run as this performance indicator is a constant within each sub-sample. Hence, the ratio for each sub-sample and its corresponding bootstrapped standard errors were calculated for patients in the common support. (e) GLM with Gamma distribution and log link function. (f) p-values come from the GLM regression, except for "Ratio of the average time-to-relapse between the most deprived deciles and the richer decile" for which bootstrap standard errors were calculated. *P <0.05; ** P <0.05.

Table 6.5 displays standardised performance scores, criteria weights and the overall MCDA scores for Bucks EIP and Oxf EIP. Results are also summarised graphically in **Figure 6.2**. In Bucks EIP services, the highest performance score was achieved on the quality of life and well-being criterion (0.769) whereas the equity criterion scored the lowest (0.500). In contrast, in Oxf EIP services, the highest score was achieved on the equity criterion (0.866) and the lowest in the 'quality of life and well-being' criterion (0.639). Differences between the two EIP services were negligible in the final health outcome and quality of care criteria.

In terms of the overall scores, calculated by combining the weights and standardised performance scores, the main contribution comes from performance on final health outcomes (> 30% in all scenarios)

and patient experience, measured as the probability of disengage with the service (> 27% in all scenarios) (see **Figure 6.2**). When cost was included in the MCDA score, its contribution is of 22.5% in Oxf EIP services and 19.2% in Bucks EIP. The largest differences between the two EIP services were found in the scores on annual average cost (0.106 vs 0.127), quality of life and well-being (0.093 vs 0.077) and the equity criterion (0.014 vs 0.024). When cost was not included in the MCDA, results were mainly driven by differences in quality of life and well-being (0.135 vs 0.113), and equity (0.048 vs 0.083).

Total MCDA scores can range from 0 to 1, with a higher score indicating better performance. In both scenarios, the MCDA total score indicates that Oxf EIP outperforms Oxf Bucks (0.552 vs 0.563 and 0.543 vs 0.550 when cost is not included). The cost-per-value ratios, obtained by dividing the MCDA total score (without cost) by the annual average cost, suggests that Oxf EIP (£ 10,437.9) provides more value per unit cost than Bucks EIP (£ 12,654.6).

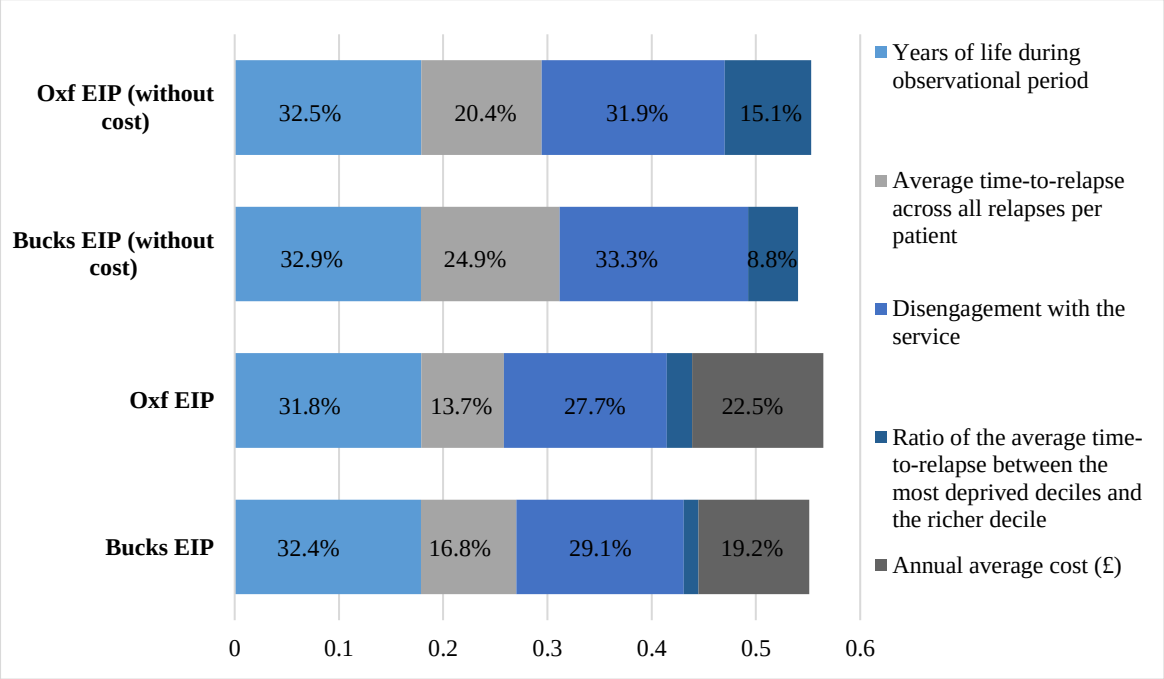
Table 6.5. Standardised performance scores, weights and MCDA scores

Assessment criteria	Performance indicators (a)	Standardised performance score (s.e.)		Relative weights (s.e.) (b)		MCDA scores (with cost)		MCDA scores (without cost)	
		Bucks EIP	Oxf EIP	with cost	With out cost	Bucks EIP	Oxf EIP	Bucks EIP	Oxf EIP
Final or intermediate health outcomes	Years of life during observational period	0.707 (0.004)	0.707 (0.004)	0.253 (0.019)	0.253 (0.016)	0.179	0.179	0.179	0.179
Quality of life and well-being considerations	Average time-to-relapse across all relapses per patient	0.769 (0.038)	0.639 (0.045)	0.120 (0.013)	0.176 (0.014)	0.093	0.077	0.135	0.113
Patient experience	Disengagement with the service	0.717 (0.065)	0.697 (0.067)	0.224 (0.020)	0.252 (0.019)	0.161	0.156	0.181	0.176
Equity	Ratio of the average time-to-relapse between the most deprived deciles and the richest decile	0.500 (0.142)	0.866 (0.092)	0.028 (0.021)	0.096 (0.017)	0.014	0.024	0.048	0.083

Cost*	Annual average cost	0.642 (0.029)	0.767 (0.025)	0.165 (0.013)	-	0.106	0.127		
Overall total score						**0.552**	**0.563**	**0.543**	**0.550**
Cost-per-value ratio (£)								**12,654.6**	**10,437.9**

(a) Standardised performance scores based on relative standardization with patient experience and total costs being reverse coded with a lower standardised score referring to a higher non-standardised score. (b) Relative weight of 'Size of the target population" criterion (20.8% (0.015) with cost, and 22.40% (0.02) without cost) was not included in this case study. When calculating the overall score, the relative weight of 'Size of the target population" multiplied by zero.

Figure 6.2. Standardised performance scores, with and without cost



6.3.3. Sensitivity analysis

Results of the probabilistic sensitivity analysis (PSA) are presented in Table 6. Figure 3 summarises the distribution of MCDA total scores for both EIP services under evaluation. In 60% of iterations (600 out of 1,000), Oxf EIP (mean: 0.5629) had a higher total MCDA score than Bucks EIP (mean: 0.5491). Results are similar when cost is not included in the MCDA, with Oxf EIP outperforming Bucks EIP

services in 54% of the iterations. When cost is not included, standard deviations increase slightly, which may reflect greater variability in performance measures unrelated to cost.

Even though Oxf EIP services seem to perform better than Bucks EIP, the MCDA score distribution of both EIP services are very similar (see **Figure 6.3**). There is considerable overlap in the interquartile ranges, suggesting that there is no clear difference between the two alternatives under evaluation. However, differences between the two EIP services are more apparent when comparing the cost-per-value ratios. The cost-per-value ratio, which relates the cost of EIP services to the value they deliver, is £12,877.29 for Bucks EIP and £10,550.33 for Oxf EIP, indicating that Oxf EIP services provide more value per pound spent. The standard deviation and confidence intervals suggest that this finding is consistent across simulations, with the ratio for Oxf EIP services being more favourable in 82% of the iterations.

The CMAC (**Figure 6.4**) shows that the probability of Oxf EIP being effective and affordable, from the perspective of general public in England, given a certain target population and available budget. The probability that Oxf EIP is effective and affordable, relative to Bucks EIP, increases to a maximum of 60% at a budget of £ 39,700,000 for the annual average cost of 3,142 people eligible to be referred to Oxf EIP services (i.e., £ 12,635.26 per patient)[62].

Table 6.6. Results of the probabilistic sensitivity analysis (PSM)

	MCDA with cost		MCDA without cost		Cost-per-value ratio	
Statistics	**Bucks EIP**	**Oxf EIP**	**Bucks EIP**	**Oxf EIP**	**Bucks EIP**	**Oxf EIP**
Mean	0.5491	0.5629	0.5371	0.5452	£12,877.29	£10,550.33
Std.Dev	0.0297	0.0320	0.0353	0.0361	£1,775.55	£1,774.19
95% IC (low)	0.5473	0.5609	0.5349	0.5430	£12,767.25	£10,440.36
95% IC (up)	0.5510	0.5649	0.5393	0.5475	£12,987.34	£10,660.29
Times Oxf > Bucks	600		544		178	

[62] As indicated in Figure 6.A.2 of Appendix 6.A, 3,142 EIP patients were active in 2016. BOB ICS 2022/23 allocation was £2.9 billion. According to the National funding profile and data from the NHS Mental Health Implementation Plan, the NHS spends around 9% if its total budget on mental health services. In BOB ICS, this is equivalent to £261,000,0000. From this value, allocation to EIP services correspond to 15.24%, which is equivalent to £39,776,400 for BOB-ICS (115, 378).
£ 12,635.26 (in 2022) corresponds to £11,193.86 in prices 2019.

% Oxf > Bucks	60%	54%	18%

95% IC, confidence interval at 95%.

Figure 6.3. PSM results of MCDA overall scores

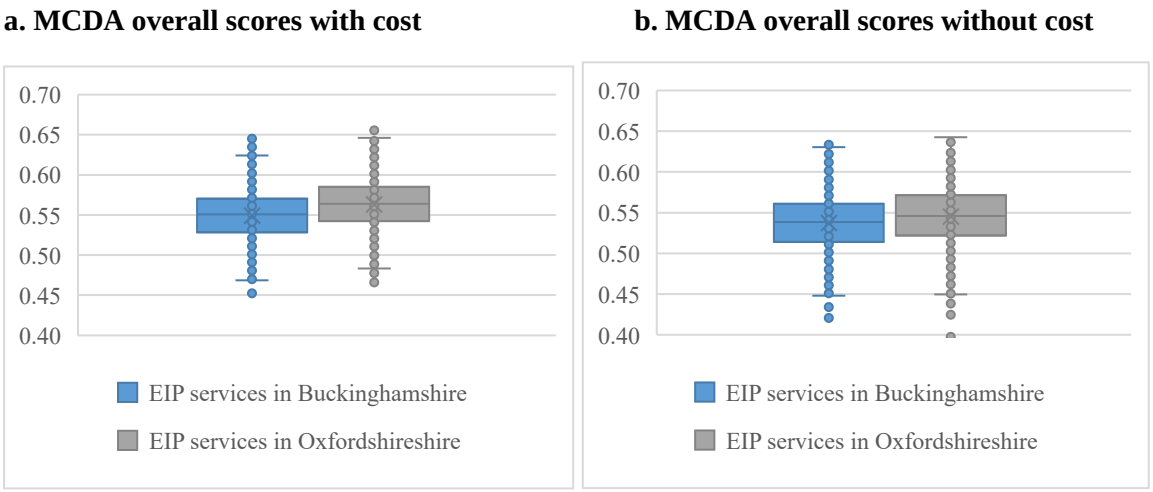
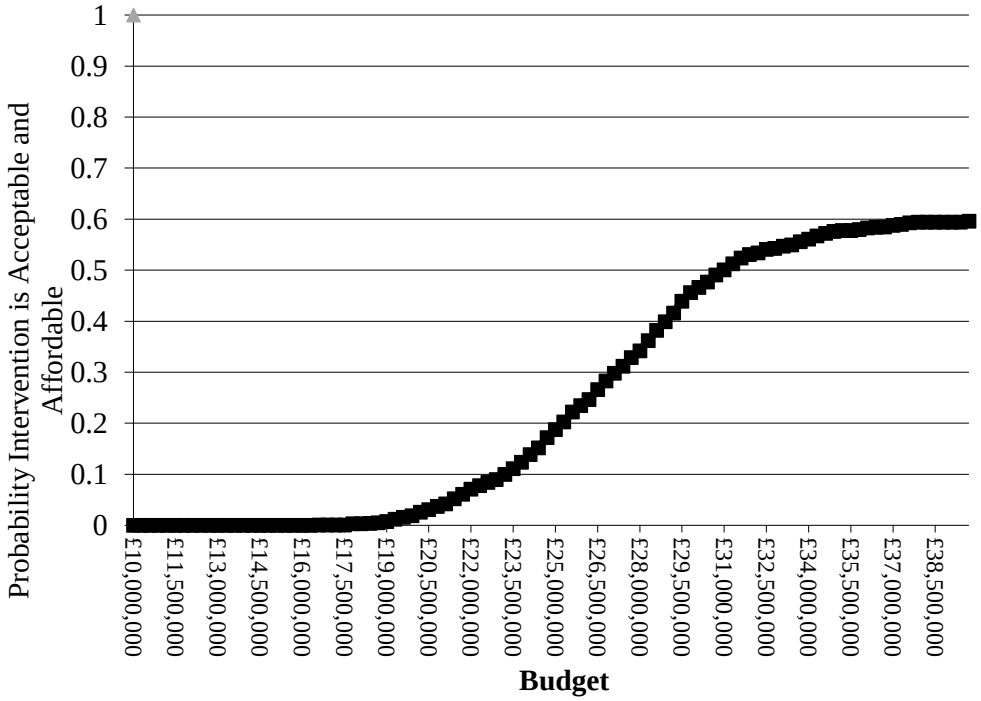


Figure 6.4. Conditional multi attribute acceptability curve (CMAC) showing the probability of Oxf EIP to be accepted as the preferred EIP service over a range of available budgets.



6.4. Discussion & Conclusion

Findings of this case study suggest that, during the period from 2016 to 2022, the EIP services provided by Oxfordshire outperformed those provided by Buckinghamshire. This conclusion remains valid when cost is not included in the MCDA overall score and when using the cost-per-value ratio, suggesting that Oxf EIP services are more cost-effective than Bucks EIP. Results are mainly driven by the significantly lower annual per capita cost of EIP services in Oxfordshire and the lesser socioeconomic inequality in quality of life outcomes, measured in terms of average time to relapse. The sensitivity analyses, however, indicate that while Oxf EIP services may have a slight edge in the MCDA total scores, difference between the two services is not pronounced. In light of these results, BOB-ICS commissioners may be inclined to prefer arrangements and procedures from Oxf EIP in a scenario where they have to determine which EIP service to scale-up across BOB-ICS. Nevertheless, it could be relevant for local commissioners to explore the drivers of the differences in average time to relapse between the two services, with Bucks EIP outperforming Oxf EIP. Some strategies implemented by Bucks EIP teams to minimize time to relapse (e.g., ensuring medication adherence, providing robust psychosocial support, monitoring early warning signs) might be more effective than the ones used by EIP teams in Oxfordshire, and could be considered when scaling up the intervention to other areas in England.

This case study highlights the importance of presenting MCDA results both with and without the inclusion of the cost criterion. Differences in MCDA total scores (with cost) between the two EIP services were relatively small, despite the statistically significant difference in costs. In both EIP services, the contribution of the cost criterion to the total score was below 25%. This contrast with the observed relatively large differences in cost-per-value ratios, confirming that Oxfordshire EIP services are more cost-effective than EIP services in Buckinghamshire. Providing decision-makers with MCDA scores, with and without cost, as well as with cost-per-value ratios, enriches the decision-making process. With cost in the MCDA, the relative contribution of costs is made explicit. Given the emphasis

on cost reduction in the context of new models of care and the importance that general population attain to cost of healthcare interventions—as revealed in the DCE conducted in **Chapter 5**—cost becomes a crucial criterion in the evaluation. By excluding cost from MCDA, it is possible to estimate the expenditure required to achieve one unit of 'value', which can help to validate the efficiency ("cost-effectiveness") of an intervention, relative to other. This can be particularly useful when differences in MCDA scores are small, such as in the case study presented in this chapter.

This case study has demonstrated the feasibility of implementing the decision-based framework that this thesis develops to assess new models of care in England. Based on the analysis conducted, the chapter has demonstrated that local commissioners can potentially leverage routinely collected data, such as electronic health records of patients with psychosis, to systematically assess new modes of care and make informed investment decisions grounded in evidence. This adds to existing literature in the MCDA field that has demonstrated the usefulness of MCDA to assess healthcare interventions, particularly in the context of complex interventions (36, 37, 149, 408). In addition to offer a total score, MCDA decomposes the multiple outcome of interest of decision-makers to support multifactorial decisions, making explicit the relative importance and the contribution of each one of them. This allows a more comprehensive understanding of the performance or value for money of an intervention, allowing associations with key performance indicators (KPIs) frequently used at the local level (e.g., associations between disengagement with the service and medication adherence KPIs).

This case study has also highlighted gaps in current datasets, which align with concerns expressed by local decision-makers from South East England, interviewed in 2021 (see **Chapter 2**). When deciding performance indicators, HoNOS and DIALOG scores were potential options as these tools enable

clinicians to measures health and social functioning of people with severe mental illness (SMI)[63]. According to NHS England guidelines, patients should be assessed with HoNOS or other tool as a minimum during initial assessment, at six and 12 months, annually, and upon discharge (270). Nevertheless, the electronic health records used for this study did not included HoNOS or DIALOG scores for more than half of the patients. This underscores the clear need to enhance data collection for this crucial measure. The need to improve data collection was further highlighted by its incomplete records on ethnicity. This gap is particularly concerning given the responsibility that ICSs have for reducing unwarranted clinical variation and addressing health inequalities. Furthermore, there remains a significant scope to enhance the gathering of data on patient experience. In this case study, the 'patient experience' criterion was framed in terms of disengagement with the service, which is pertinent when treating patients with psychosis. As local commissioners may need to compare models of care addressing different conditions, it is crucial to reach a consensus on non-condition-specific Patient-Reported Experience Measures (PREMs) and strength efforts to collect data in this regard.

Strengths and limitations

Several methodological strengths are worth noting. First, the study identified performance indicators to operationalise each of the assessment criterion, overcoming data limitations in an area - mental health - where quantifiable and objective data have traditionally been more scarce and less analysed compared to physical health (409, 410). The results of the PSM are indicative of its successful implementation in the study, effectively balancing the covariates between the two sub-samples, the Bucks EIP and Oxf EIP services. This increased the robustness of the service evaluation undertaken and underlined the reliability of the findings. In addition, the presentation of the MCDA uncertainty results using the CMAC showed how the preference for EIP services in BOB-ICS changes across a range of available

[63] HoNOS is a clinician-rated outcome measure that covers safety, substance use, physical health, symptoms and social issues. DIALOG is a service user-rated outcome and experience measure, which focuses on quality of life, care needs and treatment satisfaction (270).

budgets, considering all outcomes of interest together (i.e., the assessment criteria) and the preferences of the general public. As the CMAC uses the available budget to be allocated to different EIP services for a given population, the results could inform local commissioners when making decisions about mental health budget allocation (37, 149). While the integration of budget constraints in MCDA is an ongoing area of debate (48, 411), the use of CMAC in this study offers a practical approach to navigate these constraints, providing a valuable tool for informed local decision-making in mental health resource allocation.

This study is highly relevant to the current healthcare decision-making landscape in England. Following the statutory establishment of ICSs across England in July 2022, there has been a marked emphasis on data-driven population health management, strongly advocated by NHS England (412). Integrating comprehensive data with in-depth population health analytics is pivotal. It helps local commissioners understand the needs of their communities, enabling them to create, adapt, and implement a range of evidence-based interventions effectively. This study, in particular, harnesses a rationing framework aligned with the preferences of England's general population, making use of routinely collected data available to local commissioners. This approach is especially crucial in the mental health sector, considering the significant burden of mental health conditions in the country and the increasing demand for integrated approaches to address the needs of patients with severe mental illness (378). The current lack of comprehensive rationing frameworks for assessing mental health programmes underscores the importance of this research.

Nevertheless, it is crucial to acknowledge the limitations of the service evaluation developed in this chapter. First, when conducting the study it was not possible to engage in discussions with the commissioners of BOB-ICS, which hindered access to more comprehensive details about the delivery of EIP services within BOB-ICS. Additionally, information that would allow for a better understanding of the differences between the two services was not publicly available. Consequently, it was not possible

to fully comprehend differences in the care services delivered in Oxfordshire and Buckinghamshire. Additionally, findings were not communicated and discussed with local commissioners, resulting in an inability to acquire feedback and validation concerning the practicality and pertinence of the framework proposed in this thesis. While the chapter demonstrated the feasibility of implementing the decision-based framework to assess new models of care, its applicability and pertinence are yet to be confirmed.

It could be argued that the performance indicators used to evaluate the two EIP services may not accurately reflect the criteria outlined in the MCDA. In addition, questions may be raised about the consistency between the broad definition of the evaluation criteria and the definitions used in the choice experiments to elicit the criteria weights (**Chapter 5**). For example, while few would question the use of 'years of life gained during the observation period' to operationalise the 'final health outcome' criterion, the suitability of 'time to relapse' as a proxy for 'quality of life improvements' may be more questioned, as it may not fully capture the multifaceted nature of quality of life improvements experienced by people with psychosis. Similarly, 'disengagement with the service', although potentially indicative of an unsatisfactory patient experience, could be due to various other factors not captured in the dataset, such as personal circumstances or symptom severity. However, it could also be argued that a good service would adjust medication or accommodate patient preferences to prevent disengagement. In addition, there is a discrepancy between the framing of the cost criterion in the DCE (presented as additional income tax per year) and its operationalisation in the service evaluation, which used annualised average costs over the observation period.

As discussed in **Chapter 4**, the aim of this thesis was to develop a framework based on routinely collected data and social values that could be applied to different new models of care. As such, the evaluation criteria were broadly defined and, in defining the attributes of the DCE (**Chapter 5**), consideration was given to the ease with which the attributes of the experience could be understood by the public and to the fact that in England the public indirectly bears the cost of healthcare through taxes.

These two approaches adopted in developing the decision-support framework come at the cost of accepting a degree of discrepancy between the definition of the evaluation criteria or DCE attributes and the data available for operationalising the criteria and estimating the performance scores.

Since the start of this thesis in 2021, considerable effort was invested in obtaining comprehensive datasets to assess two or more new models of care. In principle, the intention was to access data to assess diabetes integrated care services in Hampshire and Isle of Wight Sustainability and Transformation Partnership (STP). Despite numerous consultations with clinicians and completion of the necessary ethical approval process, access to the data was not granted by local commissioners and the commissioning support unit (CSU) that was supporting the STP at that time. This was attributed to the ongoing transformation of the local commissioning system in England, with STP commissioning teams transitioning to new ICB/ICS legal structures. Databases were also undergoing a transition from CSU servers to ICS population-based servers, which further complicated access. The COVID-19 pandemic increased demands and workload on both clinicians and local commissioners, resulting in limited opportunities for meetings with them for the development of this thesis.

Conclusion

This case study in mental health has demonstrated that it is possible to use the MCDA-based framework developed in this thesis to generate evidence on the performance of new models of care in England. Although there remains a significant scope to enhance data quality and accessibility, this framework has the potential to support efforts in the country to promote local population health management, informed by data analysis.

Chapter 7. Discussion and conclusion

7.1. Introduction

This thesis aimed to develop a systematic and accessible framework that incorporates evidence on multiple aspects of decision-making to support the local commissioning of new models of care in England. The establishment of Integrated Care Systems (ICSs) as pivotal healthcare stakeholder has encouraged the spread of new models of care across the country. However, the evaluation of these models poses multiple challenges due to their complex nature that is not accounted for in traditional economic evaluation methods. In a context of rising expectations for integrated care and tight healthcare budgets, it is imperative for local commissioners to improve efficiency by investing in value and evidence-based healthcare. For that, it is essential that ICSs strengthen their evaluation culture and capacity, ensuring evaluations are not only methodically sound but also embedded within the local commissioning process, thereby promoting evidence-based decision-making.

In this final chapter, **Section 7.2** presents a summary of the key findings of the thesis. This is followed by a discussion of the limitations of the research in terms of the methods used and the potential applicability of the decision-support framework developed (**Section 7.3**). Additionally, the chapter considers the implications and their potential consequences for local healthcare policy-making in England (**Section 7.4**). Finally, the chapter concludes with some steps forward and concluding remarks (**Sections 7.5 and 7.6**).

7.2. Summary of key findings

Chapter 2 explored the policy initiatives that have shaped the integration agenda in England over the past three decades. These initiatives have gradually led to the current commissioning structures, with

ICSs as the centrepiece of today's local commissioning landscape. In light of the statutory duties assigned to ICSs, including greater flexibility for setting priorities, this novel structure provides local decision-makers with the opportunity to improve the local commissioning of integrated care in the country. In this endeavour, ICSs can learn from challenges faced by the old local commission structures as well as from the well-established national-level health technology assessment framework. In particular, the interviews conducted and the policy documents reviewed, highlighted the need for improvements in the collection and systematic use of data, and for an evidence-based decision-support framework, with a clear set of criteria, to inform local investment decisions. The interviews confirmed the relevance of this thesis, and findings from this chapter were used to define the structure of the decision-support framework, including the fundamental principles that should form the basis of the decision-support framework and the criteria that should be used for assessing new models of care.

The structure of the decision-support framework was further shaped by an updated systematic literature review on the application of multi-criteria decision analysis (MCDA) in different healthcare decision-contexts (**Chapter 3**). This review confirmed the increasing use of MCDA in healthcare literature, with almost half of the studies informing priority-setting decisions. Criteria commonly used to evaluate care delivery include safety, cost, and quality. However, there are variations in their use across and within decision-contexts. In terms of the MCDA approaches, almost half of the studies used the linear additive model whereas scales and the analytical hierarchy process were the most used techniques for scoring and weighting, respectively. Nevertheless, few studies considered axiomatic properties and/or adequately justified the selection of their chosen techniques. Many authors did not differentiate between the MCDA approach adopted and the scoring and weighting methods used. These flaws in the application and reporting across all MCDA steps compromise the transparency and legitimacy of the MCDA-based recommendations. The review concluded by underscoring the need for a guideline that directs the conduct and reporting of MCDA, tailored to the specificities of various decision-contexts and methods, thereby enhancing the method's utility and practical applicability in healthcare.

Chapter 4 outlined the foundations and components of the decision-support framework, drawing on insights from **Chapters 2 and 3**. The framework addresses the challenges local commissioners encounter in evaluating new models of care, developing a framework anchored in three principles: relevance, theoretical soundness, and practical applicability. In response to these principles, the framework integrates the monitoring and evaluation stages inherent in the commissioning cycle. To capture both the immediate results (output indicators) and longer-term changes (outcomes and impacts), the framework's monitoring and assessment criteria follow a logic model approach (230, 239). Under the monitoring component, local commissioners can categorise the most relevant key performance indicators based on the four monitoring criteria: (i) patient lifestyle and self-management behaviour, (ii) quality of care or features of the process of care delivery, (iii) compliance with national standards, and, (iv) equity in quality of care. The evaluation component of the framework is based on an MCDA with six assessment criteria, which were selected following the axiomatic properties of decision theory. This is one of the central elements of the framework as the assessment criteria reflect the value-based health care approach that ICSs follow, and that should guide local resource allocation decisions (413). As stated by Lewis (2022) when examining the strengths and limitations of value-based healthcare and its application in the UK context:

> "Understanding our outcomes is important. Rather than measuring outcomes as a dataset with which to reward or penalise providers of healthcare, they are arguably far more valuable as a set of information through which we begin to understand the needs of our patients and understand how we may direct our resources more effectively to meet those needs." (25, p.214)

Six assessment criteria were identified: (i) Final or intermediate health outcomes, (ii) considerations for quality of life and well-being, (iii) patient experience, (iv) equity, (v) size of the target population; and (vi) cost. These criteria extend beyond health-related outcomes and are comparable to those suggested by other studies in the context of integrated care in high-income countries (245). These broadly defined assessment criteria provide local commissioners with the flexibility to choose appropriate performance indicators for each criterion depending on the availability of routinely collected data. The data-driven

nature of the proposed decision-support framework responds to calls from the NHS England, urging ICSs to leverage data when making priority-setting decisions. The NHS Long Term Plan (2019), underscores the expectation for ICSs to base their decisions on data and evidence, highlighting the necessity of effective data use to understand population health needs, design services, and improve patient outcomes (24). The Health and Care Act 2022 further reinforces this directive, providing a legal foundation that supports ICSs in using data to guide their decision-making processes (10).

The use of MCDA facilitated the integration of societal preferences into the decision-making process through the weighting of the assessment criteria. **Chapter 5** presented the results of a discrete choice experiment (DCE) conducted to determine the relative importance of assessment criteria that should guide local investment decisions based on public preferences in England. The study's findings suggest that people place a relatively high importance on life expectancy, with people willing to pay an additional tax increase of £24 for one additional year of life. Improvements in health-related quality of life received a much lower relative importance score than 'Additional years of life'. In line with other studies, this finding suggests that, from a societal perspective, some Quality-Adjusted Life Years (QALYs) gains due to improvements in longevity are more valuable than QALY gains associated to improvements in health-related quality of life (278, 362, 363). Furthermore, the results indicate that the public values ‘Patient experience’ at least as much as improvements in quality of life, which highlights the need for measuring patient experience in the NHS and providing this data to decision-makers. These findings question the use of QALYs in certain decision-making contexts, reinforcing the arguments for using MCDA to provide a more comprehensive measure of overall value (278, 362, 363). Given the multifaceted nature of patient experience, which encompasses various aspects of care delivery (359), the chapter concludes with a call for consensus on a set of universal patient reported experience measures (PREMs) and the standardization of routinely collecting such data to support the systematic monitoring and assessment of new models of care.

The findings reported in **Chapter 5** emphasise the importance that the English population attaches on additional healthcare costs associated with care programmes, suggesting that a decision-support framework that omits cost implications may be incomplete. Arguably, this finding advocates for incorporating healthcare costs as a criterion in the MCDA's overall score, allowing local commissioners to explicitly account for the public's valuation of costs and their willingness to pay for improvements in other criteria (36, 122, 149, 264). However, the DCE revealed that although cost did not receive the highest value and its inclusion did not alter preference ordering, its addition in the choice task affected the valuation of other criteria disproportionately. This could favour the argument against including the cost criterion in the MCDA's overall score, with proponents of this approach arguing that cost is not a value element per se (210, 265, 266).

In line with the results obtained from the DCE and in recognition of the usefulness of both approaches, the decision-support framework proposes two value scores to rank competing alternatives and inform budget allocation decisions: the MCDA total score, including the cost criterion, and the cost-per-value ratio (CVR), calculated as the ratio between the unstandardized total cost of the alternative and the MCDA score excluding the cost criterion. The relative weights of the six criteria can be used to obtain a single composite measure for all different alternatives, making more explicit the contribution of each criterion, including healthcare costs (36-37). The relative weights derived from the DCE without cost can be used to calculate a cost-per-value ratio, with the denominator based on a multi-attribute benefit score. This approach facilitates estimating the expenditure required to achieve one unit of 'value', which can help to validate the efficiency of an intervention, relative to other (210, 266). This dual-scoring system allows local commissioners to better understand how costs influence the overall value of interventions and assess the balance between economic and intrinsic value considerations.

The case study reported in **Chapter 6** demonstrated the feasibility of implementing the decision-support framework developed in this thesis. The evaluation of two early intervention in psychosis (EIP) services

operating in Buckinghamshire, Oxfordshire, Berkshire West Integrated Care System (BOB-ICS), using routinely collected data, suggested that during the period from 2016 to 2022, the EIP services provided by Oxfordshire (Oxf EIP) outperformed those provided by Buckinghamshire (Bucks EIP). This conclusion remained valid when using the cost-per-value ratio, indicating that Oxf EIP services are more cost-effective than Bucks EIP. Therefore, in a scenario in which local commissioners had to determine which EIP service to scale-up across BOB-IC, arrangements and procedures from Oxfordshire EIP services might be preferred. This case study also underscores the usefulness to commissioners of presenting MCDA results both with and without the cost criterion. Differences in MCDA total scores (with cost) between the two EIP services were relatively small, despite the statistically significant difference in costs. This contrast with the observed relatively large differences in the CVRs, confirming that Oxfordshire EIP services are more cost-effective than EIP services in Buckinghamshire.

This case study illustrated how local commissioners can leverage routinely collected data, such as electronic health records of patients with psychosis, to systematically assess new modes of care and make evidence-based investment decisions. This is particularly relevant as the study was conducted in mental health, which has less developed data systems than physical care in England. Furthermore, this case study contributes to the increasing body of literature that demonstrates the ability of MCDA to support the local decision-making process (23, 36, 37, 245). MCDA is highly effective in providing a broader measurement of effectiveness that includes outcomes beyond years of life and health-related quality of life, as well as public preferences for these outcomes. Additionally, MCDA facilitates the generation of economic evidence in a systematic and ongoing manner. This enables timely and responsive decision-making suitable for local contexts, where short-term planning and adaptability are often crucial (23, 36, 37, 149, 408).

7.3. Challenges and limitations of this research

The decision-support framework developed in this thesis is founded on robust principles that aim to facilitate evidence-based decision-making at the local level in England, particularly in the context of integrated care. However, it is important to acknowledge the challenges in applying this framework, as well as the methodological limitations discussed in each chapter. This section outlines the main limitations of this thesis and the remaining challenges for the framework to effectively inform commissioning of new models of care in England.

First, the development of the decision-support framework places significant emphasis on the evaluation component but falls short in elaborating on the monitoring component. This is in response to the need for rigorous evaluation of new models of care, which contrasts with the progress that has been made on the development of monitoring tools and key performance indicators (KPIs) at the local and national levels in the country, as shown in **Chapter 2**. However, it is essential to acknowledge that the priority-setting frameworks tend to overlook the relevance of monitoring. Some of these frameworks fail to distinguish between output indicators and longer-term outcomes and impacts, which is critical for making informed short-term and long-term decisions, ensuring both immediate efficiency and sustained benefits (277). Furthermore, national initiatives such as 'Data Saves Lives' advocate for an integrated approach to health monitoring, emphasising the use of digital technologies and data analytics for effective healthcare delivery and decision-making (414). Given the critical role of monitoring in the commissioning cycle, this aspect warrants further research.

Second, although good practice guidelines in MCDA in healthcare were followed (47, 48), the assessment component of the decision-support framework is not entirely free of limitations. To start, the selection of the assessment criteria, which represent the elements of value-based healthcare, was subject to the limitations encountered when conducting the semi-structured interviews (**Chapter 2**) and

the updated systematic literature review (**Chapter 3**). The interview participant pool did not include representatives from national bodies such as NHS England or the Department of Health and Social Care. Their insights could have enriched the understanding of the numerous changes to the local commissioning of healthcare in the country. Additionally, in the literature review all searches were conducted by one reviewer, which might have affected the identification of relevant studies. A more significant limitation, however, lies in the review's focus on MCDA in healthcare, rather than on value-based care specifically. This may have resulted in a skewed identification of value elements towards those highlighted in MCDA literature, rather than a more targeted value-based care perspective. Despite these limitations, the use of both sources – interviews and systematic literature review - allowed the construction of a comprehensive list of potential assessment criteria. The final selection process, conducted methodically through a stepwise approach and grounded in decision theory axioms (192, 211, 232), would have nonetheless benefited from the direct involvement of local commissioners, potentially enhancing the framework's applicability.

Another limitation of the framework's assessment criteria is its exclusive focus on healthcare, overlooking the integration of health and social care. This omission is particularly relevant given England's ongoing agenda towards the integration of these sectors. As outlined in **Chapter 2**, local authorities, responsible for social care services, are integral to ICSs. Therefore, they are expected to work in collaboration with Integrated Care Boards, social care providers, and the voluntary, community, and social enterprise (VCSE) sector to develop strategies that improve health and social care services and overall wellbeing in their ICS areas. However, the different funding and delivery models of health and social care in the country pose challenges. For example, while health care is provided by the NHS, and is generally free at the point of use for all UK residents, social care services are means-tested and not necessarily free at the point of use (10, 13, 16). Despite these differences, a combined framework that addresses the broader care pathway, encompassing the wider wellbeing of the population, could potentially contribute to the integration agenda in the country.

The absence of explicit constraints in the decision-support framework developed could be considered as another limitation. This oversight might affect the framework's applicability in real-world healthcare settings, where limitations such as budget and practical feasibility play a crucial role in the local decision-making processes (29, 229, 276). To enhance the framework's practical utility, future developments should focus on integrating these constraints. Incorporating such elements could provide a more holistic and realistic tool for healthcare decision-making, aligning theoretical robustness with practical exigencies.

The framework's reliance on quasi-experimental methods could arguably be considered a limitation, given that systematic application of these methods requires access to data and advanced analytical expertise. Studies, including the interviews summarised in **Chapter 2**, have highlighted the lack of capabilities at the local level as well as the difficulties that commissioners face in terms of data access (28, 29). This concern, however, is mitigated to some extent by the ongoing investments ICSs are making in population health analytics, aiming to optimize data utilization. Local analytical teams are emerging nationwide, and the NHS has published guidelines to support ICSs in developing local intelligence functions (412). However, not all ICSs are progressing at the same pace (14, 159) and the application of quasi-experimental methods and MCDA methods at the local level demands substantial resources and support. Expertise in MCDA methods, stakeholder engagement, and clarity on integrating MCDA results into decision-making processes are some of the factors required (47, 48). Consequently, to facilitate the practical application of the decision-support framework, the development of an automated tool using algorithms is essential. This tool would empower local commissioners, who may lack coding or economic modelling expertise but possess an understanding of clinical data, to systematically apply quasi-experimental methods. Alongside the tool development, comprehensive training and guidelines are necessary to ensure effective use across different contexts, coupled with measures to guarantee easier access to datasets.

Another important aspect for consideration within the framework's assessment component is the exclusive reliance on the general public for deriving criteria weights when conducting the DCE (**Chapter 5**). Incorporating preferences of different stakeholder groups could have enriched the assessment of new models of care as not all criteria might hold the same importance to all stakeholders (355). International experiences and relevant literature suggest that the success of new care models is highly dependent on stakeholder alignment with the model's goals and achievements (ibid). However, the sample used in the DCE conducted was representative of the English general population, and as such, it can be argued that all types of stakeholders were included. Additionally, a former study exploring patients' and other stakeholders' preferences for outcomes of integrated care did not find statistically significant differences (293). It was also decided to focus on preferences from the general population as members of the public should voice their opinions in the priority-setting process in return for paying taxes (231).

The potential violation of the preference independence principle also warrants consideration. As indicated in **Chapter 4**, when computing MCDA total scores using the MAVT-based additive model the weight attached to one criterion should not depend on the performance on other criteria (211, 236). In **Chapter 5** the violation of this principle was statistically tested by assessing all potential interaction terms between attributes using multinomial logit models. Given that some interaction terms were statistically significant, correlated mixed logit models were used to estimate criteria weights (295, 345). However, no significant improvement in the model fit were found and the relative importance scores (i.e., criteria weights) did not change substantially. Although, based on these results, it can be reasonably assumed that the criteria are preferentially independent of each other, a certain level of mutual preference dependence cannot be completely ruled out.

The decision-support framework may have limited usefulness in the context of disinvestment decisions, a topic of growing interest in times of austerity and financial constraints. Disinvestment can be defined as the withdrawing of health resources from any existing health care practices, procedures, technologies or pharmaceuticals that are deemed to deliver little or no health gain for their cost (416). However, the literature on disinvestment is still in its early stages, with various definitions proposed and scarce evidence of successful implementation of 'disinvestment' projects in the local settings (417). The decision-support framework of this thesis was developed as a prioritisation tool, and as such, the assessment criteria defined may not encompass all the necessary criteria for making disinvestment decisions in local healthcare settings (418). While resource allocation and disinvestment in healthcare share some common elements, they differ in terms of their objectives, guidelines, and the unique challenges they encounter (e.g., unlike resource allocation, disinvestment decisions requires assessing the impact of withdrawing services) (418, 419). Nevertheless, the framework developed in this thesis could aid in identifying healthcare interventions that offer relatively low value and, therefore, might be considered as potential candidates for disinvestment.

A key limitation of this thesis was the inability to thoroughly investigate the usefulness and applicability of the decision-support framework. Despite the participation of local stakeholders in the interviews and exchanges in the some meetings and conferences, it was not possible to obtain proper feedback from ICS members on the suitability and potential applicability of the framework to inform local commissioning decisions. While the case study in **Chapter 6** demonstrated the feasibility of implementing the decision-support framework for assessing new models of care in England, the findings were not socialised and discussed with local commissioners. This limited the opportunity to receive feedback and validation on the practicality and relevance of the proposed framework. Part of this thesis was conducted during the COVID-19 pandemic which significantly hindered in-person interactions and reduced the availability and capacity of local commissioners to engage in external research projects. Additionally, during the period of this thesis, substantial changes were introduced to

the local healthcare commissioning architecture, including the transition from Clinical Commissioning Groups (CCGs) to ICSs, placing further demands on local commissioners and constraining opportunities for comprehensive workshops or meetings. The need for appropriate consultation with local commissioners and other relevant stakeholder remains essential to demonstrate the usefulness and applicability of the decision-support framework.

7.4. Implications of this research

Contribution to the literature

As stated in **Chapter 1**, the aim of this thesis was to contribute to four research areas: the application of MCDA in healthcare, discrete choice experiments in health economics, the economic evaluation of new models of care in mental health, and local priority-setting in the context of integrated care.

First, the evolution of integrated care in England and results of the 26 semi-structured interviews conducted have been published in peer-reviewed journals (**Chapter 2**). The examination of policies and reforms that have affected care integration over the past 30 years was published in a comparative study (18). The second study discusses the investment decision process in integrated care amid the transition toward ICSs. It highlights challenges in the local commissioning of healthcare, and provides recommendations to inform ICSs in their healthcare commissioning role (229). This was the first study to explore how local commissioners in England invest in integrated care. Both studies have thereby made a contribution to the field of local priority-setting in the context of integrated care.

The results of the updated systematic literature review (**Chapter 3**) have also been published in a peer-reviewed journal. This is first review that includes grey literature and provides an extensive examination of the use of MCDA methods in various healthcare decision-making contexts. The set of MCDA criteria identified could be used as a basis to start developing a uniform set of top-level criteria by decision

context in healthcare (240). However, it also identifies deficiencies in the application and reporting of all MCDA stages, which undermines the credibility and systematic implementation of MCDA in healthcare. Therefore, the review recommends the development of a guide that recognises the unique characteristics of different decision contexts and methods for conducting and reporting MCDA. Overall, this review contributes to the existing literature on MCDA in healthcare.

The DCE in **Chapter 5** contributes to the field of choice experiments in healthcare. It is the first study in exploring the factors that local healthcare commissioners and members of the public in England would give precedence to inform decisions in the context of integrated care. The relative weights and marginal rate of substitution derived in the study, including willingness to pay estimates, could be used when conducting other type of evaluations, such as social return on investment or cost-benefit analyses (3, 23). Furthermore, the experiment conducted was the first one in using a split-sample design that expresses costs in terms of income tax increase. Previous studies have used an out-of-pocket expenditure or changes in insurance premiums based definition. In principle, this suggests that the type of payment vehicle used to define the monetary attribute does not seem to affect choice preferences in health (310).

The service evaluation conducted in **Chapter 6** contributes to the field of evaluating new models of care in mental health. The evaluation of two EIP services adds to the relatively few comprehensive and robust evaluations in the mental health sector (242, 376, 385). Additionally, it adds to existing literature in the MCDA field that has demonstrated the usefulness of MCDA to assess healthcare interventions, particularly in the context of complex interventions (36, 37, 149, 408). This evaluation is particularly relevant considering the significant burden of mental health conditions in the country and the current lack of comprehensive rationing frameworks for assessing mental health programmes (378, 421).

Finally, the decision-support framework developed in this thesis has contributed to the field of local priority-setting in the context of integrated care. This thesis introduces the first framework that integrates the monitoring and evaluation stages inherent in the commissioning cycle (**Chapter 4)**. The framework proposed defines a comprehensive set of value-based elements that are applicable and allows for explicit recognition of public preferences in the local priority-setting process. Results of the choice experiment conducted (**Chapter 5**) and the case study (**Chapter 6**) highlight the need of reporting the MCDA total score both with and without the cost criterion, as people do place significant value on the additional healthcare costs that a health intervention entails, but it has an uneven impact on the other criteria or elements of value. The case study also demonstrated that it is possible to use the MCDA-based framework developed in this thesis to generate evidence on the performance of new models of care in England.

Potential policy implications

The analysis of the commissioning experience from the CCGs, including the identification of potential areas for improvement, may help ICSs to integrate care across different services and allocate the NHS budgets to maximise population health, improve patient experience, and decrease health inequalities. It is expected that insights from the semi-structured interviews conducted (**Chapter 2**) may encourage open discussions within the emerging ICSs about the barriers and enablers of local commissioning of integrated care. Study results were already presented to the local commissioners involved and used for discussions within some of the ICS' teams (229).

The updated review conducted in **Chapter 3** provides guidance for potential future MCDA applications in healthcare. The overview covers decision-contexts, criteria, methods, stakeholders involved, and sensitivity analyses. This information can assist both researchers and decision makers in selecting appropriate methods and criteria relevant to their decision-context. The thesis extensively discusses the importance of adhering to good practice guidelines when using an MCDA in healthcare. **Chapter 4**,

which covers the MCDA-based assessment component of the proposed framework, and the case study summarised in **Chapter 6**, provide an example of how to report and make methodological decisions when conducting an MCDA.

The thesis emphasises the need for advancements in the collection, access, and utilization of clinical and patient-reported outcomes data, particularly concerning patient experience. The high relative importance that the general population in England attaches to patient experience (**Chapter 5**) underscores the need for measuring patient experience in the NHS, in a systematic way, and providing this data to decision-makers. Ongoing initiatives to routinely collect data on continuity and coordination of care, responsiveness to patient concerns, the opportunity of care, and professionals-patient communication should be strengthened (361, 422). The semi-structured interviews conducted (**Chapter 2**) have highlighted a keen interest in data-driven commissioning of integrated care, underlining the imperative to enhance the availability and quality of data for such purposes. This improvement should focus on making routinely collected data easily available and accessible to local commissioners, and ideally this information should be electronically available for ongoing analysis and not merely confined to annual audit reports (25, 412, 414).

The framework developed in this thesis holds potential in supporting the local commissioning of new models of care in England. While the local commissioning of healthcare operates under a complex and interactive system that hardly follows a rigid orderly logic (27, 157, 423, 424), a value-based framework is desirable for enhancing efficiency, transparency, and accountability. If the framework's applicability and usefulness are tested and barriers to adopting priority-setting tools, as identified in this thesis and elsewhere, are addressed, the proposed framework could support explicit and transparent local decision-making processes (424). It could also offer opportunities for formal deliberation, allowing stakeholders to engage with comprehensive information that can underpin their decision-making.

The findings of this thesis and the decision-support framework developed are aligned with the priorities set by NHS England for 2023/24 (425). The immediate priority is to recover core services and productivity, and the proposed framework serves this purpose by helping local commissioners identify inefficiencies and make better-informed decisions when replicating or adapting care models. The second priority is to make progress in delivering the key ambitions of the NHS Long Term Plan. This involves advancing local empowerment and accountability, implementing new models of care, and improving mental health services. The proposed framework facilitates the local monitoring and evaluation of new models of care, filling the gap of traditional analytical methods. The framework builds on principles of relevance, theoretical soundness, and practicality, with clearly defined components that can potentially increase transparency and legitimacy of local commissioning decisions (426). Additionally, the case study conducted, which evaluates the performance of two EIP services in BOB-ICS, contributes to the commitment to enhance mental health services in the country.

Lastly, while the decision-support framework was tailored for local commissioners in England, its findings and methodologies have potential applicability in various other decision-making contexts. As previously mentioned, findings from the updated systematic literature review (**Chapter 2**) can be used to guide the application of MCDA in various healthcare decision-making contexts, such as clinical decision-making, regulatory decisions, and research and development, in high, middle and low-income countries. Additionally, with the growing global focus on healthcare integration, as evidenced by the increasing number of documents on integrated care released by the World Health Organisation (WHO) and other international bodies (22, 427), the lessons and findings from this research could be valuable to local healthcare decision-makers in other high-income economies. For instance, while the relative criteria weights might not be directly transferrable to other high-income countries due to documented cross-country differences in preferences for integrated care (293), the assessment criteria or elements

of value-based healthcare defined in this framework could hold relevance. Despite their potential applicability, these criteria should be validated by decision-makers in each respective country.

7.5. Steps forward

This section outlines actions to advance the development of the proposed decision-support framework and provides suggestions for future research based on the findings and limitations discussed earlier in this chapter.

First, it is essential to present and widely discuss each component of the framework with local commissioners and other relevant stakeholders. This includes presenting the results of the service evaluation conducted in **Chapter 6**, in order to make the necessary adjustments to the framework, and with this, ensure the usefulness and applicability of the decision-support framework. Furthermore, it is crucial to evaluate the relevance of the monitoring component of the framework in relation to the existing monitoring tools and performance indicators utilised by ICSs. This consultation process could also offer an opportunity to explore strategies for incorporating all pertinent dimensions of patients' wellbeing into the framework, given its relevance in the context of integrated care. Similarly, this process can investigate methods for integrating social care considerations into the framework. Although the assessment criteria are broadly defined, they may not encompass particular aspects of social care, such as the availability of social services or the wellbeing of caregivers. As indicated in previous sections, integrated care is becoming increasingly important as new models of care are developed. This involves coordinating health and social services to provide a more comprehensive approach. To capture patient experience, a composite measure could be developed that takes into account both health and social care, as well as the experiences of patients, their families, and carers. However, the feasibility of this measure would depend on the availability of data.

The potential impact of healthcare investment decisions on other sectors could be further explored. This could be particularly relevant for ICSs, as local commissioners are encouraged to collaborate with various stakeholders – including local authorities, social care providers, the VCSE sector, and entities involved in health and wellbeing like education, housing, and emergency services (7, 12). For instance, it may be worth considering incorporating carbon emissions concerns into the decision-support framework. The health and social care system accounts for a significant amount of resources and carbon emissions, making up approximately 4-5% of England's total carbon footprint (428). While the public may not view the NHS as a contributor to the climate crisis, the NHS has committed to reaching carbon net zero and local commissioners should consider methods to include such commitment in their decision-making process (ibid).

Third, conducting a service evaluation of EIP services in other ICSs across England, similar to the one conducted for BOB-ICS, can significantly contribute to the literature by providing comparative insights and a broader understanding of EIP effectiveness and efficiency in diverse healthcare settings. This comparative approach would allow for the identification of best practices, variations in service delivery, and the impact of different local policies and resource allocations on patient outcomes. Such evaluations would also offer valuable data for refining EIP models, addressing regional disparities, and informing national strategies for mental health services. **Chapter 6** proposes performance indicators, with data sources potentially available for different provider collaborative across the country. Additionally, it suggests methodological approach to assess the performance of these models of care in mental health. This could facilitate the replication of the evaluation to different regions across England.

Fourth, the thesis identified gaps in current datasets through semi-structured interviews (**Chapter 2**) and the case study (service evaluation) in mental health (**Chapter 6**). However, a more detailed analysis of the available dataset, limitations in accessing the data systematically, periodicity, and database quality would benefit ICSs. Special consideration should be given to patient experience data, as there

is a need to establish a universal set of PREMs and standardize the collection of data to support the systematic monitoring and evaluation of new models of care. Furthermore, an evaluation of how this data is being used could help identify the capabilities or professionals (profiles) required for ICSs to make the most of the available dataset.

Fifth, to make the decision-support framework more practical and increase its usage across the country, an online automated tool could be developed. This tool should include comprehensive guidelines and associated software that allows for the application of quasi-experimental methods in a user-friendly manner. In addition to facilitating the use of the decision-support framework, the suggested online tool could also promote exchanges between ICSs across England.

Sixth, the thesis underscores the need to develop comprehensive methodological and reporting guidelines for the implementation of MCDA in healthcare. This guide should include a universal classification of MCDA methods, guidance on their application, and a common set of criteria or macro domains for each decision context in healthcare. Additionally, a checklist with standards for reporting, similar to the Consolidated Health Economic Evaluation Reporting Standards statement used in traditional economic evaluations could be provided (429). The creation of this methodological and reporting guide should draw upon empirical research and the accumulated expertise of scholars and decision-makers in healthcare. Such a guide would enhance transparency and clarity in MCDA applications, thereby bolstering the credibility and systematic implementation of MCDA across different healthcare decision contexts.

Last, future choice experiments could investigate preference heterogeneity among stakeholders in integrated care in England. A study measuring the relative preferences for outcomes of integrated care across eight European countries, for instance, found that although preferences do not differ much

between patients, caregivers, professionals, and payers/policy-makers, some particular differences could inform policies related to the delivery of care (293). Further research on the impact of the monetary attribute on people's preferences is also needed, as results might vary by type of respondent, particularly in publicly funded health systems, with payers/decision-makers being more sensitive to the programme's cost than patients or carers (ibid).

7.6. Concluding remarks

This thesis has developed a decision-support framework designed to enhance the local commissioning of integrated care in England, underpinned by evidence. The framework, informed by literature reviews and insights from commissioners, is founded on the principles of relevance, theoretical soundness, and practicality. It stands out as the first to integrate monitoring and evaluation stages, inherent in the commissioning cycle. In addition, it defines a comprehensive yet applicable set of value-based elements, while allowing for the explicit recognition of public preferences in the local priority-setting process. The case study conducted demonstrated the framework's feasibility in mental health, a key priority area for NHS England. Despite the encountered limitations and challenges, the empirical work presented in this thesis has generated new evidence in the application of MCDA and DCEs in healthcare, the evaluation of new models of care in mental health, and the local priority-setting in the context of integrated care. These findings offer valuable insights for researchers seeking to further this field and for local commissioners in England. This is especially relevant in a context where expectations and demands for integration are higher than ever, and where local commissioners, operating under stringent budgets, are compelled to enhance efficiency alongside improving healthcare quality and outcomes.

Appendices

Appendices – Chapter 2

Appendix 2.A. Semi-structured interview guide

1. GENERAL BACKGROUND

1. Can you tell me a bit about your role? Which team are you part of? And, for how long have you been in the CCG/STP/ICS?
2. Are you involved in the decision-making process around integrated care programmes?

2. INTEGRATED CARE PROGRAMMES

1. Could you tell me which integrated care programmes (or programmes with elements of integration) are in place at the moment in the CCG/STP/ICS, and how do they work?
2. Who set the budget across health programmes within the CCG? How are financial flows determined and influenced? Is there a specific budget for integration?
3. How would you define integrated care programmes?

3. ASSESSMENT OF HEALTH INTERVENTIONS

1. In the context of integrated care programmes, how are health interventions assessed?
2. What kind of data or information is normally used for that purpose, and where does the data come from?
3. What would you say are the main constraints or challenges that the CCG/STP/ICS face when assessing health interventions?

4. COMMISSIONING

1. Who are generally involved when commissioning integrated care programmes? Any organisation from outside?
2. How would you describe the approaches taken in this local authority to develop priorities for integrated care spending or disinvestment? Is there any decision-making tool in place? What kind of information is used to inform prioritisation decisions?
3. What sorts of factors influence commissioners to invest in one integrated care programme over others? How is this going to change in the light of the transition?
4. What do you see as the main barriers that the CCG/STP/ICS face when making commissioning decisions of integrated care programmes?

Appendix 2.B. List of interviewees

Participant No.	Role
1	Programme manager – CCG/STP
2	Pathway lead - CCG/STP
3	Executive lead - CCG/STP
4	Pathway lead - CCG/STP
5	Accountable officer - CCG/STP
6	Director - CCG
7	Programme manager – CCG
8	Head/lead commissioner – CCG
9	Public health registrar - County Council
10	Programme manager – CCG
11	Chair/Executive lead – CCG
12	Director – CCG
13	Chief officer - CCG

14	Director – CCG
15	CSU member
16	Director – CCG/STP
17	Former Chair/Executive lead – CCG
18	Head/lead commissioner – CCG
19	CSU member
20	Chair/Executive lead – CCG/STP
21	Chair/Executive lead - CCG
22	Head/lead commissioner - CCG
23	Head/lead commissioner – CCG/STP
24	Director - CCG
25	Public health registrar - County Council
26	CSU member

Appendix 2.C. Local structures under the new commissioning context

Structure	Definition
Integrated Care Board	"A new type of NHS body that bring the NHS together with partners across their area. Integrated care boards are responsible for developing a plan for meeting the health needs of the population, managing the NHS budget, and arranging for the provision of NHS services in their area" [1]
Integrate Care Partnership	"Each integrated care board and their partner local authorities will be required to establish an integrated care partnership, bringing together health, social care, public health (and potentially representatives from the wider public space where appropriate). Integrated care partnerships will be tasked with promoting partnership arrangements to address the health, social care and public health needs of their system" [1]
Health and Wellbeing Board	"A committee of the partner local authority, often considered to be at 'place' level (although 'places' are locally determined, this will not be the case in all areas). Established by the Health and Social Care Act 2012 they must include listed representatives from the local authority, local Healthwatch, and each integrated care board in the local authority's area, alongside such other persons as the local authority considers appropriate. They produce a joint strategic needs assessment and a joint local health and wellbeing strategy."[1]
Place-based partnerships	"Collaborative arrangements formed by the organisations responsible for arranging and delivering health and care services in a locality or community. Place-based partnerships will remain as the foundations of integrated care systems as they are put on a statutory footing (subject to legislation), building on existing local arrangements and relationships" [2]
Provider collaboratives	"Provider collaboratives are partnership arrangements involving at least two trusts working at scale across multiple places, with a shared purpose and effective decision-making arrangements, to: • reduce unwarranted variation and inequality in health outcomes, access to services and experience • improve resilience by, for example, providing mutual aid • ensure that specialisation and consolidation occur where this will provide better outcomes and value. Provider collaboratives work across a range of programmes and represent just one way that providers collaborate to plan, deliver and transform services. Collaboratives may support the work of other collaborations including clinical networks, Cancer Alliances and clinical support service networks." [3]

1. Department of Health & Social Care. Annex B: glossary. Definitions of terms found in 'Guidance on the preparation of integrated care strategies' 2022 [cited August 2, 2022; updated July 29, 2022]. Available from: https://www.gov.uk/government/publications/guidance-on-the-preparation-of-integrated-care-strategies/annex-b-glossary
2. NHS England and NHS Improvement. Thriving places. Guidance on the development of place-based partnerships as part of statutory integrated care systems. London, UK; 2021 September 2, 2021. [cited August 2, 2022]. Available from: https://www.england.nhs.uk/wp-content/uploads/2021/06/B0660-ics-implementation-guidance-on-thriving-places.pdf

3. NHS England and Improvement. Working together at scale: guidance on provider collaboratives. 2021 August, 2021. Report No.: PAR754. [cited August 2, 2022]. Available from: https://www.england.nhs.uk/wp-content/uploads/2021/06/B0754-working-together-at-scale-guidance-on-provider-collaboratives.pdf

Appendix 2.D. Supplementary quotes and information provided by interviewees

Table 2.D.1. Examples of integrated care programmes in place in two ICSs from South East England, mentioned by interviewees

Condition-specific programs	Programs across conditions
Integrated diabetes care programme (No.1) *Cardiovascular disease programme (No.16)* *Musculoskeletal programme (No. 22)* *Ophthalmology programme (No. 8)*	*Aging Well programme (No. 1)* *Enhanced Health in Care Homes (No. 1)* *Urgent Community Response (No.11)* *Primary Care Transformation Programme (No.13)*

- *"you know, you don't refer to all them as programmes, formally speaking, but rather you, you use another terminology, other name to refer to all them. Yeah. It'll be, it'll be, it'll be, it'll be 'projects', 'programmes', 'workstreams' (No. 2)*
- *"[…] I think of it sometimes because, you know, we've got very different organisations with all their own individual cultures, objectives, all the rest of it. And they, it, it, you know, the reality of integration is that you've got your own governance processes within those organisations, so you have to … (noise) Sorry, that's the front door, you have to, you know, you have to meet the requirements of your own individual organisations as, which means that integration quite often means additional work rather than, it's an 'as well as' rather than an 'instead of'.(No. 24)*
- *"But I would, I would say that probably most of the, most of the integration work is on a more of a partnership basis, through things like memorandums of understanding and joint working, rather than vertical integration end of the sphere." (No. 23)*
- *"I think that has been limited to date, in my view, by the combination of a lack of, of real understanding of need, of, you know, of, of population data and also the, the, the, the, the great difficulty that systems and senior managers have responding to this because of the combination of the, you know, what, the, the cultural challenges of, of people being locked into roles, you know, where this is what we do, I do this, blah blah blah, and the, the central demands that come down all the time, you know, you've got to launch this service or you've got to do that. So I think it, it is very challenging although, you know, there are starting to be some evidence of success" (No. 17)*

Investment decision process

- *"So, how it normally works is that the commissioners, like the colleagues you've already spoken to at the CCG, they will identify a need for a service, and they will, will sort of work on the, the basic structure and the requirements of that service in the service spec" (No.12)*
- *"R: So how much flexibility would you say the CCG has to, to set local priorities?*

  I: Very little really. It's very centrally driven and finance driven. So although we spend a lot of time, you know, trying to behave as commissioners and coming up with good ideas, you have to demonstrate that the, the main demonstration is that it essentially saves money or that it will save money, not that it's necessarily the best thing to do." (No. 21)
- *"But I think in terms of the decision making, I mean, you would need to go through the CCG decision-making processes about something like this. So say you were gonna commission a new service, for instance, or you were gonna fundamentally change the way you commission something, and particularly if it has a financial implication, what you would, what you would do is, is, you would do a, a report for the, the Executive Committee. I mean, you, you would do a report generally. So I would, in my case, for instance, I would normally take that, you know, I'd, I'd work up a report or a briefing around this, and a proposal, if you like, with, including any, kind of, financial and contractual implications. Now, I would, in the first instance, sort of, work that up and take that, you know, through to, you know, like, the head of Planned Care, who's my boss, and take, you know, for comment. And, and then take it through to, say, the, the Planned Care Programme Board within the CCG in the first instances. After that, you know, including, kind of, you know, any comments around, around that, I'd then, sort of, take it through to our director for, for our, for us, which is the, the, the Deputy Chief Executive of the CCG, who we would take that to. And then she will take a view on it and, you know, and, you know, and decide whether it should go through, it will go through to the Executive Committee for a formal, kind of, decision, or, you know, and, you know, to maybe get advice on this from the Director of Governance as to whether it needs to go through to a formal decision making at the, the CCG board, the full board meeting" (No. 10)*
- *"I don't think it'll change hugely, but I think that's also to be determined, because somebody's still gonna have to make decisions about how the money is spent" (No. 16)*
- *"so whilst we're awaiting legislation under the white paper, we are working on the assumption that we as an integrated care system will receive the allocation for the total population of [name of ICS]. And that will sum to £[.] billion, there or there about. And then we're trying to work through how we will, how we will allocate that resource and whether that allocation will be routed through a series of contracts for the provision of services. Or whether we will be delegating the responsibility for securing services for a population through to different sub-ICS structures like local authority pooled budgets, or provider collaboratives, or an alternative set of arrangements. So that's the bit of design we're going on at the moment." (No.5)*
- *"Although until now, until the merge of the CCG came into effect all the governance of decision-making was in the local CCG's. So, very little was done centrally." (No. 6)*

Prioritisation of interventions

- *"National priorities, areas of local needs, that kind of thing. So areas of, you know, that have been identified as areas of local need. Or quite often like if there's a problem, it, it, it, there, if there's a*

problem with a service, although that wouldn't be investing in one over and above another, it would just be focusing on one." (No. 24)

- *"I would say, and it is cynically, there are a number of drivers for the areas of work we develop. There are, there are very clear national drivers established from central government […] There are other things that come in because we are, we know there are challenges, failures within our existing system […] There are other things that then start to get more interest, pressure based, personal interest based, which, sort of, come from a wide number of sources. So, you might have a particular clinician that's really keen on doing one thing, and has good networks, good ways of communicating that, good ways of on-boarding people to their, to their issue. And that develops, sort of, a, a momentum of itself. So, there are individual, individuals who, who are championing a cause in a very powerful way that that gets done" (No. 7)*

Assessment of interventions

- *"there's a number of ways that we would do that within the CCG. So you've got, you've got something like, for instance, the Integrated Performance Report […] you will, there'll, there'll also be, you know, certain metrics and things that will be gathered from things like the Quality and Outcomes framework in Primary Care […] So, and you'll also have things like the [place] Joint Strategic Needs assessment […]" (No. 10)*
- *"So, there'd be hard metrics of, of, of, you know, patient metrics of money, attendances, is one way. And that's, you know, and then performance against constitutional standards, which we always have to look at. But then there would be patient reported outcome measures, and patient confidence measures." (No. 11)*
- *"[…] the differential impact of Covid by ethnicity over the past 12 months has highlighted the importance of data and we know that some of our data, in terms of impact and burden of disease, either at very high resolution within different parts of the county or for different potentially more marginalised in one of the population groups, isn't as good as it should be." (No. 25)*
- *"So the traditional commissioning assessment of things is through I guess three main possible routes, the, the top route usually being, "Is it costing us more than we can afford? Which is a very familiar problem and one that CCGs are obsessed about for the length of their existence. The other one is a quality perspective, so are we seeing the kind of outcomes we would expect for patient care? Are we seeing serious events that we would want to prevent? That kind of stuff. And the third one is probably general performance, so for example around cancer care, we work closely with the services delivering cancer care, particularly looking down to tumour site, add (?) waiting times and the processes in place within providing organisations to make sure waiting times are not any longer than they absolutely have to." (No. 18)*
- *"It will vary by programme. So, if they're … so, I don't, I don't think we have any, well in fact I know, we don't have kind of KPIs related to "integration" and what integration means and how you might measure that and how you might understand success with that. So, that isn't … doesn't happen. But then of the programmes that we might class as being integrated, then they will have with their own individual KPIs in terms of process and outcome measures." (No. 25)*

Challenges in the commissioning process

- *"*[…] *So I, I think again, I mean CSUs are used because CCGs have to use them […] we need a, a better system. It needs to be, they need to be owned by the integrated care systems not some external*

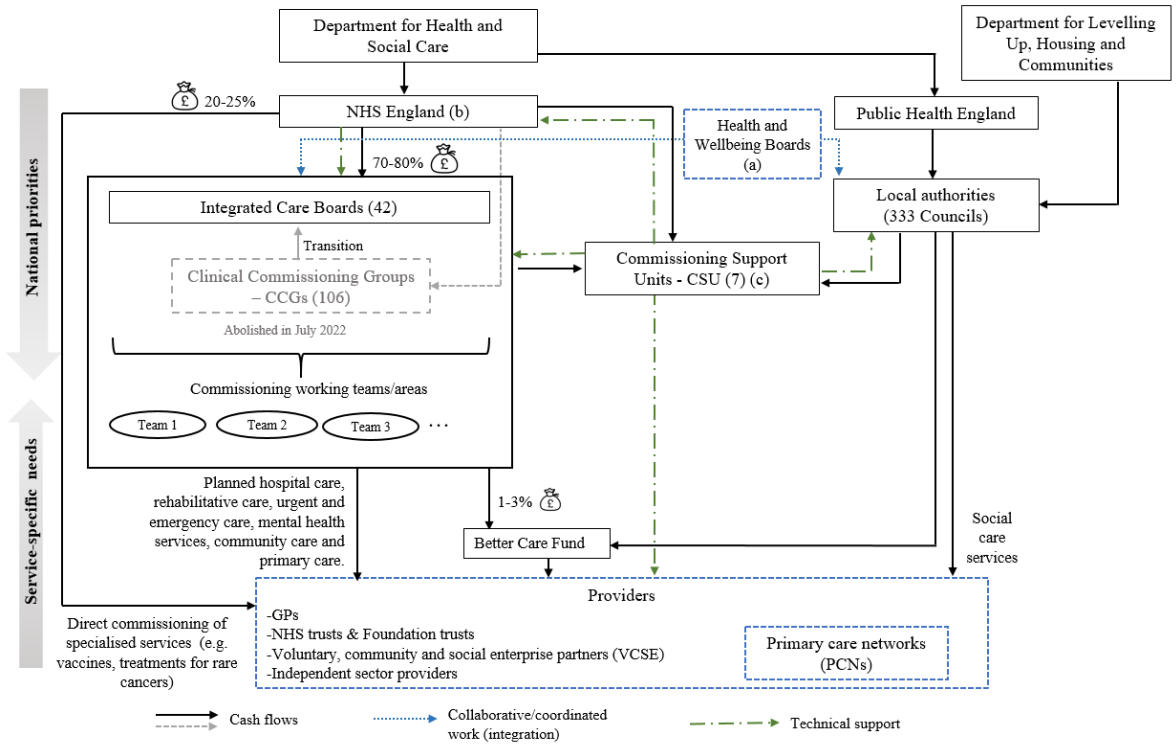
consultancy advisory which just peddles the same stuff and sells it five times to different people, different CCGs" (No. 17)

- *"You, you know, we need to start commissioning by people's needs, rather than, than services. I suppose some of the restrictions to that are the, the length of contracts that we have for different providers who provide these services. They don't always end up the, same time, so, you know, we, we, we'll renew a contract, but what it doesn't, what we don't take into consideration is actually, "Right, these, these people need this bundle of care, and what, what contracts are we going to, to consolidate and integrate to, to achieve that? We're, we're, we don't tend to end contracts early. We don't tend to, to do a lot of procurement, to, to achieve that." (No. 12)*
- *"Historic focus on organisational targets can be distracting from taking a step back and prioritising a … prioritising something. The procurement rules are very constraining and really don't encourage people to do … make many changes, if any. And the third thing that's a constraint is the … the complexity of delivering an integrated out … care pathway outcome. It demands so many facets that you can't easily evaluate the integrated component. You can measure activity by organisation, but … but we don't have joined up data sharing yet. So, it … to do any integrated pathway at the moment relies on every organisation sharing their component part of an integrated pathway. Joined up data would enable that to be more easily done, so that we measure the outcome by the person's experience of the pathway, rather than by the organisation's intervention. (No. 20)*
- *"the money side of things is obviously a massive issue, because they have got so many things that come down from NHS England that they are meant to achieve, and, you know, and, and the funding just isn't there, and at the same time, they also know what their local population needs, and there may not be funding available to do that. So, you know, that's a, that's the, that's one of the big things. And, and having the right staff is quite often an issue, you know? There's … I've been on projects over the years in the NHS where it has surprised me about how little has been achieved because it's decisions by committee, rather than just getting on and achieving something" (No. 15)*
- *"I think the other thing I've not mentioned that is an important driver and one that's deeply frustrating is because of the financial regime we work within, everything has to deliver within a year which is not realistic." (No. 18)*
- *"And we started to make progress, some areas more than others. But then the pandemic hit, and it really meant we, we had to quite rightly pause everything, as a lot of services got stood down." (No. 8)*
- *"Well, I think obviously there's always a finite financial envelope, isn't there, there's always, you know, there's not an unlimited pot of money that we can spend and I think we all acknowledge, you know, the, the impact of Covid and how, you know, the impact that that, the expenditure on Covid is going to have on, on the NHS and what it might mean for future budgets and just generally, you know, the, the, the amount that the government spends on the health service can't just keep going up and up and up and up, you know, there are limits on what we can spent." (No. 2)*
- *"Yeah. We are working much closer with one another. The information constraints and also the governance structures around Covid are going to help that as well. So, the different boards, the Health Protection Board, part of … sort of the local silver and bronze structures, etcetera." (No. 25)*

- *"The transition from CCGs to integrated care systems is creating more of an opportunity for providers to collaborate together and to together take a responsibility for their population, rather than to assume it's the CCG's responsibility." (No. 20)*
- *"Well, I think on paper it should definitely help. Whether it, whether it will I think remains to be seen. But the fact that there's less different organisations should help. There's only the one CCG that eventually, I think, provide commissioners sit around one table, on paper looks like it should help with those sort of conversations." (No. 3)*
- *"I still think it's early days and it's not clear yet what will be done centrally at scale and what will still be left to local decision. I think I think probably that that's the big challenge for next year is is to describe that narrative about how you run programs at scale, while still having local accountability and local decision-making. This is one of those things there's not a right answer. You've got to have the local you've got to have the at-scale, the Community program I'm involved in is a really good case in point. You know so…we work as as a group of commissioners across [locations' name] footprint defining care pathways, doing evaluation impacts… uhm but the conversations about how we move the money or support are still local conversations, because of course you've got local trusts are the building blocks to this, and they're the ones which are feeling the impact. So, I think there's a real opportunity… I think that the ICS Financial framework, which is being developed, will be key to this, because It's all about governance; If we get the governance right and get the right decisions taken at the right point." (No. 6)*

Appendix 2.E. The commissioning decision-context before the abolishment of CCGs

Figure 2.E.1. The commissioning decision-context before the abolishment of CCGs



Percentages (%) in the figure refer to proportion of the total NHS England budget. (a) Some have other joint commissioning structures, such as Health, Education and Social Care, Social Care boards. (b) Since April 2019, NHS England and NHS Improvement work together as a single organisation. (c) NICE, the NIHR and the Academic Health Science Networks could also offer some technical support to CCGs and providers.

Appendices – Chapter 3

Appendix 3.A. Results of searches

Table 3.A.1. Academic studies

Date searched	Search engine	Search terms	Number of Results	Total results
11-Nov-20	Excerpta Medica Databases - EMBASE (a)	"Multi-criteri* decision*', `Multicriteri* decision*´, `Multiple criteri* decision*´, `Multicriteri* analysis´ `MCDM´ or `MCDA´";	1,679	**4,072**
		`Benefit risk assessment*´; `Risk benefit assessment*´	1,018	
		`Multi-objective optim*´ `Analytical hierarchy process´, `Multi-attribute utility´, `Multiattribute utility theory´, 'Weighted product method´, 'Analytic network process´, `Goal programming', `Elimination and choice expressing reality´	1,171	
		`Preference ranking organization method of enrichment evaluation OR PROMETHEE´, 'weighted product model', 'measuring attractiveness by a categorical based evaluation technique OR MACBETH', technique for order preference by similarity to ideal solution'	204	
11-Nov-20	Ovid MEDLINE (a)	"Multi-criteri* decision*', `Multicriteri* decision*´, `Multiple criteri* decision*´, `Multicriteri* analysis´ • `MCDM´ or `MCDA´"; `Benefit risk assessment*´; `Risk benefit assessment*´	1,778	**2,856**
11-Nov-20		`Multi-objective optim*´ `Analytical hierarchy process´, `Multi-attribute utility´, `Multiattribute utility theory´, 'Weighted product method´, 'Analytic network process´, `Goal programming', `Elimination and choice expressing reality´	913	
11-Nov-20		`Preference ranking organization method of enrichment evaluation OR PROMETHEE´, 'weighted product model', 'measuring attractiveness by a categorical based evaluation technique OR MACBETH', technique for order preference by similarity to ideal solution'	165	
23-Nov-20	Google Scholar - first 100 hits (b)	"Multi-criteria decision" health; "Multicriteria decision" health; "Multiple criteria decision" health and "Multicriteria analysis" health	31	**31**

Table 3.A.2. Grey literature: Governmental entities

Date searched	Organisation	Total	Studies retrieved
23-Nov-20	UK National Health Service NHS Google: nhs.uk	13	7
23-Nov-20	Public Health England https://www.gov.uk/ (using directly the search engine from the website)	3	2
23-Nov-20	Public Health England Google: gov.uk	123	0
23-Nov-20	National Institute for Health and Care Excellence-NICE Google: nice.org.uk	21	10
24-Nov-20	National Institute for Health Research-NIHR (UK) Google: nihr.ac.uk	50	16
23-Nov-20	Health Technology Wales HTW (UK) Google: healthtechnology.wales	0	0
12-Nov-20	Healthcare Improvement Scotland - HIS (UK) Google: healthcareimprovementscotland.org	3	0
12-Nov-20	Health Information and Quality Authority – Ireland Google: hiqa.ie	0	0
12-Nov-20	Social & Health Services and Labour Market – Denmark Google: defactum.net	0	0
25-Nov-20	Finnish Coordinating Center for Health Technology Assessment – Finland Google: fincchta.fi	0	0
25-Nov-20	Norwegian Institute of Public Health - Norway Google: fhi.no	3	3
25-Nov-20	Swedish Agency for Health Technology Assessment and Assessment of Social Services, Sweden Google: sbu.se	0	0
25-Nov-20	Swiss Federal Office of Public Health - Switzerland Google:bag.admin.ch	2	1
26-Nov-20	The Federal Joint Committee (Gemeinsamer Bundesausschuss) – Germany Google: g-ba.de	0	0
26-Nov-20	Institute for Quality and Efficiency in Health Care, IQWiG – Germany Google:iqwig.de	13	5
26-Nov-20	Agencia de Evaluación de Tecnologias Sanitarias – Spain Google:isciii.es	7	5
26-Nov-20	The Agency for Regional Healthcare - Italy Google: agenas.it	0	0
26-Nov-20	Haute Autorité de santé (HAS) or French National Authority for Health – France Google:has-sante.fr	4	2
26-Nov-20	KCE - Belgian Health Care Knowledge Centre Google: kce.fgov.be	42	20
26-Nov-20	Zorginstituut Nederland - The Netherlands Google: zorginstituutnederland.nl	6	3
26-Nov-20	The Netherlands Organisation for Health Research and Development - The Netherlands Google: zonmw.nl	0	0
26-Nov-20	Austrian Institute for Health Technology Assessment – Austria HTA Austria Google: aihta.at	14	0

26-Nov-20	SEC - Department of HTA at the State Expert Centre of the Ministry of Health, Ukraine Google: moz.gov.ua	0	0
26-Nov-20	Agency for Health Technology Assessment and Tariff System - Poland Google: aotm.gov.pl	2	0
26-Nov-20	Canadian Agency for Drugs and Technologies in Health - Canada https://www.cadth.ca/ (using directly the search engine from the website)	34	25
26-Nov-20	Institute of Health Economics – Canada Google: ihe.ca	18	5
26-Nov-20	Institut national d'excellence en santé et services sociaux - Canada Google: inesss.qc.ca	0	0
26-Nov-20	Agency for Healthcare Research and Quality – USA Google: ahrq.gov	51	10
26-Nov-20	National Committee for Technology Incorporation – Brazil Google: conitec.gov.br	1	0
12-Nov-20	Instituto de Evaluación de Tecnologías en Salud IETS - Colombia Google: iets.org.co	0	0
26-Nov-20	Institute of Health Technology Assessment and Research, Peru Google: essalud.gob.pe/ietsi	0	0
26-Nov-20	Ministry of Public Health – Uruguay Google: gub.uy/ministerio-salud-publica/	0	0
27-Nov-20	Institute for Clinical Effectiveness and Health Policy IECS - Argentina Google: iecs.org.ar	1	0
27-Nov-20	Health Technology Reference Group, Australia and New Zealand Google: coaghealthcouncil.gov.au	0	0
27-Nov-20	Adelaide Health Technology Assessment – Australia Google:adelaide.edu.au/ahta/	0	0
27-Nov-20	Australian Safety and Efficacy Register of New Interventional Procedures -Surgical - Australia Google: surgeons.org/racs/research-and-audit/asernip-s	0	0
27-Nov-20	Health Technology Assessment Section, Ministry of Health Malaysia, Malaysia Google:moh.gov.my	2	0
27-Nov-20	Agency for Care Effectiveness – Singapore Google: ace-hta.gov.sg	0	0
27-Nov-20	National Evidence-based healthcare Collaborating Agency, Korea Google: neca.re.kr	0	0
27-Nov-20	Center for Drug Evaluation, Taiwan - Republic of China Googlecde.org.tw	0	0
27-Nov-20	Center for Healthcare Quality Assessment and Control - Russia Google: rosmedex.ru	1	0
27-Nov-20	Ministry of Public Health of the Republic of Kazakhstan, Republican Centre for Health Development, Kazakhstan Google:.rcrz.kz	2	0
27-Nov-20	Israeli Center for Technology Assessment in Health Care – Israel Google:gertnerinst.org.il	1	1
27-Nov-20	National Authority for Assessment and Accreditation in Healthcare, Tunisia Google:ineas.tn	0	0
27-Nov-20	Charlotte Maxeke Research Cluster - South Africa Google:cmerc.org	0	0
27-Nov-20	European medicines agency – EMA Google:ema.europa.eu	30	12
Total		447	127

Table 3.A.3. Grey literature: Networks, societies and international bodies

Date searched	Organisation	Total	Studies retrieved
30-Nov-20	World Health Organization-WHO Google: who.int	134	67
30-Nov-20	Professional Society for Health Economics and Outcomes Research- ISPOR www.ispor.org	Value outcomes spotlight: 46 Value in Health: 74 Value in health regional issues: 2 Website: 96	12
15-Nov-20	European Network for Health Technology Assessment - EUnetHTA Google: eunethta.eu	1	0
15-Nov-20	EuroScan International Network Google: euroscan.org	0	0
01-Dec-20	Health Technology Assessment international - HTAi Google: htai.org	32	5
01-Dec-20	International Symposium on the Analytic Hierarchy Process Google: isahp.org	127	16
01-Dec-20	International Network of Agencies for Health Technology Assessment - INHATA Google: inahta.org	5	1
Total		517	101

Table 3.A.4. Grey literature: Trip Medical Database

Date searched	Search terms	Total	Evidence-bases synopses	Guidelines	Studies retrieved
27-Nov-20	"multi-criteria decision" OR "multicriteria decision" OR "multiple criteria decision" OR "multicriteria analysis"	263	19	3	19
27-Nov-20	"Benefit risk assessment" OR "Multi-objective optimization" OR "Analytical hierarchy process" OR "Multi-attribute utility" OR "Multiattribute utility theory" OR "Weighted product method" OR "Analytic network process" OR "Goal programming" OR "Elimination and choice expressing reality"	858	33	29	42
27-Nov-20	"Preference ranking organization method of enrichment evaluation" OR "'weighted product model" OR "measuring attractiveness by a categorical based evaluation technique" OR "technique for order preference by similarity to ideal solution"	13	1	0	0
Total		1134	53	32	61
			85		

Appendix 3.B. Definition of the decision-contexts

Decision context	Definition & rationale
Priority-setting	MCDA studies that have informed resource allocation decisions or investment decisions in which due to resource constraints, priorities need to be set. This decision-context or type of decision was also identified by Marsh et al (2014) (39). As a step further, the present review distinguishes between three types of priority-setting or investment decisions, based on the type of alternative from which priorities need to be set: (a) Prioritization of healthcare interventions for coverage, reimbursement or funding. Eg., Rank healthcare services to define the contents of a national health benefits package. (b) Prioritization of diseases or health conditions for coverage, reimbursement or funding. Eg., Locate the most important health problems for reimbursement decisions. (c) Prioritization of patients to access health care. Eg., To help prioritize hospital admission of patients affected by Covid-19.
Clinical decision making	MCDA studies that aim to inform medical treatment or selection of screening alternatives for screening or disease diagnosis. These studies tend to inform physicians when treating patients, patients in the context of shared medical decision-making, or health authorities when developing medical guidelines. This type of decision was also identified by Marsh et al (2014) (39). As a step further, the present review distinguishes between two types of clinical decision-making:

	(a) Treatment or prescription decisions. Eg., To help clinicians decide the best approach for the treatment of femoroacetabular impingement. (b) Screening decisions. Eg. To help clinicians and patients to make decisions on colon cancer screening.
Regulatory decisions	MCDA studies that have conducted benefit-risk assessments (BRA) aiming at issuing recommendations for health authorities to make regulatory decisions, such as licensing decisions or market authorization decisions. Eg., To decide if an antidepressant pharmaceutical can be given market authorization. This decision-context or type of decision was also identified by Marsh et al (2014) (39).
Planning and R&D	MCDA studies that aim to guide research and development (R&D) decisions (a decision type also identified by Marsh et al (2014) (39)), and studies that support planning decisions across the production, sourcing, and organisation of healthcare (not classified in previous categories). More precisely, this review distinguishes between the following types of decisions: (a) Capital investment decisions. Eg,. To compare various bioprinters, so as to recommend the most suitable one to a decision maker. (b) Location / Reallocation decisions. Eg. To help health authorities to decide where to build a new hospital. (c) Guide pharmaceutical developments. Eg. To support decision making during drug development by assessing the value of compounds in clinical development for the treatment of nonalcoholic steatohepatitis (NASH). (d) Research purposes. Eg., To identify the priority questions related to environment, agriculture and infectious disease of poverty to guide research over the next decade. (e) Regulation of addictive substances. Eg. To review drug harms (including alcohol and tobacco) and provide better guidance to policy makers in health, policing, and social care. (f) Decisions on hospital wastes management decisions. Eg. To evaluate five different technologies used for infectious medical waste treatment and help health care facilities to select the optimum one. (g) Other investing planning decisions. Eg. To guide the selection of healthcare providers or to choose best length of stay policy alternatives.

Appendix 3.C. Studies included in the review

#	Study	Source	Year	Location	Decision context			
					Priority-setting	Clinical decision making	Regulatory decisions	Planning and R&D
1	Jehu-Appiah, et al [1]	Value in Health	2008	Ghana	X			
2	Miot, et al [2]	Cost Eff Resour Alloc	2012	South Africa	X			
3	Marsh, et al [3]	Journal of Public Health	2013	United Kingdom	X			
4	Nobre, et al[4]	Statistics in Medicine	1999	Brazil	X			
5	Le Gales, and Moatti [5]	International Journal of Technology Assessment In Health Care	1990	France	X			
6	Dolan and Bordley[6]	Med Decis Making	1994	United States		X		
7	Wenstøp and Magnus [7]	Health Policy	2001	Norway	X			
8	Hummel, et al[8]	J Rehabil Res Dev	2005	The Netherlands		X		
9	Dolan et al [9] Dolan and Frisina[10]	Health Expectations Medical Decision Making	2005 2002	United States		X		
10	Van Wijk, et al [11] – MCDA 1	American Journal of Cardiovascular Drugs	2006	The Netherlands		X		
11	Van Wijk, et al [11] - MCDA 2	American Journal of Cardiovascular Drugs	2006	The Netherlands		X		
12			2006			X		

	Singh, et al [12]	BMC Medical Informatics and Decision Making		United States				
13	Baltussen and Niessen[13] Baltussen et al [14]	Cost Effectiveness and Resource Allocation Health Economics	2006 2006	Ghana	X			
14	Baltussen, et al[15]	Health Policy and Planning	2007	Nepal	X			
15	Wilson, et al [16]	Journal of Health Services Research & Policy	2007	United Kingdom (England)	X			
16	van Til, et al [17]	Archives of Physical Medicine and Rehabilitation	2008	The Netherlands		X		
17	Dalalah and Magableh [18]	Telemedicine and e-Health	2008	Jordan		X		
18	González-Zapata, et al [19]	Public Health Nutrition	2009	Cyprus, Finland, France, Greece, Hungary, Italy, Poland, Spain and the United Kingdom	X			
19	Felli, et al [20]	Medical Decision Making	2009	United States			X	
20	Kroese, et al [21]	Public Health Genomics	2010	United Kingdom	X			
21	Goetghebeur, et al [22]	Cost Effectiveness and Resource Allocation	2010	Canada				X
22	Baeten, et al [23]	Value in Health	2010	The Netherlands	X			

23	Tony, et al [24]	BMC Health Services Research	2011	Canada	X			
24	Diaby, et al [25]	Applied Health Economics and Health Policy	2011	Coˆ te d'Ivoire	X			
25	Cunich, et al [26]	The Patient: Patient-Centered Outcomes Research	2011	Australia		X		
26	Airoldi, et al [27] Airoldi, et al [28]	LSE Working paper series Medical Decision Making	2011 2014	United Kingdom	X			
27	Youngkong, et al [29] Youngkong, et al [30]	Health Research Policy and Systems BMC Health Services Research	2012	Thailand	X			
28	Youngkong, et al [31]	Value in Health	2012	Thailand	X			
29	van Valkenhoef, et al [32]	Journal of Clinical Epidemiology	2012	The Netherlands			X	
30	Guest, et al [33]	Journal of Neurosurgery: Spine SPI	2012	United States				X
31	Hummel, et al[34]	The Patient: Patient-Centered Outcomes Research	2012	Germany	X	X		
32	Hummel, et al [35]	Journal of Orthopaedic Research	2012	The Netherlands	X			
33	Holdsworth, et al [36] – MCDA 1	Health policy and planning	2012	Morocco	X			
34	Holdsworth, et al[36] – MCDA 2	Health policy and planning	2012	Tunisia	X			

35	Goetghebeur, et al[37]	Med Decis Making	2012	Canada	X			
36	Erjaee, et al[38]	Hong Kong Journal of Paediatrics	2012	Iran		X		
37	Defechereux, et al[39]	BMC Health Services Research	2012	Norway	X			
38	Broekhuizen, et al[40]	Value in Health	2012	The Netherlands			X	
39	Diaz-Ledezma, and Parvizi [41]	Clinical Orthopaedics and Related Research	2013	United States		X		
40	Brass, et al[42]	The Journal of Clinical Pharmacology	2013	United States			X	
41	Wagner, et al[43]	Advances in Therapy	2018	Spain		X		
42	Broekhuizen, et al [44]	Value in Health	2018	The Netherlands		X		
43	Rutten-van Molken, et al [45]	BMC Health Services Research	2018	Eight European countries: Austria, Croatia, Germany, Hungary, the Netherlands, Norway, Spain, and the United Kingdom.	X			
44	Tarimcilar and Khaksari [46]	Socio-Economic Planning Sciences	1991	United States				X
45	Sinuany-Stern, et al [47]	Locarion Science	1995	Israel				X
46	Dolan [48]	Medical Decision Making	1995	United States		X		
47			2013	Europe	X			

	Sussex, et al [49]	Value in Health						
48	Sidi and Harel [50]	Value in Health	2020	United States			X	
49	Wen, et al [51]	Value in Health	2014	Europe			X	
50	Mujoomdar, et al[52]	Canadian Agency for Drugs and Technologies in Health	2012	Canada				X
51	Kaltoft, et al [53]	Health Economics Review	2015	Australia		X		
52	Liverpool Clinical Commissioning Group[54]	UK National Health Service NHS	2017	United Kingdom				X
53	Williams, et al[55]	British Thoracic Society and the Primary Care Respiratory Society UK	2012	United Kingdom	X			
54	Agapova, et al [56]	Academic Radiology	2017	United States			X	
55	Zozaya, et al [57]	BioDrugs	2018	Spain	X			
56	Zhang, et al [58]	International Journal of Environmental Research & Public Health [Electronic Resource]	2018	China	X			
57	Ritrovato, et al [59]	Value in Health	2015	Italy	X			
58	Lvovschi, et al [60]	PLoS ONE	2020	France	X			
59	Ender Sir and Batur Sir [61]	BMC Medical Informatics & Decision Making	2019	Turkey		X		

60	Chapple, et al [62]	European Urology Focus	2020	Europe			X	
61	Inotai, et al [63]	BMC Health Services Research	2018	Indonesia	X			
62	Yu, et al [64]	Paediatric Drugs	2018	China	X			
63	Wagner, et al [65]	Advances in Therapy	2018	United States		X		
64	Tacconelli, et al [66] Tacconelli and Magrini [67] WHO [68]	The Lancet Infectious Diseases WHO WHO	2018 2017 2017	Global / WHO				X
65	Si Mehand, et al [69] WHO [70]	Emerging Infectious Diseases World Health Organization -WHO	2018 2017	Global / WHO	X			
66	Mehand, et al [71] WHO [72]	Antiviral Research World Health Organization -WHO	2018	Global / WHO	X			
67	Sabir et al [73]	Journal of the Pakistan Medical Association	2018	Pakistan	X			
68	Roldan, et al [74]	International Journal of Technology Assessment in Health Care	2018	Spain	X			
69			2018			X		

	Postmus, et al [75]	The Oncologist		United Kingdom				
70	Mendoza-Sanchez, et al [76]	PLoS ONE [Electronic Resource]	2018	Colombia			X	
71	Jimenez, et al [77]	Orphanet Journal Of Rare Diseases	2018	Spain	X	X		
72	Fdez-Olivares, et al [78]	Artificial Intelligence in Medicine	2019	Spain		X		
73	Bots and Hulshof [79] – Case 1	Journal of Multi-Criteria Decision Analysis	2000	The Netherlands	X			
74	Bots and Hulshof [79] – Case 2	Journal of Multi-Criteria Decision Analysis	2000	The Netherlands				X
75	Eiring, et al[80]	Systematic Reviews	2018	Norway		X		
76	Nutt, et al [81]	The Lancet	2010	United Kingdom				X
77	Rossi, et al[82]	Ergonomics	2019	Italy	X			
78	Samanlioglu and Kaya[83]	Journal of Healthcare Engineering	2020	Global	X			
79	Howard, et al[84]	Australian Health Review	2019	Australia	X			
80	De Nardo, et al[85]	International Journal of Infectious Diseases	2020	Italy	X			
81	Ozsahin, et al[86]	Archives of Psychiatry and Psychotherapy	2020	Turkey		X		
82	Cleemput, et [87]	KCE - Belgian Health Care Knowledge Centre	2014	Belgium	X			

	Cleemput, et al [88] Cleemput, et al [89]	KCE - Belgian Health Care Knowledge Centre Health Policy	2016 2018					
83	Bouwknegt, et al[90]	Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin	2018	Europe	X			
84	Rahman, et al [91]	Modeling Earth Systems and Environment	2020	Bangladesh	X			
85	Ezeife, et al[92]	Cancer	2020	Canada	X			
86	Sussex, et al[93]	OHE Research Paper	2013	Europe	X			
87	Otten, et al[94]	Microbial Risk Analysis	2020	Canada	X			
88	Richman, et al[95]	The Journal of urology	2005	United States	X			
89	Samanlioglu [96]	Journal of Healthcare Engineering	2019	Turkey	X			
90	European Centre for Disease Prevention and Control[97]	European Centre for Disease Prevention and Control - Technical Guidance	2017	Europe	X			
91	Ozkurt et al [98]	International Journal of Environmental Research & Public Health [Electronic Resource]	2019	Syria				X

92	Improta, et al [99]	BMC Medical Research Methodology	2019	Italy	X			
93	Tang [100]	Tang Health Research Policy and Systems	2013	China				X
94	Hummel, et al[101]	Applied Health Economics & Health Policy	2013	The Netherlands	X			
95	Stevens, et al [102]	Spatial and Spatio-Temporal Epidemiology	2013	Countries in Southern, East ern and Southeastern Asia (China, Romania, Bangladesh, Thailand, Vietnam, etc)	X			
96	Ramli, et al [103]	Therapeutics & Clinical Risk Management	2013	Malaysia	X			
97	Cox, R.; Sanchez, J.; Revie, C. W. [104]	PLoS ONE	2013	Canada	X			
98	Bentsi-Enchill and Clemens [105] SAGE [106]	World Health Organization World Health Organization	2013	Global / WHO				X
99	Del Rio Vilas, et al. [107]	Pathogens and Global Health	2013	The region of the Americas	X			
100	Ghandour, et al [108]	International Journal of Public Health	2015	Palestine, Syria, Tunisia and Turkey	X			
101	Kolasa, et al [109]	Orphanet Journal Of Rare Diseases	2016	Poland	X			
102	Angelis, et al [110]	Medical Decision Making	2020	United Kingdom (England), France, and Germany				X

103	Angelis, et al [111]	Social Science & Medicine	2020	Sweden, Andalusia/Spain, Poland and Belgium	X			
104	Babashov, et al [112]	Medical Decision Making	2020	Canada	X			
105	Costa, et al [113]	Revista de Saude Publica	2020	Brazil				X
106	Castro-Jaramillo, et al [114]	International Journal of Technology Assessment in Health Care	2016	Colombia	X			
107	Sullivan and Hansen [115]	Value in Health	2017	New Zealand	X			
108	Hongoh, et al [116]	PLoS ONE	2017	Canada	X			
109	Stromme, et al [117]	Developing World Bioethics	2014	Countries where where Neglected Tropical Diseases (NTDs) are endemic (WHO)	X			
110	Marcelon, et al [118] EMA Committee for Medicinal Products for Human Use [119]	Expert Review of Vaccines European medicines agency – EMA	2016 2014	Europe			X	
111	Reddy, et al [120]	Public Health	2014	United Kingdom				X
112	Tervonen, et al [121]	Medical Decision Making	2015	The Netherlands			X	
113	Arslan [122]	Ankara Universitesi Eczacilik Fakultesi Dergisi	2020	Turkey				X
114			2016	Iran	X			

	Mohamadi, et al [123]	Shiraz E Medical Journal						
115	Paolucci, et al [124]	Health Policy & Planning	2014	China	X			
116	Guarga, et al [125]	Orphanet Journal Of Rare Diseases	2019	Spain	X			
117	Araz [126]	Journal of Systems Science and Systems Engineering	2013	United States	X			
118	Roselli, et al [127]	Computational & Mathematical Methods in Medicine	2020	Brazil		X		
119	Lin, et al [128]	Journal of Multi-Criteria Decision Analysis	2013	United States	X			
120	Rahimi, et al [129]	Shiraz E-Medical Journal	2017	Iran				X
121	WHO [130]	World Health Organization (WHO)	2013	Global / WHO				X
122	FAO & WHO [131]	Food and Agriculture Organization (FAO) of the United Nations World Health Organization (WHO)	2014	Global WHO/FAO	X			
123	De Laet, et al [132]	KCE - Belgian Health Care Knowledge Centre	2016	Belgium	X			
124	Rello, et al [133]	Intensive Care Med	2010	Europe	X			
125	Kwon, et al [134]	Expert Review of Pharmacoeconomics & Outcomes Research	2017	South Korea (Republic of Korea)	X			

126	Sayan, et al [135]	Computational & Mathematical Methods in Medicine	2020	Turkey		X		
127	Ruggeri, et al [136]	International Journal of Environmental Research & Public Health [Electronic Resource]	2020	Italy	X			
128	Carnero and Gomez [137]	BMC Medical Informatics & Decision Making	2016	Spain				X
129	Marsh, et al [138]	Value in Health	2017	United States			X	
130	Maseleno, et al [139]	Investigacion Clinica (Venezuela)	2017	Malaysia	X			
131	Dehe and Bamford [140] Dehe, et al [141]	Expert Systems with Applications University of Huddersfield Repository	2015 2011	United Kingdom (England)				X
132	Schug, et al [142]	Journal of pain research	2020	Europe			X	
133	Moore, et al [143]	Journal of Pharmacy & Pharmacology	2017	United Kingdom			X	
134	Carter, et al [144]	Medical Decision Making	1999	United States		X		
135	Angelis, et al [145]	BMC Medical Informatics & Decision Making	2017	United Kingdom (England)	X	X		
136	Byun, et al [146]	Therapeutics & Clinical Risk Management	2016	South Korea			X	
137			2017	Turkey		X		

	Hancerliogullari, et al[147] – MCDA 1	BMC Medical Informatics & Decision Making						
138	Hancerliogullari, et al [147] - MCDA 2	BMC Medical Informatics & Decision Making	2017	Turkey		X		
139	Hsu, et al [148] – MCDA 1	International Journal of Clinical Practice	2015	Taiwan			X	
140	Hsu, et al [148] – MCDA 2	PLoS ONE [Electronic Resource]	2015	Taiwan			X	
141	Gopi, et al[149]	Journal of Pharmaceutical Sciences and Research	2015	India				X
142	Garau, et al[150]	PharmacoEconomics Open	2018	Italy	X			
143	Rello, et al [151]	European Journal of Clinical Microbiology & Infectious Diseases	2019	Global	X			
144	Dias nad Santos [152]	Value in Health Regional Issues	2020	Brazil		X		
145	Özcan, et al[153]	Journal of Biochemical & Molecular Toxicology	2020	Turkey				X
146	Osborn, et al[154]	BMC Infectious Diseases	2020	Global	X			
147	Tervonen, et al [155]	Value in health	2017	European Union		X		
148	Rajabi, et al [156]	Geospatial Health	2013	Iran	X			
149	Peacock, et al [157]	Social science & medicine	2007	Australia	X			

150	Tsiachristas, et al [158]	International Journal of Technology Assessment in Health Care	2013	The Netherlands	X			
151	Bettinger, et al[159]	Annals of pharmacotherapy	2007	United States	X	X		
152	Vidal, et al [160]	Expert Systems with Applications	2010	France	X			
153	Chung, et al [161]	American Journal of Health-System Pharmacy	2010	Korea		X		
154	Ivlev, et al [162]	European Journal of Operational Research	2015	Czech Republic		X		
155	Buttigieg, et al [163]	Journal of Health Organization & Management	2018	Malta				X
156	Oketch-Rabah, et al [164]	Journal of Dietary Supplements	2020	United States				X
157	Mokrini, et al [165]	Health Services Management Research	2020	Morocco				X
158	Cruden, et al [166]	Prevention Science	2020	United States	X			
159	Chen, et al [167]	Health and Technology.	2020	Taiwan				X
160	Rossetti and Selandari [168]	Computers & Industrial Engineering	2001	United States	X			
161	Liberatore, et al [169]	Computers & Operations Research	2003	United States		x		
162	Cho and Kim [170]	The International journal of health planning and management	2003	Korea	X			

163	Chang, et al [171]	Expert Systems with Applications	2004	Taiwan	X			
164	Kip, et al [172]	BMC Family Practice	2019	Netherlands		X		
165	Li, et al [173]	Statistics in Medicine	2019	United States			X	
166	Vermersch, et al [174]	Clinical Therapeutics	2019	Europe			X	
167	Mortimer, et al[175]	Health Expectations	2019	Australia	X			
168	Eghbali-Zarch, et al [176]	Artificial Intelligence in Medicine	2018	Iran		X		
169	Hsu, et al [177]	PLoS ONE [Electronic Resource]	2019	Taiwan	X			
170	Holmes, et al [178]	Clinical Pharmacology & Therapeutics	2019	United Kingdom			X	
171	Choi, et al [179]	Medical Decision Making	2019	United States		X		
172	de Paula, et al[180]	Molecules	2018	Brazil				X
173	Lin and Wang[181]	Journal of Healthcare Engineering	2019	China				X
174	Bollaerts, et al[182]	Vaccine	2019	Denmark, Spain, Italy and the UK.			X	
175	Badia, et al [183]	International Journal of Technology Assessment in Health Care	2019	Europe		X		
176	Akcan and Guldes [184]	Journal of Healthcare Engineering	2019	Turkey				X

177	Alsalem, et al [185]	Journal of Medical Systems	2019	Malaysia		X		
178	Nutt, et al [186]	European Addiction Research	2014	Global				X
179	Wahlster, et al [187] Wahlster, et al [188]	Health Research Policy & Systems WHO	2015 2013	Germany		X		
180	Diaz-Ledezma, et al [189]	Clinical Orthopaedics & Related Research	2014	United States	X	X		
181	Aenishaenslin, et al [190]	PLoS ONE [Electronic Resource]	2015	Switzerland	X			
182	van Amsterdam, J, et al [191]	Journal of Psychopharmacology	2015	Europe				X
183	Almahdi, et al [192]	Journal of Medical Systems	2019	Malaysia		X		
184	Abdel-Basset, et al [193]	Journal of Medical Systems	2019	Egypt		X		
185	Martelli, et al [194]	Journal of Biomedical Informatics	2016	France	X			
186	Hongoh, et al [195]	International Journal of Environmental Research & Public Health [Electronic Resource]	2016	Canada and Burkina Faso	X			
187	Suner, et al [196]	BMC Medical Informatics & Decision Making	2017	Turkey	X			
188	Levitan, et al [197]	Therapeutic Innovation & Regulatory Science	2014	United States			X	
189			2017	India				X

	Kyavars and Subramanian [198]	Indian Journal of Pharmaceutical Education and Research						
190	Hongoh, et al [199]	PLoS ONE	2016	Canada	X			
191	Al-Badriyeh, et al [200]	Clinical Therapeutics	2016	Qatar	X			
192	Mobinizadeh, et al [201]	Daru: Journal of Faculty of Pharmacy, Tehran University of Medical Sciences	2016	Iran	X			
193	Claudio, et al [202]	IIE Transactions on Healthcare Systems Engineering	2014	United States	X			
194	Wagner, et al [203]	BMC Cancer	2017	France, Italy and Spain.	X			
195	Iskrov, et al [204]	Frontiers in Public Health	2016	Bulgaria	X			
196	Yang, et al [205]	International Journal of Environmental Research & Public Health [Electronic Resource]	2020	China	X			
197	de Greef-van der Sandt, et al [206]	Clinical Pharmacology & Therapeutics	2016	The Netherlands			X	
198	Kalhor, et al [207]	Journal of Biology and Today's World	2016	Iran				X
199	Nixon, et al. [208]	Biometrical Journal	2016	Europe			X	
200	Voudrias [209]		2016	Greece				X

		Journal of the Air & Waste Management Association						
201	Goetghebeur [210]	Therapeutic Innovation & Regulatory Science	2019	Europe			X	
202	Alimohammadzadeh, et al [211]	Iranian Journal of Radiology	2015	Iran				X
203	Kunasekaran and Krishnamoorthy [212]	Journal of Advanced Pharmaceutical Technology & Research	2014	India				X
204	Maro, et al [213]	PLoS ONE [Electronic Resource]	2014	United States				X
205	Mahfoud, et al [214]	American Journal of Applied Sciences	2016	Morocco	X			
206	Carnero [215]	Shock and Vibration	2016	Spain				X
207	Bowers, et al [216]	BMC Health Services Research	2018	United Kingdom	X			
208	Qian, et al [217]	International Journal of Environmental Research & Public Health [Electronic Resource]	2016	China				X
209	Ortiz, et al [218]	BMC Medical Informatics & Decision Making	2015	Colombia				X
210	Ozturk and Tozan [219]	Journal of Healthcare Engineering	2015	Turkey		X		
211	Panattoni, et al [220]	The Patient: Patient-Centered Outcomes Research	2020	United States		X		

212	Ozturk, et al[221]	International Journal of Environmental Research & Public Health [Electronic Resource]	2020	Turkey	X			
213	Ozsahin [222]	Healthcare Informatics Research	2020	Turkey		X		
214	Parsia and Sorooshian [223]	Journal of Infection and Public Health	2020	Iran				X
215	Shojaei, et al [224]	Cost Effectiveness & Resource Allocation	2020	Iran				X
216	Gómez and Carnero [225]	IFAC-PapersOnLine	2016	Spain				X
217	Parsia and Sorooshian [226]	International Journal of Environmental Research & Public Health	2020	Iran				X
218	Ren, et al [227]	Computers & Industrial Engineering	2020	China		X		
219	Kalid, et al [228]	Journal of Medical Systems	2018	Malaysia	X			

References from the updated systematic literature review:

1. Jehu-Appiah C, Baltussen R, Acquah C, et al. Balancing Equity and Efficiency in Health Priorities in Ghana: The Use of Multicriteria Decision Analysis. Value in Health. 2008;11(7):1081-1087.

2. Miot J, Wagner M, Khoury H, Rindress D, Goetghebeur MM. Field testing of a multicriteria decision analysis (MCDA) framework for coverage of a screening test for cervical cancer in South Africa. Cost effectiveness and resource allocation : 2012;10(1):2-2.

3. Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. J Public Health (Oxf). Sep 2013;35(3):460-6.

4. Nobre FF, Trotta LTF, Gomes LFAM. Multi-criteria decision making – an approach to setting priorities in health care. Statistics in Medicine. 1999;18(23):3345-3354.

5. Le Gales C, Moatti JP. Searching for consensus through multi-criteria decision analysis. Assessment of screening strategies for hemoglobinopathies in southeastern France. Int J Technol Assess Health Care. 1990;6(3):430-49.

6. Dolan JG, Bordley DR. Isoniazid Prophylaxis:The Importance of Individual Values. Medical Decision Making. 1994;14(1):1-8.

7. Wenstøp F, Magnus P. Value focused rationality in AIDS policy. Health Policy. Jul 2001;57(1):57-72.

8. Hummel JM, Snoek GJ, van Til JA, van Rossum W, Ijzerman MJ. A multicriteria decision analysis of augmentative treatment of upper limbs in persons with tetraplegia. J Rehabil Res Dev. Sep-Oct 2005;42(5):635-44.

9. Dolan JG. Patient priorities in colorectal cancer screening decisions. Health Expectations. 2005;8(4):334-344.

10. Dolan JG, Frisina S. Randomized Controlled Trial of a Patient Decision Aid for Colorectal Cancer Screening. Medical Decision Making. 2002;22(2):125-139.

11. Van Wijk BLG, Klungel OH, Heerdink ER, de Boer A. A Comparison of Two Multiple-Characteristic Decision-Making Models for the Comparison of Antihypertensive Drug Classes. American Journal of Cardiovascular Drugs. 2006;6(4):251-258.

12. Singh S, Dolan JG, Centor RM. Optimal management of adults with pharyngitis – a multi-criteria decision analysis. BMC Medical Informatics and Decision Making. 2006;6(1):14.

13. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Effectiveness and Resource Allocation. 2006;4(1):14.

14. Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. Health Econ. Jul 2006;15(7):689-96.

15. Baltussen R, ten Asbroek AHA, Koolman X, Shrestha N, Bhattarai P, Niessen LW. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? Health Policy and Planning. 2007;22(3):178-185. d

16. Wilson ES, J.; Macleod, C.; Fordham, R. Prioritizing health technologies in a Primary Care Trust. Journal of Health Services Research & Policy. 2007;12(2):80-85.

17. van Til JA, Renzenbrink GJ, Dolan JG, Ijzerman MJ. The Use of the Analytic Hierarchy Process to Aid Decision Making in Acquired Equinovarus Deformity. Archives of Physical Medicine and Rehabilitation. 2008;89(3):457-462.

18. Dalalah DM, S. A Remote Fuzzy Multicriteria Diagnosis of Sore Throat. Telemedicine and e-Health. 2008;14(7):656-665.

19. González-Zapata LI, Alvarez-Dardet C, Ortiz-Moncada R, et al. Policy options for obesity in Europe: a comparison of public health specialists with other stakeholders. Public Health Nutrition. 2009;12(7):896-908.

20. Felli JC, Noel RA, Cavazzoni PA. A Multiattribute Model for Evaluating the Benefit-Risk Profiles of Treatment Alternatives. Medical Decision Making. 2009;29(1):104-115.

21. Kroese M, Burton H, Whittaker J, Lakshman R, Alberg C. A Framework for the Prioritization of Investment in the Provision of Genetic Tests. Public Health Genomics. 2010;13(7-8):538-543.

22. Goetghebeur MM, Wagner M, Khoury H, Rindress D, Grégoire J-P, Deal C. Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM decision-making framework to growth hormone for Turner syndrome patients. Cost effectiveness and resource allocation : C/E. 2010;8:4-4.

23. Baeten SA, Baltussen RMPM, Uyl-de Groot CA, Bridges J, Niessen LW. Incorporating Equity–Efficiency Interactions in Cost-Effectiveness Analysis—Three Approaches Applied to Breast Cancer Control. Value in Health. 2010;13(5):573-579.

24. Tony M, Wagner M, Khoury H, et al. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. BMC Health Services Research. 2011;11(1):329.

25. Diaby V, Lachaine J. An application of a proposed framework for formulary listing in low-income countries. Applied Health Economics and Health Policy. 2011;9(6):389-402.

26. Cunich M, Salkeld G, Dowie J, et al. Integrating Evidence and Individual Preferences Using a Web-Based Multi-Criteria Decision Analytic Tool. The Patient: Patient-Centered Outcomes Research. 2011;4(3):153-162.

27. Airoldi M, Morton A, Smith J, Bevan G. Healthcare prioritisation at the local level: a socio-technical approach. 2011:

28. Airoldi M, Morton A, Smith JA, Bevan G. STAR--people-powered prioritization: a 21st-century solution to allocation headaches. Med Decis Making. Nov 2014;34(8):965-75.

29. Youngkong S, Teerawattananon Y, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. Health Research Policy and Systems. 2012;10(1):6.

30. Youngkong S, Baltussen R, Tantivess S, Koolman X, Teerawattananon Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. BMC Health Services Research. 2010;10(1):197.

31. Youngkong S, Baltussen R, Tantivess S, Mohara A, Teerawattananon Y. Multicriteria Decision Analysis for Including Health Interventions in the Universal Health Coverage Benefit Package in Thailand. Value in Health. 2012;15(6):961-970.

32. van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multicriteria benefit–risk assessment using network meta-analysis. Journal of Clinical Epidemiology. 2012;65(4):394-403.

33. Guest J, Harrop JS, Aarabi B, et al. Optimization of the decision-making process for the selection of therapeutics to undergo clinical testing for spinal cord injury in the North American Clinical Trials Network. J Neurosurg Spine. 2012;17(1 Suppl):94-101.

34. Hummel MJM, Volz F, van Manen JG, et al. Using the Analytic Hierarchy Process to Elicit Patient Preferences. The Patient: Patient-Centered Outcomes Research. 2012;5(4):225-237. 5

35. Hummel JM, Boomkamp ISM, Steuten LMG, Verkerke BGJ, IJzerman MJ. Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis. Journal of Orthopaedic Research. 2012;30(9):1453-1458.

36. Holdsworth M, El Ati J, Bour A, et al. Developing national obesity policy in middle-income countries: a case study from North Africa. Health policy and planning. 2012;28(8):858-870.

37. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging Health Technology Assessment (HTA) and Efficient Health Care Decision Making with Multicriteria Decision Analysis (MCDA):Applying the EVIDEM Framework to Medicines Appraisal. Medical Decision Making. 2012;32(2):376-388.

38. Erjaee A, Bagherpour M, Razeghi S, Dehghani SM, Imanieh MH, Haghighat M. A Multi-Criteria Decision Making Model for Treatment of Helicobacter pylori Infection in Children. Hong Kong Journal of Paediatrics. 2012;17:237-242.

39. Defechereux T, Paolucci F, Mirelman A, et al. Health care priority setting in Norway a multicriteria decision analysis. BMC Health Services Research. 2012;12(1):39.

40. Broekhuizen H, Ijzerman MJ, Groothuis-Oudshoorn KG, Hauber AB. PRM79 Integrating Patient Preferences and Clinical Trial Data in a Bayesian Model for Quantitative Risk-Benefit Assessment. Value in Health. 2012;15(7):A474.

41. Diaz-Ledezma C, Parvizi J. Surgical Approaches for Cam Femoroacetabular Impingement: The Use of Multicriteria Decision Analysis. Clinical Orthopaedics and Related Research®. 2013;471(8):2509-2516.

42. Brass EP, Lofstedt R, Renn O. A Decision-Analysis Tool for Benefit–Risk Assessment of Nonprescription Drugs. The Journal of Clinical Pharmacology. 2013;53(5):475-482.

43. Wagner M, Samaha D, Cuervo J, et al. Applying Reflective Multicriteria Decision Analysis (MCDA) to Patient-Clinician Shared Decision-Making on the Management of Gastroenteropancreatic Neuroendocrine Tumors (GEP-NET) in the Spanish Context. Advances in Therapy. Aug 2018;35(8):1215-1231.

44. Broekhuizen H, Groothuis-Oudshoorn CGM, Vliegenthart R, Groen HJM, MJ IJ. Assessing Lung Cancer Screening Programs under Uncertainty in a Heterogeneous Population. Comparative Study. Value in Health. 2018;21(11):1269-1277.

45. Rutten-van Mölken M, Leijten F, Hoedemakers M, et al. Strengthening the evidence-base of integrated care for people with multi-morbidity in Europe using Multi-Criteria Decision Analysis (MCDA). BMC Health Services Research. 2018/07/24 2018;18(1):576.

46. Tarimcilar MM, Khaksari SZ. Capital budgeting in hospital management using the analytic hierarchy process. Socio-Economic Planning Sciences. 1991;25(1):27-34.

47. Sinuany-Stern Z, Mehrez A, Tal A-G, Shemuel B. The location of a hospital in a rural region: The case of the Negev. Location Science. 1995;3(4):255-266.

48. Dolan JG. Are Patients Capable of Using the Analytic Hierarchy Process and Willing to Use It to Help Make Clinical Decisions? Medical Decision Making. 1995;15(1):76-80.

49. Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. A Pilot Study of Multicriteria Decision Analysis for Valuing Orphan Medicines. Value in Health. 2013/12/01/ 2013;16(8):1163-1169.

50. Sidi Y, Harel O. Comprehensive Benefit–Risk Assessment of Noninferior Treatments Using Multicriteria Decision Analysis. Value in Health. 2020;23(12):1622-1629.

51. Wen S, Zhang L, Yang B. Two approaches to incorporate clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. Comparative Study

Research Support, Non-U.S. Gov't. Value in Health. Jul 2014;17(5):619-28.

52. Mujoomdar M, Russell E, Dionne F, et al. Optimizing Health System Use of Medical Isotopes and Other Imaging Modalities [Internet]. 2012. Ottawa: Canadian Agency for Drugs and Technologies in Health.

53. Kaltoft MK, Turner R, Cunich M, Salkeld G, Nielsen JB, Dowie J. Addressing preference heterogeneity in public health policy by combining Cluster Analysis and Multi-Criteria Decision Analysis: Proof of Method. Health Economics Review. 2015;5(1):10.

54. Liverpool CCG. Healthy Liverpool Programme: Review of Services Provided by Liverpool Women's NHS Foundation Trust [Internet]. 2017. https://www.liverpoolccg.nhs.uk/media/2862/pcbc_-_final_draft_-_11_january_2017.pdf

55. Williams W, Baxter N, Holmes S, Restrick L, Scullion S, Ward M. IMPRESS Guide to the relative value of interventions for people with COPD: A population-based approach to improving outcomes for people with chronic obstructive pulmonary disease based on the cost of delivering those outcomes. 2012.

56. Agapova M, Bresnahan BW, Linnau KF, et al. Using the Analytic Hierarchy Process for Prioritizing Imaging Tests in Diagnosis of Suspected Appendicitis. Academic Radiology. 2017;24(5):530-537.

57. Zozaya N, Martinez-Galdeano L, Alcala B, et al. Determining the Value of Two Biologic Drugs for Chronic Inflammatory Skin Diseases: Results of a Multi-Criteria Decision Analysis. Biodrugs. 2018;32(3):281-291.

58. Zhang C, Wu X, Wu D, Liao H, Luo L, Herrera-Viedma E. An Intuitionistic Multiplicative ORESTE Method for Patients' Prioritization of Hospitalization. Research Support, Non-U.S. Gov't. International Journal of Environmental Research & Public Health [Electronic Resource]. 2018;15(4):17.

59. Ritrovato M, Faggiano FC, Tedesco G, Derrico P. Decision-Oriented Health Technology Assessment: One Step Forward in Supporting the Decision-Making Process in Hospitals. Value in Health. 2015;18(4):505-511.

60. Lvovschi VE, Maignan M, Tazarourte K, et al. Multiple criteria decision analysis approach to consider therapeutic innovations in the emergency department: The methoxyflurane organizational impact in acute trauma pain. PLOS ONE. 2020;15(4):e0231571.

61. Sir E, Batur Sir GD. Evaluating treatment modalities in chronic pain treatment by the multi-criteria decision making procedure. BMC Medical Informatics and Decision Making. 2019;19(1):191.

62. Chapple CR, Mironska E, Wagg A, et al. Multicriteria Decision Analysis Applied to the Clinical Use of Pharmacotherapy for Overactive Bladder Symptom Complex. European Urology Focus. 2020;6(3):522-530.

63. Inotai A, Brixner D, Maniadakis N, et al. Development of multi-criteria decision analysis (MCDA) framework for off-patent pharmaceuticals – an application on improving tender decision making in Indonesia. BMC Health Services Research. 2018;18(1):1003.

64. Yu Y, Jia L, Meng Y, et al. Method Development for Clinical Comprehensive Evaluation of Pediatric Drugs Based on Multi-Criteria Decision Analysis: Application to Inhaled Corticosteroids for Children with Asthma. Pediatric Drugs. 2018;20(2):195-204.

65. Wagner M, Samaha D, Khoury H, et al. Development of a Framework Based on Reflective MCDA to Support Patient–Clinician Shared Decision-Making: The Case of the Management of

Gastroenteropancreatic Neuroendocrine Tumors (GEP-NET) in the United States. Advances in Therapy. 2018;35(1):81-99.

66. Tacconelli E, Carrara E, Savoldi A, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. The Lancet Infectious Diseases. 2018;18(3):318-327.

67. Tacconelli E, Magrini N. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. 2017. https://www.aidsdatahub.org/sites/default/files/resource/who-global-priority-list-antibiotic-resistant-bacteria.pdf

68. World Health Organization. Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis. 2017. 9789240026438 (electronic version) 2017. https://apps.who.int/iris/handle/10665/311820

69. Mehand MS, Millett P, Al-Shorbaji F, Roth C, Kieny MP, Murgue B. World Health Organization Methodology to Prioritize Emerging Infectious Diseases in Need of Research and Development. Emerging Infectious Disease journal. 2018;24(9)

70. WHO. 2017 Annual review of diseases prioritized under the Research and Development Blueprint. 2017.

71. Mehand MS, Al-Shorbaji F, Millett P, Murgue B. The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. Antiviral Research. 2018;159:63-67.

72. WHO. 2018 Annual review of diseases prioritized under the Research and Development Blueprint. 2018.

73. Sabir M, Ali Y, Muhammad N. Forecasting incidence of dengue and selecting best method for prevention. J Pak Med Assoc. 2018;68(9):1383-1386.

74. Roldán ÚB, Badia X, Marcos-Rodríguez JA, et al. Multi-criteria decision analysis as a decision-support tool for drug evaluation: a pilot study in a pharmacy and therapeutics committee setting. International Journal of Technology Assessment in Health Care. 2018;34(5):519-526.

75. Postmus D, Richard S, Bere N, et al. Individual Trade-Offs Between Possible Benefits and Risks of Cancer Treatments: Results from a Stated Preference Study with Patients with Multiple Myeloma. The Oncologist. 2017;23:theoncologist.2017-0257.

76. Mendoza-Sanchez J, Silva F, Rangel L, et al. Benefit, risk and cost of new oral anticoagulants and warfarin in atrial fibrillation; A multicriteria decision analysis. PLOS ONE. 2018;13(5):e0196361.

77. Jiménez A, Ais A, Beaudet A, Gil A. Determining the value contribution of selexipag for the treatment of pulmonary arterial hypertension (PAH) in Spain using reflective multi-criteria decision analysis (MCDA). Orphanet Journal of Rare Diseases. 2018;13(1):220.

78. Fdez-Olivares J, Onaindia E, Castillo L, Jordán J, Cózar J. Personalized conciliation of clinical guidelines for comorbid patients through multi-agent planning. Artificial Intelligence in Medicine. 2019/05/01/ 2019;96:167-186.

79. Bots PWG, Hulshof JAM. Designing multi-criteria decision analysis processes for priority setting in health policy. Journal of Multi-Criteria Decision Analysis. 2000;9(1-3):56-75.

80. Eiring Ø, Brurberg KG, Nytrøen K, Nylenna M. Rapid methods including network meta-analysis to produce evidence in clinical decision support: a decision analysis. Systematic Reviews. 2018;7(1):168.

81. Nutt DJ, King LA, Phillips LD. Drug harms in the UK: a multicriteria decision analysis. The Lancet. 2010;376(9752):1558-1565.

82. Rossi D, Marciano F, Cabassa P. A multi-criteria methodology for evaluating alternative ultrasound devices. Ergonomics. 2019;62(10):1301-1312.

83. Samanlioglu F, Kaya BE. Evaluation of the COVID-19 Pandemic Intervention Strategies with Hesitant F-AHP. Journal of Healthcare Engineering. 2020;2020:8835258.

84. Howard S, Scott IA, Ju H, McQueen L, Scuffham PA. Multicriteria decision analysis (MCDA) for health technology assessment: the Queensland Health experience. Aust Health Rev. Oct 2019;43(5):591-599.

85. De Nardo P, Gentilotti E, Mazzaferri F, et al. Multi-Criteria Decision Analysis to prioritize hospital admission of patients affected by COVID-19 in low-resource settings with hospital-bed shortage. International Journal of Infectious Diseases. 2020;98:494-500.

86. Ozsahin I, Abebe S, Mok G. A multi-criteria decision-making approach for schizophrenia treatment techniques. Archives of Psychiatry and Psychotherapy. 06/25 2020;22:52-61.

87. Cleemput I, Devriese S, Kohn L, et al. Incorporating societal preferences in reimbursement decisions – Relative importance of decision criteria according to Belgian citizens. 2014.

88. Cleemput I, Devriese S, Christiaens W, Kohn L. Multi-criteria decision analysis for the appraisal of medical needs: a pilot study. Health Services Research (HSR). 2016. KCE Reports 272.

89. Cleemput I, Devriese S, Kohn L, Westhovens R. A multi-criteria decision approach for ranking unmet needs in healthcare. Health Policy. 2018;122(8):878-884.

90. Bouwknegt M, Devleesschauwer B, Graham H, Robertson LJ, van der Giessen JW. Prioritisation of food-borne parasites in Europe, 2016. Euro Surveill. Mar 2018;23(9)

91. Rahman M, Islam A, Islam MN. Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. Modeling Earth Systems and Environment. 2021;7:3.

92. Ezeife DA, Dionne F, Fares AF, et al. Value assessment of oncology drugs using a weighted criterion-based approach. Cancer. 2020;126(7):1530-1540.

93. Sussex M, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. Multi-Criteria Decision Analysis to Value Orphan Medicines. OHE Research Paper. 2013;13/03

94. Otten A, Fazil A, Chemeris A, Breadner P, Ng V. Prioritization of vector-borne diseases in Canada under current climate and projected climate change. Microbial Risk Analysis. 2020/04/01/ 2020;14:100089.

95. Richman MB, Forman EH, Bayazit Y, Einstein DB, Resnick MI, Stovsky MD. A novel computer based expert decision making model for prostate cancer disease management. J Urol. Dec 2005;174(6):2310-8.

96. Samanlioglu F. Evaluation of Influenza Intervention Strategies in Turkey with Fuzzy AHP-VIKOR. Journal of Healthcare Engineering. 2019:9486070.

97. ECDC. ECDC tool for the prioritisation of infectious disease threats. Handbook and manual. 2017.

98. Ozkurt P, Koyuncu M, Hallak J. Primary Health Care Center (PHCC) Location-Allocation with Multi-Objective Modelling: A Case Study in Idleb, Syria. International Journal of Environmental Research and Public Health. 2019;16:811.

99. Improta G, Perrone A, Russo MA, Triassi M. Health technology assessment (HTA) of optoelectronic biosensors for oncology by analytic hierarchy process (AHP) and Likert scale. BMC Medical Research Methodology. 2019;19(1):140.

100. Tang L. Helping the decision maker effectively promote various experts' views into various optimal solutions to China's institutional problem of health care provider selection through the organization of a pilot health care provider research system. Health Research Policy and Systems. 013;11(1):11.

101. Hummel JM, Steuten LGM, Groothuis-Oudshoorn CJM, Mulder N, Ijzerman MJ. Preferences for Colorectal Cancer Screening Techniques and Intention to Attend: a Multi-Criteria Decision Analysis. Applied Health Economics and Health Policy. 2013;11(5):499-507.

102. Stevens KB, Gilbert M, Pfeiffer DU. Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia: A spatial multicriteria decision analysis approach. Spatial and Spatio-temporal Epidemiology. 2013;4:1-14.

103. Ramli A, Aljunid SM, Sulong S, Md Yusof FA. National Drug Formulary review of statin therapeutic group using the multiattribute scoring tool. Ther Clin Risk Manag. 2013;9:491-504.

104. Cox R, Sanchez J, Revie CW. Multi-Criteria Decision Analysis Tools for Prioritising Emerging or Re-Emerging Infectious Diseases Associated with Climate Change in Canada. PLOS ONE. 2013;8(8):e68338.

105. Bentsi-Enchill A, Clemens J. Immunization Implementation Research Priority Setting. Executive Summary. 2013. Working Group on Immunization Implementation Research Prioritization.

106. SAGE. Strategic Advisory Group of Experts (SAGE) 5-7 November 2013. Accessed 22 January, 2021

107. Del Rio Vilas VJ, Burgeño A, Montibeller G, Clavijo A, Vigilato MA, Cosivi O. Prioritization of capacities for the elimination of dog-mediated human rabies in the Americas: building the framework. Pathog Glob Health. Oct 2013;107(7):340-5.

108. Ghandour R, Shoaibi A, Khatib R, et al. Priority setting for the prevention and control of cardiovascular diseases: multi-criteria decision analysis in four eastern Mediterranean countries. International Journal of Public Health. 2015;60(1):73-81.

109. Kolasa K, Zwolinski KM, Kalo Z, Hermanowski T. Potential impact of the implementation of multiple-criteria decision analysis (MCDA) on the Polish pricing and reimbursement process of orphan drugs. Orphanet Journal of Rare Diseases. 2016;11(1):23.

110. Angelis A, Thursz M, Ratziu V, et al. Early Health Technology Assessment during Nonalcoholic Steatohepatitis Drug Development: A Two-Round, Cross-Country, Multicriteria Decision Analysis. Medical Decision Making. 2020;40(6):830-845.

111. Angelis A, Linch M, Montibeller G, et al. Multiple Criteria Decision Analysis for HTA across four EU Member States: Piloting the Advance Value Framework. Soc Sci Med. 2020;246:112595.

112. Babashov V, Ben Amor S, Reinhardt G. Framework for Drug Formulary Decision Using Multiple-Criteria Decision Analysis. Med Decis Making. 2020;40(4):438-447.

113. Costa I, Maêda S, Teixeira L, Gomes CF, Santos M. Choosing a hospital assistance ship to fight the covid-19 pandemic. Revista de saude publica. 2020;54:79.

114. Castro H, Goetghebeur M, Moreno-Mattar O. Testing multi-criteria decision analysis for more transparent resource-allocation decision making in Colombia. International Journal of Technology Assessment in Health Care. 2016;32:1-8.

115. Sullivan T, Hansen P. Determining Criteria and Weights for Prioritizing Health Technologies Based on the Preferences of the General Population: A New Zealand Pilot Study. Value in Health. 2017;20(4):679-686.

116. Hongoh V, Gosselin P, Michel P, et al. Criteria for the prioritization of public health interventions for climate-sensitive vector-borne diseases in Quebec. PLOS ONE. 2017;12(12):e0190049.

117. Strømme EM, Bærøe K, Norheim OF. Disease Control Priorities for Neglected Tropical Diseases: Lessons from Priority Ranking Based on the Quality of Evidence, Cost Effectiveness, Severity of Disease, Catastrophic Health Expenditures, and Loss of Productivity. Developing World Bioethics. 2014;14(3):132-141.

118. Marcelon L, Verstraeten T, Dominiak-Felden G, Simondon F. Quantitative benefit–risk assessment by MCDA of the quadrivalent HPV vaccine for preventing anal cancer in males. Expert Review of Vaccines. 2016;15(1):139-148.

119. EMA. Assessment report. Gardasil & Silgard. Human Papilloma Virus Vaccine. 2014.

120. Reddy BP, Kelly MP, Thokala P, Walters SJ, Duenas A. Prioritising public health guidance topics in the National Institute for Health and Care Excellence using the Analytic Hierarchy Process. Public Health. 2014;128(10):896-903.

121. Tervonen T, Naci H, van Valkenhoef G, et al. Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing among Statins in Primary Prevention. Medical Decision Making. 2015;35(7):859-871.

122. Arslan M. Application of AHP method for the selection o pharmaceutical warehouse location. Ankara Universitesi Eczacilik Fakultesi Dergisi. 2020:253-264.

123. Mohamadi E, Tabatabaei S, Olyaeemanesh A, et al. Coverage Decision-Making for Orthopedics Interventions in the Health Transformation Program in Iran: A Multiple Criteria Decision Analysis (MCDA). Shiraz E-Medical Journal. 2016.

124. Paolucci F, Mentzakis E, Defechereux T, Niessen LW. Equity and efficiency preferences of health policy makers in China—a stated preference analysis. Health Policy and Planning. 2014;30(8):1059-1066.

125. Guarga L, Badia X, Obach M, et al. Implementing reflective multicriteria decision analysis (MCDA) to assess orphan drugs value in the Catalan Health Service (CatSalut). Orphanet Journal of Rare Diseases. 2019;14(1):157.

126. Araz OM. Integrating complex system dynamics of pandemic influenza with a multi-criteria decision making model for evaluating public health strategies. Journal of Systems Science and Systems Engineering. 2013;22(3):319-339.

127. Roselli LRP, Frej EA, Ferreira RJP, Alberti AR, de Almeida AT. Utility-Based Multicriteria Model for Screening Patients under the COVID-19 Pandemic. Computational and Mathematical Methods in Medicine. 2020;2020:9391251.

128. Lin CS, Harris SL. A Unified Framework for the Prioritization of Organ Transplant Patients: Analytic Hierarchy Process, Sensitivity and Multifactor Robustness Study. Journal of Multi-Criteria Decision Analysis. 2013;20(3-4):157-172.

129. Rahimi F, Goli A, Rezaee R. Hospital location-allocation in Shiraz using Geographical Information System (GIS). Shiraz E Medical Journal. 08/01 2017;In Press:57572.

130. World Health Organization & TDR Thematic Reference Group on Environment Agriculture and Infectious Diseases of Poverty. Research priorities for the environment, agriculture and infectious diseases of poverty: technical report of the TDR Thematic Reference Group on Environment, Agriculture and Infectious Diseases of Poverty. 2013. Technical report series; 976. 9789241209762

0512-3054. 2013. https://apps.who.int/iris/handle/10665/78129

131. World Health Organization & Food Agriculture Organization of the United Nations. Multicriteria-based ranking for risk management of food-borne parasites: report of a Joint FAO/WHO expert meeting, 3-7 September 2012, FAO Headquarters, Rome, Italy. Microbiological risk assessment series;23. FAO, World Health Organization; 2014.

132. De Laet C, Hanquet G, Hendrickx E. Multi criteria decision analysis to select priority diseases for Newborn blood screening. Health Services Research (HSR). 2016. KCE Reports 267.

133. Rello J, Lode H, Cornaglia G, Masterton R, The VAPCBC. A European care bundle for prevention of ventilator-associated pneumonia. Intensive Care Medicine. 2010;36(5):773-780.

134. Kwon S-H, Park S-K, Byun J-H, Lee E-K. Eliciting societal preferences of reimbursement decision criteria for anti cancer drugs in South Korea. Expert Review of Pharmacoeconomics & Outcomes Research. 2017;17(4):411-419.

135. Sayan M, Sarıgul F, Sanlidag T, Uzun B, Uzun Ozsahin D, Ozsahin I. Capacity Evaluation of Diagnostic Tests For COVID-19 Using Multicriteria Decision-Making Techniques. Computational and Mathematical Methods in Medicine. 08/06 2020;2020:1-8.

136. Ruggeri M, Cadeddu C, Roazzi P, Mandolini D, Grigioni M, Marchetti M. Multi–Criteria–Decision–Analysis (MCDA) for the Horizon Scanning of Health Innovations an Application to COVID 19 Emergency. International Journal of Environmental Research and Public Health. 2020;17(21):7823.

137. Carnero MC, Gómez A. A multicriteria decision making approach applied to improving maintenance policies in healthcare organizations. BMC Medical Informatics and Decision Making. 2016;16(1):47.

138. Marsh K, Zaiser E, Orfanos P, et al. Evaluation of COPD Treatments: A Multicriteria Decision Analysis of Aclidinium and Tiotropium in the United States. Value in Health. 2017;20(1):132-140.

139. Maseleno A, Tang A, Mahmoud M, Othman M, Saputra S, Muslihudin M. Fuzzy AHP Method to Determine Headache Types based on Symptoms. Investigación clínica. 2017;58:235-245.

140. Dehe B, Bamford D. Development, test and comparison of two Multiple Criteria Decision Analysis (MCDA) models: A case of healthcare infrastructure location. Expert Systems with Applications. 2015;42(19):6717-6727.

141. Dehe B, Bamford D, Bamford J. An Application of a MCDA Model for Healthcare Site Selection. presented at: 22nd Annual Production and Operations Management Society (POMS) Conference 2011; 2011; Reno, Nevada, USA.

142. Schug S, Pogatzki-Zahn E, Phillips LD, et al. Multi-Criteria Decision Analysis to Develop an Efficacy-Safety Profile of Parenteral Analgesics Used in the Treatment of Postoperative Pain. Journal of pain research. 2020;13:1969-1977.

143. Moore A, Crossley A, Ng B, Phillips L, Sancak Ö, Rainsford KD. Use of multicriteria decision analysis for assessing the benefit and risk of over-the-counter analgesics†. Journal of Pharmacy and Pharmacology. 2017;69(10):1364-1373.

144. Carter KJ, Ritchey NP, Castro F, Caccamo LP, Kessler E, Erickson BA. Analysis of Three Decision-making Methods: A Breast Cancer Patient as a Model. Medical Decision Making. 1999;19(1):49-57.

145. Angelis A, Montibeller G, Hochhauser D, Kanavos P. Multiple criteria decision analysis in the context of health technology assessment: a simulation exercise on metastatic colorectal cancer with multiple stakeholders in the English setting. BMC medical informatics and decision making. 2017;17(1):149-149.

146. Byun J-H, Kwon S-H, Ha J-H, Lee E-K. A benefit-risk assessment model for statins using multicriteria decision analysis based on a discrete choice experiment in Korean patients. Therapeutics and clinical risk management. 2016;12:965-974.

147. Hancerliogullari G, Hancerliogullari KO, Koksalmis E. The use of multi-criteria decision making models in evaluating anesthesia method options in circumcision surgery. BMC medical informatics and decision making. 2017;17(1):14-14.

148. Hsu JC, Tang DH, Lu CY. Risk-benefit assessment of oral phosphodiesterase type 5 inhibitors for treatment of erectile dysfunction: a multiple criteria decision analysis. Int J Clin Pract. 2015;69(4):436-43.

149. Gopi G, Kannan K. Application of Analytical Hierarchy Process to Select Best Method for the Preparation of Antidiabetic Drug Loaded Nanoparticles. Journal of Pharmaceutical Sciences and Research. 2015;7:175-181.

150. Garau M, Hampson G, Devlin N, Mazzanti NA, Profico A. Applying a Multicriteria Decision Analysis (MCDA) Approach to Elicit Stakeholders' Preferences in Italy: The Case of Obinutuzumab for Rituximab-Refractory Indolent Non-Hodgkin Lymphoma (iNHL). PharmacoEconomics - Open. 2018/06/01 2018;2(2):153-163.

151. Rello J, Kalwaje Eshwara V, Lagunes L, et al. A global priority list of the TOp TEn resistant Microorganisms (TOTEM) study at intensive care: a prioritization exercise based on multi-criteria decision analysis. European Journal of Clinical Microbiology & Infectious Diseases. 2019;38(2):319-323.

152. Dias Q, Santos M. Multi-Criteria Model for Evaluating Drugs to Prevent Deep Venous Thrombosis Associated With Orthopedic Surgery: A Hospital-Based Case Study. Value in Health Regional Issues. 2020;23:105-111.

153. Özcan E, Ökten S, Eren T. Decision making for promising quinoline-based anticancer agents through combined methodology. Journal of Biochemical and Molecular Toxicology. 2020;34(9):e22522.

154. Osborn J, Roberts T, Guillen E, et al. Prioritising pathogens for the management of severe febrile patients to improve clinical care in low- and middle-income countries. BMC Infectious Diseases. 2020;20(1):117.

155. Tervonen T, Ustyugova A, Sri Bhashyam S, et al. Comparison of Oral Anticoagulants for Stroke Prevention in Nonvalvular Atrial Fibrillation: A Multicriteria Decision Analysis. Value in Health. 2017;20(10):1394-1402.

156. Rajabi M, Mansourian A, Bazmani A. Susceptibility mapping of visceral leishmaniasis based on fuzzy modelling and group decision-making methods. Geospat Health. Nov 2012;7(1):37-50.

157. Peacock SJ, Richardson JRJ, Carter R, Edwards D. Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). Social Science & Medicine. 2007;64(4):897-910.

158. Tsiachristas A, Cramm JM, Nieboer A, Rutten-van Mölken M. Broader economic evaluation of disease management programs using multi-criteria decision analysis. Int J Technol Assess Health Care. 2013;29(3):301-8.

159. Bettinger TL, Shuler G, Jones DR, Wilson JP. Schizophrenia: Multi-Attribute Utility Theory Approach to Selection of Atypical Antipsychotics. Annals of Pharmacotherapy. 2007;41(2):201-207.

160. Vidal L-A, Sahin E, Martelli N, Berhoune M, Bonan B. Applying AHP to select drugs to be produced by anticipation in a chemotherapy compounding unit. Expert Systems with Applications. 2010;37(2):1528-1534.

161. Chung S, Kim S, Kim J, Sohn K. Use of multiattribute utility theory for formulary management in a health system. American Journal of Health-System Pharmacy. 2010;67(2):128-135.

162. Ivlev I, Vacek J, Kneppo P. Multi-criteria decision analysis for supporting the selection of medical devices under uncertainty. European Journal of Operational Research. 2015;247(1):216-228.

163. Buttigieg SC, Gauci D, Bezzina F, Dey PK. Post-surgery length of stay using multi-criteria decision-making tool. J Health Organ Manag. 2018;32(4):514-531.

164. Oketch-Rabah HA, Hardy ML, Patton AP, et al. Multi-Criteria Decision Analysis Model for Assessing the Risk from Multi-Ingredient Dietary Supplements (MIDS). Journal of Dietary Supplements. 2021;18(3):293-315.

165. Mokrini AE, Aouam T. A fuzzy multi-criteria decision analysis approach for risk evaluation in healthcare logistics outsourcing: Case of Morocco. Health Services Management Research. 2020;33(3):143-155.

166. Cruden G, Frerichs L, Powell BJ, Lanier P, Brown CH, Lich KH. Developing a Multi-Criteria Decision Analysis Tool to Support the Adoption of Evidence-Based Child Maltreatment Prevention Programs. Prevention Science. 2020/11/01 2020;21(8):1059-1064.

167. Chen T-C, Lin Y-C. A FAHP-FTOPSIS approach for bioprinter selection. Health and Technology. 2020;10.

168. Rossetti MD, Selandari F. Multi-objective analysis of hospital delivery systems. Computers & Industrial Engineering. 2001;41(3):309-333.

169. Liberatore MJ, Myers RE, Nydick RL, et al. Decision counseling for men considering prostate cancer screening. Computers & Operations Research. 2003;30(10):1421-1434.

170. Cho K-T, Kim S-M. Selecting medical devices and materials for development in Korea: the analytic hierarchy process approach. The International Journal of Health Planning and Management. 2003;18(2):161-174.

171. Chang C-L, Cheng B-W, Su J-L. Using case-based reasoning to establish a continuing care information system of discharge planning. Expert Systems with Applications. 2004;26(4):601-613.

172. Kip MMA, Hummel JM, Eppink EB, et al. Understanding the adoption and use of point-of-care tests in Dutch general practices using multi-criteria decision analysis. BMC Family Practice. 2019;20(1):8.

173. Li K, Luo S, Yuan S, Mt-Isa S. A Bayesian approach for individual-level drug benefit-risk assessment. Statistics in medicine. 2019;38(16):3040-3052.

174. Vermersch P, Martinelli V, Pfleger C, et al. Benefit-risk Assessment of Cladribine Using Multi-criteria Decision Analysis (MCDA) for Patients With Relapsing-remitting Multiple Sclerosis. Clinical Therapeutics. 2019;41(2):249-260.e18.

175. Mortimer D, Iezzi A, Dickins M, et al. Using co-creation and multi-criteria decision analysis to close service gaps for underserved populations. Health Expectations. 2019;22(5):1058-1068.

176. Eghbali-Zarch M, Tavakkoli-Moghaddam R, Esfahanian F, Sepehri MM, Azaron A. Pharmacological therapy selection of type 2 diabetes based on the SWARA and modified MULTIMOORA methods under a fuzzy environment. Artificial Intelligence in Medicine. 2018;87:20-33.

177. Hsu JC, Lin J-Y, Lin P-C, Lee Y-C. Comprehensive value assessment of drugs using a multi-criteria decision analysis: An example of targeted therapies for metastatic colorectal cancer treatment. PloS one. 2019;14(12):e0225938-e0225938.

178. Holmes EAF, Plumpton C, Baker GA, et al. Patient-Focused Drug Development Methods for Benefit-Risk Assessments: A Case Study Using a Discrete Choice Experiment for Antiepileptic Drugs. Clinical pharmacology and therapeutics. 2019;105(3):672-683.

179. Choi SE, Berkowitz SA, Yudkin JS, Naci H, Basu S. Personalizing Second-Line Type 2 Diabetes Treatment Selection: Combining Network Meta-analysis, Individualized Risk, and Patient Preferences for Unified Decision Support. Medical Decision Making. 2019;39(3):239-252.

180. de Paula RL, de Almeida J, Cavalcante SFA, et al. Molecular Modeling and In Vitro Studies of a Neutral Oxime as a Potential Reactivator for Acetylcholinesterase Inhibited by Paraoxon. Molecules. 12 2018;23(11)

181. Lin Q, Wang D. Facility Layout Planning with SHELL and Fuzzy AHP Method Based on Human Reliability for Operating Theatre. Journal of Healthcare Engineering. 2019;2019:8563528.

182. Bollaerts K, Ledent E, de Smedt T, et al. ADVANCE system testing: Benefit-risk analysis of a marketed vaccine using multi-criteria decision analysis and individual-level state transition modelling. Vaccine. 2020;38 Suppl 2:B65-b75.

183. Badia X, Aguarón A, Fernández A, et al. Patient involvement in reflective multicriteria decision analysis to assist decision making in oncology. Int J Technol Assess Health Care. 2019;35(1):56-63.

184. Akcan S, Güldeş M. Integrated Multicriteria Decision-Making Methods to Solve Supplier Selection Problem: A Case Study in a Hospital. Journal of Healthcare Engineering. 2019;5614892.

185. Alsalem MA, Zaidan AA, Zaidan BB, et al. Multiclass Benchmarking Framework for Automated Acute Leukaemia Detection and Classification Based on BWM and Group-VIKOR. Journal of Medical Systems. 2019;43(7):212.

186. Nutt DJ, Phillips LD, Balfour D, et al. Estimating the harms of nicotine-containing products using the MCDA approach. Eur Addict Res. 2014;20(5):218-25.

187. Wahlster P, Goetghebeur M, Schaller S, Kriza C, Kolominsky-Rabas P. Exploring the perspectives and preferences for HTA across German healthcare stakeholders using a multi-criteria assessment of a pulmonary heart sensor as a case study. Health Res Policy Syst. 2015;13:24.

188. Wahlster P, Goetghebeur M, Schaller S, Kriza C, Niederländer C, Kolominsky-Rabas P. German healthcare stakeholders perspectives regarding the value of a diagnostic device for heart failure: First results of a Multi-criteria Decision Analysis (MCDA). 2013.

189. Diaz-Ledezma C, Lichstein PM, Dolan JG, Parvizi J. Diagnosis of periprosthetic joint infection in Medicare patients: multicriteria decision analysis. Clin Orthop Relat Res. 2014;472(11):3275-84.

190. Aenishaenslin C, Gern L, Michel P, et al. Adaptation and Evaluation of a Multi-Criteria Decision Analysis Model for Lyme Disease Prevention. PLOS ONE. 2015;10(8):e0135171.

191. van Amsterdam J, Nutt D, Phillips L, van den Brink W. European rating of drug harms. J Psychopharmacol. 2015;29(6):655-60.

192. Almahdi EM, Zaidan AA, Zaidan BB, Alsalem MA, Albahri OS, Albahri AS. Mobile-Based Patient Monitoring Systems: A Prioritisation Framework Using Multi-Criteria Decision-Making Techniques. Journal of Medical Systems. 2019;43(7):219.

193. Abdel-Basset M, Manogaran G, Gamal A, Smarandache F. A Group Decision Making Framework Based on Neutrosophic TOPSIS Approach for Smart Medical Device Selection. Journal of Medical Systems. 2019;43(2):38.

194. Martelli N, Hansen P, van den Brink H, et al. Combining multi-criteria decision analysis and mini-health technology assessment: A funding decision-support tool for medical devices in a university hospital setting. J Biomed Inform. 2016;59:201-8. d

195. Hongoh V, Michel P, Gosselin P, et al. Multi-Stakeholder Decision Aid for Improved Prioritization of the Public Health Impact of Climate Sensitive Infectious Diseases. Int J Environ Res Public Health. 2016;13(4):419.

196. Suner A, Oruc OE, Buke C, Ozkaya HD, Kitapcioglu G. Evaluation of infectious diseases and clinical microbiology specialists' preferences for hand hygiene: analysis using the multi-attribute utility theory and the analytic hierarchy process methods. BMC Med Inform Decis Mak. 2017;17(1):129.

197. Levitan B, Phillips LD, Walker S. Structured Approaches to Benefit-Risk Assessment: A Case Study and the Patient Perspective. Ther Innov Regul Sci. Sep 2014;48(5):564-573.

198. Kyavars V, Subramanian S. Multi Criteria Decision Making to Select the Suitable Method for the Preparation of Nanocrystals Using an Analytic Hierarchy Process. Indian Journal of Pharmaceutical Education and Research. 2017;51(2):249-259.

199. Hongoh V, Campagna C, Panic M, et al. Assessing Interventions to Manage West Nile Virus Using Multi-Criteria Decision Analysis with Risk Scenarios. PLOS ONE. 2016;11(8):e0160651.

200. Al-Badriyeh D, Alabbadi I, Fahey M, Al-Khal A, Zaidan M. Multi-indication Pharmacotherapeutic Multicriteria Decision Analytic Model for the Comparative Formulary Inclusion of Proton Pump Inhibitors in Qatar. Clin Ther. May 2016;38(5):1158-73.

201. Mobinizadeh M, Raeissi P, Nasiripour AA, Olyaeemanesh A, Tabibi SJ. A model for priority setting of health technology assessment: the experience of AHP-TOPSIS combination approach. Daru. 2016;24:10.

202. Claudio D, Kremer GEO, Bravo-Llerena W, Freivalds A. A dynamic multi-attribute utility theory–based decision support system for patient prioritization in the emergency department. IIE Transactions on Healthcare Systems Engineering. 2014;4(1):1-15.

203. Wagner M, Khoury H, Bennetts L, et al. Appraising the holistic value of Lenvatinib for radio-iodine refractory differentiated thyroid cancer: A multi-country study applying pragmatic MCDA. BMC Cancer. 17 2017;17(1):272.

204. Iskrov G, Miteva-Katrandzhieva T, Stefanov R. Multi-Criteria Decision Analysis for Assessment and Appraisal of Orphan Drugs. Frontiers in public health. 2016;4:214-214.

205. Yang Z, Li X, Garg H, Qi M. Decision Support Algorithm for Selecting an Antivirus Mask over COVID-19 Pandemic under Spherical Normal Fuzzy Environment. International Journal of Environmental Research and Public Health. 2020;17(10):3407.

206. de Greef-van der Sandt I, Newgreen D, Schaddelee M, et al. A quantitative benefit-risk assessment approach to improve decision making in drug development: Application of a multicriteria decision analysis model in the development of combination therapy for overactive bladder. Clin Pharmacol Ther. 2016;99(4):442-51.

207. Kalhor R, Kiaei MZ, Mohebbifar R, et al. Prioritizing the options for health-care waste management in qazvin: Using a multi-criteria decision making approach. Article. Journal of Biology and Today's World. 2016;5(6):101-106.

208. Nixon R, Dierig C, Mt-Isa S, et al. A case study using the PrOACT-URL and BRAT frameworks for structured benefit risk assessment. Biom J. 2016;58(1):8-27.

209. Voudrias EA. Technology selection for infectious medical waste treatment using the analytic hierarchy process. Journal of the Air & Waste Management Association. 2016;66(7):663-672.

210. Goetghebeur MM, Wagner M, Nikodem M, Zyla A, Micaleff A, Amzal B. Pragmatic Multicriteria Decision Analysis (MCDA) Combined With Advanced Pharmacoepidemiology for Benefit-Risk Assessments of Medicines Adapted to the Real-Life Constraints of Regulators: Development and Case Study. Ther Innov Regul Sci. 2016;50(5):620-631.

211. Alimohammadzadeh K, Bahadori M, Hassani F. Application of Analytical Hierarchy Process Approach for Service Quality Evaluation in Radiology Departments: A Cross-Sectional Study. Iran J Radiol. 2016;13(1):e29424.

212. Kunasekaran V, Krishnamoorthy K. Multi criteria decision making to select the best method for the preparation of solid lipid nanoparticles of rasagiline mesylate using analytic hierarchy process. J Adv Pharm Technol Res. 2014;5(3):115-21.

213. Maro JC, Fryback DG, Lieu TA, Lee GM, Martin DB. Responding to Vaccine Safety Signals during Pandemic Influenza: A Modeling Study. PLOS ONE. 2014;9(12):e115553.

214. Mahfoud H, Barkany AE, Biyaali AE. A Hybrid Decision-Making Model for Maintenance Prioritization in Health Care Systems. American Journal of Applied Sciences. 2016;13(4)

215. Carnero C. Fuzzy Multicriteria Model for Selection of Vibration Technology. Shock and Vibration. 2016:1-18.

216. Bowers J, Cheyne H, Mould G, et al. A multicriteria resource allocation model for the redesign of services following birth. BMC Health Services Research. 2018;18(1):656.

217. Qian W, Wang ZJ, Li KW. Medical Waste Disposal Method Selection Based on a Hierarchical Decision Model with Intuitionistic Fuzzy Relations. Int J Environ Res Public Health. 2016;13(9)

218. Ortíz MA, Felizzola HA, Isaza SN. A contrast between DEMATEL-ANP and ANP methods for six sigma project selection: a case study in healthcare industry. BMC Medical Informatics and Decision Making. 2015;15(3):S3.

219. Oztürk N, Tozan H. A Fuzzy-Based Decision Support Model for Selecting the Best Dialyser Flux in Haemodialysis. J Healthc Eng. 2015;6(3):303-24.

220. Panattoni L, Phelps CE, Lieu TA, et al. Feasibility of Measuring Preferences for Chemotherapy Among Early-Stage Breast Cancer Survivors Using a Direct Rank Ordering Multicriteria Decision Analysis Versus a Time Trade-Off. Patient. 2020;13(5):557-566.

221. Öztürk N, Tozan H, Vayvay Ö. A New Decision Model Approach for Health Technology Assessment and A Case Study for Dialysis Alternatives in Turkey. Int J Environ Res Public Health. 2020;17(10).

222. Ozsahin I. Identifying a Personalized Anesthetic with Fuzzy PROMETHEE. Healthcare informatics research. 2020;26(3):201-211.

223. Parsia Y, Sorooshian S. Minimization of nosocomial infections risks by a decision algorithm for upgrading of healthcare facilities. Journal of Infection and Public Health. 2020;13(5):746-752.

224. Shojaei P, bordbar N, Ghanbarzadegan A, Najibi M, Bastani P. Ranking of Iranian provinces based on healthcare infrastructures: before and after implementation of Health Transformation Plan. Cost Effectiveness and Resource Allocation. 2020;18(1):4.

225. Gómez A, Carnero MC. Decision Support System for maintenance policy optimization in medicinal gases subsystems. IFAC-PapersOnLine. 2016;49(28):268-273.

226. Parsia Y, Sorooshian S. A Decision-Making Algorithm for Rearchitecting of Healthcare Facilities to Minimize Nosocomial Infections Risks. International journal of environmental research and public health. 2020;17(3):855.

227. Ren Z, Liao H, Liu Y. Generalized Z-numbers with hesitant fuzzy linguistic information and its application to medicine selection for the patients with mild symptoms of the COVID-19. Computers & industrial engineering. 2020;145:106517-106517.

228. Kalid N, Zaidan AA, Zaidan BB, et al. Based on Real Time Remote Health Monitoring Systems: A New Approach for Prioritization "Large Scales Data" Patients with Chronic Heart Diseases Using Body Sensors and Communication Technology. J Med Syst. 2018;42(4):69.

Appendix 3.D. The 10 most used criteria by studies informing prioritisation of interventions in high-income countries

Criteria	Examples
Cost, budget or Economic impact	"Minimise monetary costs (i.e. This includes health program expenditures, hospital costs and loss of future income): Mill. NOK (7 NOK=1 USD)." "Economic impact on public sector" "Budget impact on health plan" "Costs (c): The additional annual funding required in 2009 and 2010 both to set-up and to run the intervention, in £'000. Set-up costs included initial investments such as training and equipment and the running costs included costs such as personnel."
Health risks or Safety	"Complications during and after surgery: Complications were measured by the expected probability and severity of medical and technical complications." ; "Intervention outcomes: Improvement of safety & tolerability (+-SD)" ; "Treatment clinical safety"
Quality considerations or features of the process of care delivery	"Clinical efficiency: The difference between the sum of the examinations that require no repetitions (between sampling and analysis) and the sum of examinations requiring repetitions (between sampling and analysis) at each timestep constitutes the sub-criterion: the greater the clinical efficiency indicator, the greater the efficiency of the technology."; ".Changes in the Process of Care Delivery: Self-management support criterion (measured by the ACIC)" ; "Accessible: How important to you is it... that you can easily find out about, sign-up for, and make use of a service? Take into account eligibility, disability access, and computer use."
Quality of life or well-being considerations	"Comparative patient-perceived health PRO"; "QOL - quality of life: Impact of the new treatment on overall sense of well-being as measured by QOL instruments or feedback from patients, compared to existing treatments" "A. Therapeutic Impact - HRQoL: Health related quality of life using EQ-5D score"
Unspecified health outcomes	"Extra health benefits" "Individual health benefits" "Improvement of efficacy/effectiveness"
Quality or availability of evidence	"Certainty: confidence in the evaluation of the intervention, based on an assessment of the quality of the method and data used in the evaluation" ; "Relevance and validity of evidence" "Quality of evidence: I believe that it strictly meets the quality criteria required for a clinical trial, especially given the difficulties of including patients as it is a low prevalence disease."
Intermediate health outcomes	"Decrease in the quantity/quality of depressive symptoms by at least 50% as measured on a depression scale (e.g., HAM-D, MADRS)" "Clinical effectiveness: 2.1 Hospital readmission related to early disease, 2.2. Request of home post-discharge care, 2.3 Aesthetic appearance, 2.4. Return to normal activities, depending on gender and age, 2.5 Functional rehabilitation, 2.6 Healing" "Changes in Biomedical, Physiological, and Clinical Health Outcomes: Lung function measured in FEV1%predicted."
Feasibility/Acceptability	"Technical feasibility: quality of practical application of the tests depends on the laboratories used" "How much time is required to set up the development? (e.g., are staffing and physical space available, or would training or new construction be needed?); -How long will it

	take to see some results/health improvement from the development?; Describe any potential impact (positive and negative) on other NHS and non-NHS agencies." "Feasibility (p): Probability of success (from 0% to 100%) to achieve the assessed benefits, assuming funding is granted and taking into account: ease of implementation; availability of workforce; acceptability to stakeholder/environment"
Disease severity	"Disease survival prognosis with current standard of care" "a) high mortality often with poor prognosis e.g. cancers (2 points) b) chronic without high mortality and morbidity (1 point), c) severe invalidity, severely harm of capacities central to individuals' functioning in society – e.g., hearing, eyesight etc"; "Disease severity: severity of the cancer disease for patients treated with the proposed intervention (five-year survival rate was used as a measurement scale)"
Alternative treatment/Unmet needs	"Availability of existing treatments" "More effective and selective drugs are needed that reduce mortality and disease progression in clinical trials and in real life. Greater comfort than prostacyclines, so an oral treatment is welcome. Currently available drugs that act on the prostacyclin route are uncomfortable, and they require time and training from patients and caregivers and interfere with the proceeding of a normal daily activity" "Existence of other treatments for the underlying condition, for the specific population targeted by the drug"

Appendices – Chapter 4

Appendix 4.A. Selection of assessment and monitoring criteria

Table 4.A.1. Sequential assessment of the potential assessment criteria, in order of relevance

	Potential candidate value element	***Step 1. Non-redundancy (for evaluation purposes)***	***Step 2. Non-overlap***	***Step 3.* Operationality**	***Assessment criteria***
Identified from semi-structured interviews (**Chapter 2**)	**1. Health outcomes**	Yes	Yes	Yes	*Final or intermediate health outcomes*
	2. Quality of care / Patient experience	Yes	Yes	Yes	*Quality of care or patient experience*
	3. Cost or budget considerations	Yes	Yes	Yes	*Cost (per patient)*
	4. Efficiency or Cost-effectiveness	Yes	No		
	5. Equity	Yes	Yes	Yes	*Equity*
	6. Compliance to policies/ priorities /guidance	No			
	7. Size of the population	Yes	Yes	Yes	*Size of the target population*
Identified from systematic literature review (**Chapter 3**)	8. Cost/budget	Yes	No		
	9. Health risks or Safety	No			
	10. Quality/Process of care delivery	Yes	No		
	11. Quality of life / well-being	Yes	Yes	Yes	*Quality of life and well-being considerations*
	12. Unspecified health outcomes	Yes	No		
	13. Quality/Availability of evidence	No			
	14. Intermediate health outcomes	Yes	No		
	15. Feasibility or Acceptability	Yes	Yes	No	
	16. Disease severity	No			
	17. Alternative treatment or Unmet needs	No			
	18. Size of population	Yes	No		
	19. Poverty/equity	Yes	No		
	20. Final health outcomes	Yes	No		
	21. Compliance to policies/ priorities/guidelines	Yes	No		
	22. Cost/effectiveness	Yes	No		
	23. Innovation	No			
	24. Externalities	Yes	Yes	No	
	25. Care continuum	Yes	No		
	26. Patient lifestyle	Yes	No		
	27. Ethics	No			
	28. Age/Gender	No			

Table 4.A.2. Sequential assessment of the potential monitoring criteria, in order of relevance

	Potential candidate value element	*Step 1. Non-redundancy (for monitoring purposes)*	*Step 2. Non-overlap*	*Step 3.* Operationality	*Monitoring criteria*
Identified from semi-structured interviews (**Chapter 2**)	**1. Health outcomes**	Yes	Yes	Yes	*Intermediate health outcomes or quality of life outcomes*
	2. Quality of care / Patient experience	Yes	Yes	Yes	*Quality of care or features of the process of care delivery*
	3. Cost or budget considerations	No			
	4. Efficiency or Cost-effectiveness	No			
	5. Equity	Yes	Yes	Yes	*Equity in quality of care*
	6. Compliance to policies/ priorities /guidance	Yes	Yes	Yes	*Compliance with national standards*
	7. Size of the population	No	No		
Identified from systematic literature review (**Chapter 3**)	8. Cost/budget	No			
	9. Health risks or Safety	No			
	10. Quality/Process of care delivery	Yes	No		
	11. Quality of life / well-being	Yes	Yes	Yes	*Intermediate health outcomes or quality of life outcomes (the same indicated above)*
	12. Unspecified health outcomes	Yes	No		
	13. Quality/Availability of evidence	No			
	14. Intermediate health outcomes	Yes	No		
	15. Feasibility or Acceptability	Yes	Yes	No	
	16. Disease severity	No			
	17. Alternative treatment or Unmet needs	No			
	18. Size of population	No			
	19. Poverty/equity	Yes	No		
	20. Final health outcomes	No			
	21. Compliance to policies/ priorities/guidelines	Yes	No		
	22. Cost/effectiveness	No			
	23. Innovation	No			
	24. Externalities	Yes	Yes	No	
	25. Care continuum	Yes	No		
	26. Patient lifestyle	Yes	No		
	27. Ethics	No			
	28. Age/Gender	No			

Table 4.A.3. Rationale for selecting assessment and monitoring criteria

	Potential assessment criteria	Rationale	Evaluation or Monitoring component?
From semi-structured interviews (a)	**1. Health outcomes**	This criterion encompasses mortality, life expectancy, or quality-adjusted life years (i.e. final outcomes), as well as biomedical, physiological, and clinical health outcomes (i.e. intermediate outcomes). This was the most frequently mentioned criteria during the semi-structured interviews. It meets the *non-redundancy* property, as "improving outcomes population health and healthcare" is the main purpose of the Integrated Care Systems (ICSs) in England, and is the principal aim of any healthcare intervention (12). Additionally, it meets the *non-overlap principle* as it is the first element under assessment. Mortality or survival data is systematically collected in the country. Information about people's health is also recorded when users receive healthcare or social care, and data is available to local healthcare commissioners. Therefore, this criterion meets the *operationality* property.	For evaluation purposes, it is preferable to concentrate on final health outcomes, although some intermediate outcomes may be pertinent and worth evaluating. To monitor the implementation of new care models, it may be useful to observe the progression of several biomedical, physiological, and clinical health outcomes over time.
	2. Quality of care / Patient experience	This criterion satisfies the *non-redundancy* property. In the context of value-based healthcare, this criterion mainly relates to the patients' or carers' experiences when receiving care (12, 187). This is one of the main objectives of health systems as it increases the likelihood for individuals or populations to achieve the desired health outcomes (469). Similarly, one of the core purposes of the ICSs in England is to "tackle inequalities in outcomes, experience and access" (12). There is also no *overlap* or double counting issues with this criterion as it is possible that a patient report improvements on health outcomes while reporting a bad patient experience (or the other way around). Even though patient experience data continues to be under-reported, it is increasingly relevant for decision-makers and some data is already being collected via national and international patient surveys (361, 470, 471). Therefore, this element also complies with the *operationality* property.	This criterion is important when evaluating new care models. Additionally, monitoring quality of care or features of the delivery of care is relevant as it can provide valuable insight into the implementation of new care models. The effective use of quality of care measures for monitoring purposes is dependent on data availability, frequency, and the treated condition.
	3. Cost or budget considerations	This criterion also complies with the *non-redundancy* property. The cost of a health intervention, including new models of care, is relevant due to the limited resources that local commissioners have and the cost pressures in the health sector (1). In the local commissioning context in England, any new model of care is expected to "simultaneous improve patient experience and health status while reducing the cost of healthcare" (Berwick et al., 2008), and ICSs are required to enhance productivity and value for money (12). The efficiency or cost-effectiveness' concept lies behind these ideas (2). As effectiveness or health benefit outcomes have already been included (above), in order to avoid *overlapping*, this element of value was limited to cost or resources required to fund a new care model. This element also complies with the *operationality* property. As local decision makers are directly responsible to commission most of the health	This criterion is included for evaluation purposes. In the context of monitoring, tracking implementation cost might be of interest of decision makers. It is however subject to the type of contract or payment mechanism. For this reason, the costing criterion was not considered for the monitoring component of the framework.

		services for their populations, access to cost data is widely available for them.	
	4. Efficiency or Cost-effectiveness	This criterion is of paramount importance since local commissioners seek to maximise societal welfare by allocating tight health care budgets to the most cost-effective interventions (2). It however overlaps with criterion #1 (health outcomes) and element #3 (Cost or budget considerations) as both are embedded into the cost-effectiveness or efficiency concept.	Not applicable.
	5. Equity	The equity criterion complies with *non-redundancy* property. The health and wellbeing gap has been widely acknowledged by the NHS England and tackling health inequalities has been of the main objectives of the reforms introduced over the last years (88, 95, 99). In the context of the 2022 Health and Care Act (10), tackling inequalities in outcomes, experience and access is part of the core purpose of an ICS (12). It also complies with *non-overlap and operationality* properties. None of the value-based elements already selected address distributional issues. Equity can also be measure in terms of access to care or in terms of health outcomes. As an example, the country developed a series of 'Atlases of Variation' to identify unwarranted variation and assess the value of the healthcare provided both to populations and to individuals (472).	This criterion holds significant importance for assessing and monitoring novel care models. When it comes to evaluation, local commissioners more likely to assess the impact of new care models or simple interventions on inequalities pertaining to health or wellbeing outcomes. For tracking the performance of care models in the short or medium term, the emphasis should be on inequalities in terms of access or other aspects of quality of care.
	6. Compliance to policies/ priorities /guidance	Local healthcare commissioners in England are expected to comply with policies, priorities, guidelines and standards provided by the NHS England and the different advisory or monitoring organisations/agencies within the health system (e.g,. The Department of Health, the National Institute for Health and Care Excellence (NICE), The Care Quality Commission). Compliance to all regulations or guidelines is however not an outcome in itself but rather an element that is part of the whole commissioning process. All health interventions have to comply with NHS policies, priorities and guidance when they are being developed. The policies, priorities and guidelines are either related to the way health interventions should be implemented (i.e., quality of care or features of care delivery, budget limits) or the outputs or early outcomes that health interventions should achieve. An example of the latter are the treatment targets set by the NHS England for specific conditions (e.g,. In the context of diabetes, proportion of patient that have HbA1c ≤58mmol/mol, cholesterol <5mmol/L and blood pressure ≤140/80). Due to these reasons, the compliance to policies/ priorities /guidance element is more relevant at the monitoring than at the assessment stage of the commissioning process. To comply with the *operationality property* and avoid potential overlapping with quality of care monitoring criteria, it was decided to focus on compliance with NHS England standards set for different conditions or care models.	This criterion is included for monitoring purposes.

	7. Size of the population	The more people a healthcare intervention is covering, the better. Even though this criterion is not an outcome variable in a strict sense, it was widely mentioned during the semi-structured interviews, in the form of '*size of the population that receives that intervention*' '*Number of potential beneficiaries*' or '*prevalence of a condition or disease*' (229). Although size of target population might not be relevant when comparing care programmes for the same health condition (i.e. number of recipients would be the same for both interventions), it could be a crucial element in the decision-making process when comparing interventions across conditions. The size of the target population has also been widely used to guide priority-setting decisions in high-income countries (290, 365, 345, 473). Due to these reasons, this criterion complies with the *non-redundancy* property.	This criterion is included for evaluation purposes.
From systematic literature review (b)	8. Cost/budget	Fails to comply with the *non-overlap* property as it was already included (i.e. #3 criterion).	Already assessed and included.
	9. Health risks or Safety	According to the literature review conducted, examples of this criterion include "complications during and after surgery", "improvement of safety & tolerability (+-SD)" or "treatment clinical safety". Although it was not mentioned during the semi-structured interviews, it has been one of the most frequently used criteria when making priority-setting decisions in high-income economies. However, it has been mainly used when assessing and monitoring the effect of pharmaceuticals. Usually, the health technologies (including pharmaceuticals) that are prescribed within a new models of care, have been already granted with marketing authorisation from the Medicines and Healthcare products Regulatory Agency (MHRA), and have been already assessed and approved by NICE. Local commissioners should also fund medicines and treatments approved through the NICE Technology Appraisal (TA) or Highly Specialised Technologies Evaluation (HST) (474). In addition, the Care and Quality Commission (CQC), an independent regulator of health and adult social care in England, is responsible for monitoring, inspecting and regulating services. Due to these reasons, the 'health risks or safety' criterion fails the *non-redundancy property*.	Not applicable.
	10. Quality considerations or features of the process of care delivery	According to the literature review conducted, some examples of this criterion are "accessible: How important to you is it... that you can easily find out about, sign-up for, and make use of a service?" or "self-management support criterion (measured by the ACIC)". Although highly relevant, this criterion fails to comply with the *non-overlap* property as it was already included (i.e., #2 criterion).	Already assessed and included.
	11. Quality of life / well-being	Examples of this criterion include "comparative patient-perceived health PRO" or "QOL - quality of life: Impact of the new treatment on overall sense of well-being as measured by QOL instruments or feedback from patients, compared to existing treatments". This criterion meets the *non-redundancy* property. Improving population health and healthcare includes improving people's quality of life (12, 88, 122, 149, 187). As the country is strengthening the integration agenda, measuring patient's quality of life or	This criterion is important when evaluating new care models. Additionally, monitoring quality of life or patients' wellbeing can offer valuable insight into the implementation of these models. The effective use of quality of life/wellbeing

		well-being is increasingly important to the NHS England (92, 475). As aspects of quality of life have not been addressed by any of the value elements already included, it does not fail to comply with the *non-overlap* property. This criterion also complies with the *operationality* property. Although there is not always data on quality of life-related outcomes, there are some national surveys (e.g. Cancer Quality of life survey) as well as routinely collected records from primary care available (476).	measures for monitoring purposes is dependent on data availability, frequency, and the treated condition. In the context of monitoring, distinguishing between intermediate health outcomes and quality of life outcomes can be challenging. To address this, it is suggested to employing a single monitoring criterion to group all pertinent KPIs: Intermediate Health Outcomes or Quality of Life Outcomes.
	12. Unspecified health outcomes	In the literature review, some studies referred to "extra health benefits", "individual health benefits" or "improvement of efficacy/effectiveness". Although health outcomes are of paramount importance, both, intermediate and final health outcomes are already considered in criterion #1. Consequently, the 'unspecified health outcome element' fails to comply with the *non-overlap* property.	Not applicable.
	13. Quality/ Availability of evidence	Examples of this criterion include "certainty: confidence in the evaluation of the intervention, based on an assessment of the quality of the method and data used in the evaluation", "Relevance and validity of evidence", or "quality of evidence: I believe that it strictly meets the quality criteria required for a clinical trial, especially given the difficulties of including patients as it is a low prevalence disease". Even though this criterion has been used to guide some priority setting decisions, according to the literature review conducted, it has been mainly relevant in the context of pharmaceuticals or medical devices. In addition, the identified criteria aim to complement routinely collected data in assessing the performance of the new models of care. The quality and availability of data stand as key factors in the identification and selection of criteria for assessment and monitoring (i.e., *operationality* property). Consequently, this criterion fails to comply with the non-redundancy property.	Not applicable.
	14. Intermediate health outcomes (eg. biomedical, physiological, and clinical health outcomes)	Examples of this criterion include "decrease in the quantity/quality of depressive symptoms by at least 50% as measured on a depression scale (e.g. HAM-D, MADRS)", "clinical effectiveness: 2.1 Hospital readmission related to early disease" and "lung function measured in FEV1%predicted." Although highly relevant, this criterion fails to comply with the *non-overlap* property as it was already included (i.e., #1 element).	Already assessed and included.
	15. Feasibility or Acceptability	According to the literature review conducted, some examples of this criterion are "technical feasibility: quality	Not applicable.

		of practical application of the tests depends on the laboratories used", "how much time is required to set up the development? (e.g, are staffing and physical space available, or would training or new construction be needed?), "how long will it take to see some results/health improvement from the development?", and "Feasibility (p): Probability of success (from 0% to 100%) to achieve the assessed benefits, assuming funding is granted and taking into account: ease of implementation; availability of workforce; acceptability to stakeholder/environment". This criterion can be highly relevant for decision-makers in the context of new models of care. The feasibility of the intervention should be assessed when planning to implement or scale up a healthcare intervention, while the acceptability of the intervention should be assessed by gathering systematically data on the view of patients, carers and healthcare professionals. The acceptability aspect of an intervention has been mainly considered when assessing pharmaceuticals, and can also be considered a wide concept that relates to patient experience and quality of care. This aspect has already been included in criterion #2 (i.e., this element fails to comply with the *non-overlap* property). On the other hand, data availability on the feasibility and acceptability of an intervention is rather limited due to its complexity, making its operationalisation very difficult (i.e., fails *operationality* property).	
	16. Disease severity	Examples of this criterion include "disease survival prognosis with current standard of care" and "severity of the cancer disease for patients treated with the proposed intervention (five-year survival rate was used as a measurement scale)". Although 'severity' has been widely used to guide priority-setting decisions in high income countries, it fails the *non-redundancy* and *non-overlap* properties due to the following reasons: (i) A prioritisation of diseases process takes places at the NHS England level and is consequently reflected in the funding streams defined by the NHS England and by the ICSs (e.g., cardiovascular disease budget, mental health budget); (ii) Most new models of care cover more than just one condition, (i.e. multi-morbidities) making it difficult to judge the severity of more than one single disease. (iii) This criterion is mostly operationalised in terms of survival or life expectancy (i.e., already covered in criterion #1).	Not applicable.
	17. Alternative treatment or Unmet needs	According to the literature review conducted, examples of this criterion include "More effective and selective drugs are needed that reduce mortality and disease progression in clinical trials and in real life" and "Existence of other treatments for the underlying condition, for the specific population targeted by the drug". Although "*Alternative treatment or Unmet needs*" has also been used by some previous studies informing	Not applicable.

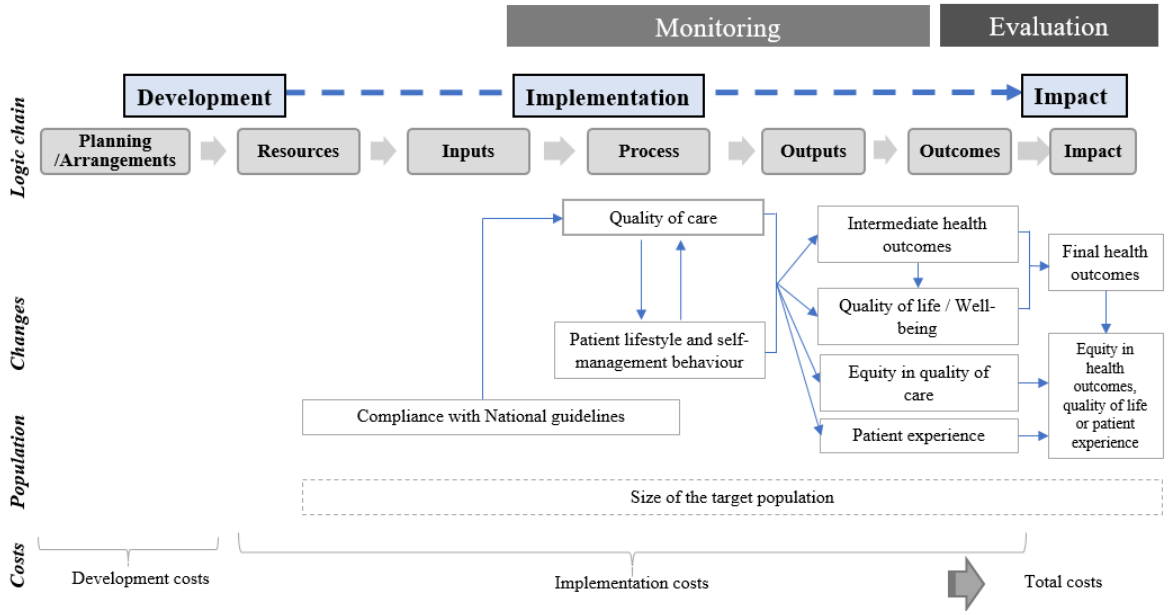
		prioritisation of interventions, it is not considered relevant to guide the local commissioning of healthcare. This criterion has been mainly used when assessing pharmaceuticals or in the context of rare disease. Most of the specialised services to support people with rare and complex conditions are directly commissioned by the NHS England. As part of this process, twice a year, the NHS England makes decisions on which new specialised treatments should be routinely commissioned (477).	
	18. Size of population	Fails to comply with the *non-overlap* property as it was already included (i.e. #7 criterion).	Already assessed and included.
	19. Poverty/equity	Fails to comply with the *non-overlap* property as it was already included (i.e. #5 criterion).	Already assessed and included.
	20. Final health outcomes	Fails to comply with the *non-overlap* property as it was already included (i.e. #1 criterion).	Already assessed and included.
	21. Compliance to policies/ priorities/ guidelines	Already assessed and disqualified (criterion #6)	Already assessed and included.
	22. Cost/effectiveness	Fails to comply with the *non-overlap* property as it was already included (i.e. #3 criterion).	Not applicable.
	23. Innovation	Examples of this criterion include "scientific advance + contribution to patient outcome", "New iteration or new indication of existing technology; substantial innovation compared with alternatives", "a new target or a novel pharmacological mechanism" and "Advanced therapy medicinal product (ATMP) including biopharmaceutical, innovative synthetic entities and delivery systems as well non-biological complex drug". Even though this criterion has been considered in the context of priority-setting, as shown in the literature review conducted, it has been used in the context of pharmaceuticals or medical devices. New models of care are innovations in itself. Consequently, this criterion fails the *non-redundancy* property.	Not applicable.
	24. Externalities	Examples of this criterion include "Human -to-human transmission", "caregiver well-being: the extent to which the new treatment improves caregiver well-being, compared to existing treatment" and "reduce the spread of the disease along the country". Although this criterion has mainly been used when assessing public health interventions, or interventions related to infectious diseases, it could be relevant in the context of new models of care, particularly when considering the potential impact of an intervention on patient's carer. This aspect, however, is potentially being already included when assessing the quality of care delivery or patient's experience. On the other hand, the *operationality* of this element is problematic as there is not systematic data available.	Not applicable.
	25. Care continuum	Examples of this criterion include "Type of medical service" and "Type of therapeutic or clinical benefit". This dimension is part of the features of care delivery. Therefore, the 'care continuum' criterion fails to comply	Not applicable.

		with the *non-overlap* property, as it was already included (i.e. #2 criterion).	
	26. Patient lifestyle	Examples of this criterion include "does this development encourage patient self-care, or a potential reduction in health service use?", "Physical functioning", "Changes in Patient Lifestyle and Self-management Behavior: Smoking .measured as the percentage of patients that had successfully quit smoking at 12 months", "a condition/decision-specific Function(al) Index (e.g., six minute walk test)". In the context of new models of care and chronic conditions, lifestyle improvement is an important objective (187). There is, however, a potential risk of overlapping or double-counting with the criterion 'quality of life / well-being" (e.g., the EQ-5D-5L includes a dimension on mobility) or with some intermediate health outcomes (e.g., diabetes interventions measure patients' cholesterol and sugar levels).	Not applicable.
	27. Ethics	According to the literature review conducted, examples of this criterion include "ethical acceptability: experts' judgments via written questionnaires", "problems pertaining the full comprehension of the informed consent" and "moral significance of biomedical and biotechnological practices at the borders of right and not right concerning who you can hire, the rights of the person, the decisions entitled to the people, the decision criteria and values at stake". This criterion has rarely being used in the priority-setting context directly. Instead, some principles or rules have been directly applied when deciding on the analytical method to set priorities, particularly when designing health benefits packages on the path towards universal health coverage (e.g., unacceptable trade-off: To expand coverage for low- or medium-priority services before there is near universal coverage for high-priority services) (478, 479). In the context of new models of care, the whole prioritisation process should be ethically acceptable. This entails an impartial, transparent and clear process, based on a clear set of prioritisation criteria, in which all individuals are treated as equals (480). In addition, all interventions or new models of care under assessment should comply with regulation and guidelines issued by NHS England. In light if this, the 'ethics' criterion does not comply with the *non-redundancy* property.	Not applicable.
	28. Age/Gender	Examples of this criterion include "interventions targeting children (<5 years)" and "interventions targeting women of reproductive age" and "interventions targeting old people (>65 years)". This criterion has been used as a way to acknowledge the vulnerability of certain population groups. Despite of its relevance when making priority-setting decisions in healthcare, in the local commissioning context in England	Not applicable.

		this criterion is likely to fail the *non-redundancy* property as ICS budgets are tied to specific funding streams, defined by the NHS England. The allocation rules include, for instance, dedicated funds for maternity services, neonatal services or children's and young people's mental health (108).	

Appendix 4.B. Graphical representation of assessment and monitoring criteria

Figure 4.B.1. Assessment and monitoring criteria from the logic chain approach



Adapted from the framework for the MCDA of disease management programs, developed by Tsiachristas et al (2013) (187). Other sources used include (239, 269)

Appendix 4.C. Example of key performance indicators to operationalise the suggested monitoring criteria in the context of Early Intervention in Psychosis (EIP) services

Table 4.C.1. Potential key performance indicators in the context of EIP services for adults

Monitoring criteria	Suggested KPIs	Description
Intermediate health outcomes or quality of life outcomes (a)	KPI 11: Improved Medication Adherence	Improved Medication Adherence: % patients with effective medication adherence. It could be based on regular reviews and tools suggested by clinicians and NICE Guidelines.
	KPI 12: Symptom Reduction	Symptom Reduction: % patients that have reported a reduction in the severity or frequency of positive (e.g., hallucinations, delusions) and/or negative symptoms (e.g., apathy, anhedonia). It could be based on self-reports or regular assessments.

Quality of care (b)	KPI_21: Psychological therapy - CBTp	% of adult patients with psychosis or schizophrenia that have been offered cognitive behavioural therapy for psychosis (CBTp).
	KPI_22: Psychological therapy - family intervention	% of adult patients with psychosis or schizophrenia for which family members have been offered family intervention.
	KPI_23: Support for carers and families	% of carers of adults with psychosis or schizophrenia that have been offered carer-focused education and support programmes.
	KPI_24: Physical health and healthy lifestyles - behavioural	% of adults with psychosis or schizophrenia that have been offered combined healthy eating and physical activity programmes and helped to stop smoking.
	KPI_25: Physical health and healthy lifestyles - screening	% of adults with psychosis or schizophrenia that have had specific comprehensive physical health assessments.
	KPI_26: Medicines management	% of adults with schizophrenia who have not responded adequately to treatment with at least two antipsychotic drugs and have been offered clozapine.
	KPI_27: Education, employment and training	Adults with psychosis or schizophrenia who wish to find or return to work have been offered supported employment programmes.
Compliance with National standards (c)	KPI_31: EIP Waiting time standards	% of people experiencing first episode psychosis treated with a NICE-approved care package within two weeks of referral. 2024 NHS standard: At least 95% of people experiencing first episode psychosis should commence a NICE-recommended package of care within two weeks of referral. The standard is targeted at people aged 14 to 65 in line with NICE recommendations. Treatment will be deemed to have commenced when the person: (a) has had an initial assessment, AND (b) has been accepted on to the caseload of an EIP service capable of providing a full package of NICE-recommended care, AND, (c) has been allocated to and engaged with an EIP care co-ordinator.
Equity in quality of care (d)	KPI 41. Gender inequalities - CBT	Female - Men ratio in terms of the proportion of patients who were not offered CBTp
	KPI 42. Gender inequalities - Clozapine	Female - Men ratio in terms of the proportion of patients who were not offered clozapine after 2 unsuccessful trials of other antipsychotic medication
	KPI 43. Ethnicity inequalities - Carer education support programme	Others-White ratio in terms of carers of people with FEP who have not started a carer education support programme *Others include: Balck, Mixed, Asian, Others.
	KPI 44. Ethnicity inequalities - Physical health	Others-White ratio in terms of carers of people with FEP who have not receive all seven physical health screenings *Others include: Balck, Mixed, Asian, Others.
	KPI 45. Ethnicity inequalities - Outcome measures	Others-White ratio in terms of not having two or more outcome measurements recorded more than once *Others include: Balck, Mixed, Asian, Others.

(a) The EIP guidance and audit reports focus on quality of care outcomes, with few references to potential intermediate health outcome KPIs. The ones proposed here are based on the references listed below. However, they should be validated and defined in collaboration with clinicians.

-NICE (2015) Psychosis and schizophrenia in adults. Quality standard.12 February 2015. https://www.nice.org.uk/guidance/qs80/resources/psychosis-and-schizophrenia-in-adults-pdf-2098901855941

-Royal College of Psychiatrists (2018) Standards for Early Intervention in Psychosis Services - 1st Edition. Editors: Anita Chandra, Emily Patterson & Sophie Hodge https://www.rcpsych.ac.uk/docs/default-

source/improving-care/ccqi/quality-networks/early-intervention-in-psychosis-teams-(eipn)/epin-standards-first-edition.pdf?sfvrsn=fd9b4a0f_2/

-O'Connell, N., O'Connor, K., McGrath, D., Vagge, L., Mockler, D., Jennings, R., & Darker, C. D. (2022). Early Intervention in Psychosis services: A systematic review and narrative synthesis of the barriers and facilitators to implementation. European Psychiatry, 65(1), e2.

-McGorry, P. D., Killackey, E., & Yung, A. (2008). Early intervention in psychosis: concepts, evidence and future directions. World psychiatry, 7(3), 148.

(b) The suggested KPIs are part of the NHS 2023 guidance "Implementing the early intervention in psychosis access and waiting time standard". https://www.england.nhs.uk/wp-content/uploads/2023/03/B1954-implementing-the-early-intervention-in-psychosis-access-and-waiting-time-standard.pdf. These KPIs are in line with the National Institute for Health and Care Excellence (NICE) recommended package of EIP care for treating and managing psychosis (NICE Quality Standard, 2015).

(c) The Early Intervention in Psychosis Audit (2021/22) reports disparities in EIP care between different groups of people with FEP to highlight inequalities where there are significant differences and to guide EIP services in addressing them. However, this does not report an inequality measure, but rather quality indicators disaggregated by gender and ethnicity, with Pearson's chi-squared significance test (see Appendix F). In this dashboard, it is suggested to use ratios to compare an area's performance with a national average. This could be supplemented with the results of the Pearson's chi-squared significance test reported by the Early Intervention in Psychosis Audit. However, it is important to consider the health condition being analysed as well as the context. Not all variation is unjustified; some may be due to legitimate differences in patient needs, preferences or local service configurations.

Royal College of Psychiatrists (2022) National Clinical Audit of Psychosis – England National Report for the Early Intervention in Psychosis Audit 2021/2022. London: Healthcare Quality Improvement Partnership. Available from: www.rcpsych.ac.uk/NCAP

Appendices – Chapter 5

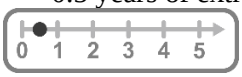
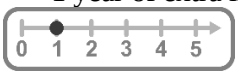
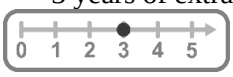
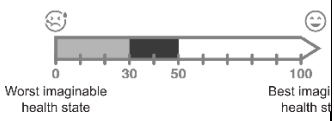
Appendix 5.A. Rationale and definition of attributes

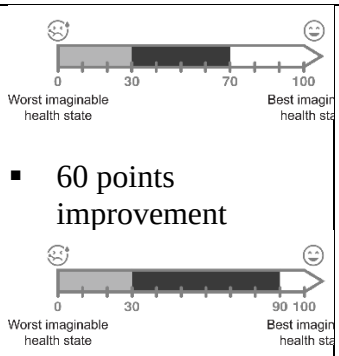
Attribute	Rationale	Definition
1. Additional years of life	To define which 'final or intermediate health outcome" to use, two systematic literature reviews were taken into account. The first one, conducted by Gu et al (2015) (309), reviewed studies that have elicited stated preferences from the public as to what factors should count in allocating publicly-funded health care. The second one, conducted by Trapero-Bertran and colleagues (2019) (481), identified attributes for designing a DCE in order to develop and validate a framework that supports decision-making on health technologies. In both reviews, '*health gains*' or '*improvements in health*' has been included in choice experiment studies, and has been framed in terms of years of life or QALYs gained. Since the second attribute of the descriptive framework of this study considers quality of life gains, it was decided to frame this first attribute in terms of additional years of life, following also similar studies on the field (290, 365, 482-486).	The number of additional years of life that each patient is expected to have with the care programme.
2. Quality of life improvements	Multiple DCE studies in the context of priority setting in healthcare, in high-income countries, have included '*quality of life and well-being considerations'* as one of the attributes (290, 473, 483, 486). The two systematic literature reviews mentioned earlier (309, 481) confirm the relevance of this attribute in the context of healthcare priority-setting.	Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression, due to the care programme.
3. Patient experience	Although patient experience has not been frequently included by choice experiment studies in the context of priority-setting (309, 481), local healthcare commissioners and stakeholders in England consider it one of the factors that should guide local investment decisions, particularly, in the context of integrated care (229). This goes in line with one of the main goals of new care models which is improving patient experience, via changes in the delivery of care (12, 88, 122, 187). The definition used was defined based on the literature (293) and the GP patient survey, run by Ipsos MORI on behalf of the CareQuality Commission and the NHS England (488, 4899).	How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, etc…
4. **Size of the target population**	The more people a healthcare intervention is covering, the better. Even though this attribute is not widely used in the context of integrated care, '*size*	The number of people who benefit directly from the

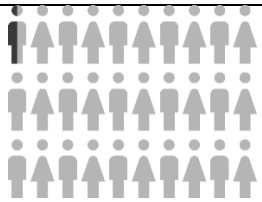
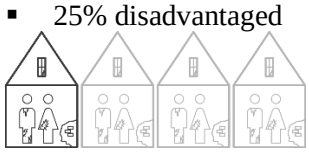
	of the population that receives that intervention' '*Number of potential beneficiaries*' or '*prevalence of a condition or disease*', was mentioned by some of the local commissioners and stakeholders during the semi-structured interviews (229), and has been widely used to guide priority-setting decisions in high-income countries (290, 345, 364, 473). These studies were taken into account to elaborate on the definition of this attribute.	care programme per 100,000 citizens.
5. Target population	Most of the stated-preference and MCDA studies, in the context of priority-setting, have framed the equity attribute in terms of whether the intervention is targeting vulnerable, marginalised or disadvantaged groups (32, 220, 276, 551, 468, 490-493). Mentzakis et al (2019), for instance, define it as the "*absence of systematic disparities in health between social groups*" (365). Bouwknegt et al (2018) refer to "*Impact on economically vulnerable communities".* Ezeife and coleagues (2020) use the wording "*Impact of the new treatment on the health of vulnerable or marginalized populations, compared to existing treatment*" (493). Linley and Hughes (2012) in the UK, acknowledge that the NHS gives special priority to improving the health of the most disadvantaged members of the population, particularly poorer people and ethnic minorities:, and in their choice experiment include the attribute "Disadvantaged populations, e.g. those from low income families" (495). Following this literature and the Core20PLUS5 approach adopted by the NHS England and NHS Improvement (294), the health equity attribute is framed in terms of the target population (i.e., Disadvantaged groups).	Proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, people from low-income households.
6. Additional budget required	Many stated-preferences and MCDA studies have framed the '*cost*' attribute in terms of the total health and social care cost per beneficiary for the health system, with 3 or 4 levels (149, 483, 496). Other authors have referred to the resources that must be mobilized to fund the intervention, and have expressed it in terms of co-payments or taxes increases (253, 255, 473, 484, 497). This option is particularly relevant in countries with a free publicly-funded healthcare system, and studies aiming at eliciting stated preferences of members of the general public. Using taxes increases as a the payment vehicle is likely to increase the cost attribute credibility, reducing the probability of attribute non-attendance (ANA) (310). Following this literature, and given that healthcare programmes (including new care models) are	Additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme.

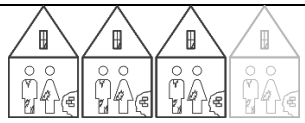
	funded by the NHS England, this attribute is framed in terms of income tax increases.	

Appendix 5.B. Attribute levels and rationale

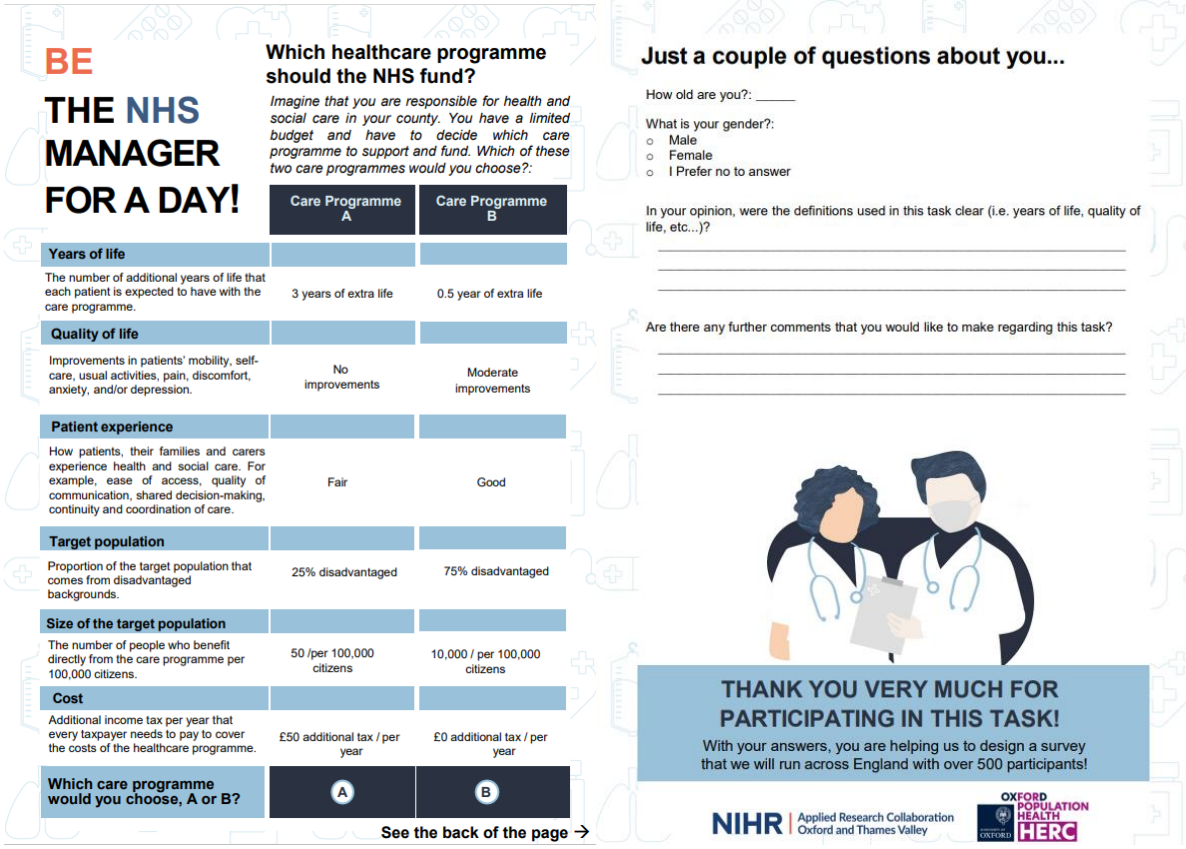
Attribute	Levels	Rationale
1. Additional years of life *Numerical*	▪ 0.5 years of extra life  ▪ 1 year of extra life  ▪ 3 years of extra life  	Studies aiming at measuring relative public preferences for guiding priority-setting decisions have used three or four levels i.e. two extremes and one or two as an intermediate central level. Mentzakis et al (2011) for instance, in the context of drugs to treat orphan diseases, used the attribute "*The impact of drug treatment on a patient's health/life-years gained (LYG)*" with the following levels: 15 years; 10 years; 5 years; 1 year (498). Similarly, Lopez-Bastida and colleagues (2019) (484) included "*Improvement in health*" as one of the attributes, and 4 levels (Large; Moderate; Small; Very small). Even though various studies have treated this attribute as categorical, using the wording "*small, medium, large*" (484, 499), it was decided to treat it as a numerical attribute in order to estimate marginal rates of substitution. For the graphical representation of the levels, Rivero-Arias et al (2022) (297) was followed. Evidence up to date suggests that new care models can potentially improve patient satisfaction, access, and healthcare quality (237, 500, 501). There is however limited evidence on the impact of integrated care on health outcomes (i.e. life expectancy, QALYs) and cost, and the few studies available suggest modest positive effects (55, 145, 502, 503). For instance, a systematic literature review and a meta-analysis on the impact on outcomes of integrated care interventions found a 6.2% statistically significant improvement in outcomes associated with integrated care (overall ratio of mean outcomes: 1.062 (95% CI 1.048–1.077)). Nevertheless, the heterogeneity in the ratio of mean outcome across the studies was similarly large (I 2=99.2%) (145). Based on this literature and following good practices (284, 299), three levels were defined.
2. Quality of life improvements *Numerical*	▪ 20 points improvement  ▪ 40 points improvement	To define the levels for this attribute, different options based on previous studies were considered. Rutten-van Mölken, (2018) (149) disaggregated it into two categorical attributes (Physical functioning and psychological well-being) and used three levels on each one of them. Hampson et al (2019) referred to "*increase in Quality of Life with New Treatment*" and used four levels: 1. "no gain", 2. "small (25% gain)", 3. "large (50% gain)" and 4. "very large (75% gain)". Morell and colleagues (2021) (291) referred to "*health improvements with the new treatment*" and used four numerical levels (i.e., 0.5, 1, 5, 10), though the variable was treated as categorical. Nicolet et al (2020) also used a numerical approach for the attribute "patients' related quality of life (HRQoL) change after new

	 ▪ 60 points improvement	treatment", with three levels (-0.2, 0.1 and 0). Rivero-Arias et al (2022) (297), in a stated-preference study aiming at estimating the value of different elements of the outcomes for a child with a surgical condition in the UK, used as one of the attributes "*Child's quality of life*" with three levels: "*Good quality of life*", "*Fair quality of life*" and "*Poor quality of life*". Other studies conducted in high-income countries have also used 3 or 4 levels and a similar approach (483, 484, 491, 494). As this is considered a numerical attribute, the EQ-5D-5L questionnaire (English version for the UK) was used in the definition. Three levels are used to indicated gains in quality of life (which has been done by various of the studies mentioned above). For the pictorial representation, the Visual Analog Scale (EQ-VAS) was used, where health status is represented on a scale of 0–1000 (261, 505). For all levels, the initial health status is set in 30, and gains are presented using a darker shade.
3. Patient experience *Categorical*	▪ Poor  ▪ Fair  ▪ Good 	Relatively few studies have explored public or patient preferences for integrated care, and the few studies published have decomposed the concept of patient experience into two or three attributes. Huppelschoten et al (2014), for instance, explored patient's willing to pay for different aspects of fertility care, with a particular focus on patient centredness, using "*Information provision*", "*Patient involvement*" and "*Continuity of physicians*" as attributes in the experiment (667). Nicolet et al (2022) studied preferences for coordinated care in Switzerland, and use 3 attributes to describe patient experiences of integrated care: 1. Professionals that have access to electronic patient data; 2. Care coordination; and 3. Access to specialist (354). Rutten-van Mölken, et al (2018) added as attributes "*Continuity of care*" and "*Person centredness*", and for each one used three levels with the working "poor", "fair" and "good" (293). None of the studies have directly used a unique "patient experience" attribute. The Picker Institute Europe proposed a set of 18 questions to measure people's experiences of integrated care in health and social care (506), but none of the questions refers to the overall patient experience. On the contrary, the GP patient survey, run by Ipsos MORI on behalf of the CareQuality Commission and NHS England, includes the question "*Overall, how would you describe your experience of your GP practice?*" with five answer options (Very good, Fairly good, Neither good nor poor, Fairly poor and Very poor) (488, 489). For the wording of the levels and the pictorials, the NHS patient survey was followed, and the attribute is treated as categorical, using only three levels (293).
4. Size of target population *Numerical*	▪ 50 / per 100,000 citizens	In previous DCE studies (345, 364, 365), authors have referred to '*Number of potential beneficiaries*' using two levels (Few & Many). Jehu-Appiah and colleagues (2008) adopted a similar wording, but with three levels (Few, Average and – Many) (172). More recent studies, conducted in high-income

	 ▪ 5,000 / per 100,000 citizens  ▪ 10,000 / per 100,000 citizens 	countries have used values directly. Hampson et al (2019), for instance, explores public preferences for health gains and cures in the UK, and for the attribute '*Number of Patients Treated*' uses four levels: 100, 200, 400 and 800 (290). Pauwels et al (2019) in Belgium, when conducting a choice experiment to elicit the trade-offs in cancer medicine characteristics, used the attribute '*Number of Patients eligible for treatment*', with three levels: 100, 5,000 and 10,000 (473). The levels of this attribute are based on this literature and on three different types of health conditions in England: ▪ Non-prevalent health conditions like rare diseases: In England, a rare disease is defined as a condition that affects less than 1 in 2,000 people (507). This is equivalent to 50 per 100,000 citizens. ▪ One of the most common chronic diseases in the UK, with its prevalence increasing, such as diabetes. In England, in 2018–2019, there were 3,319,266 people diagnosed with diabetes (508). This is equivalent to around 5,000 per 100,000 citizens. Diabetes is one of the top 10 causes of death and disability (DALYs) in England, and the condition that registered the largest increase in age-standardised DALY rates, among the top 10 causes of death and disability combined, between 2009 and 2019 (i.e. 32.5%) (511). ▪ Highly prevalent condition/disease: According to the Global Burden of Disease, the top three causes of death and disability (DALYs) in England are low back pain, ischemic heart disease, and depressive disorders (511). It is estimated that in England 6.4 million people live with circulatory diseases (467). This is equivalent to between 10,000 and 11,000 people per 100,000 citizens. Similarly, it is estimated that more than 9% of the population have generalised anxiety disorder or depressive episodes, which is also equivalent to around 10,000 per 100,000 citizens (421, 466). The levels used for this attribute are similar to the ones used by Pauwels et al (2019) (473). The darker shade of the pictorial indicates the size of the population covered by the care programme, per 100,000 citizens.
5. Target population *Categorical*	▪ 25% disadvantaged  ▪ 50% disadvantaged  ▪ 75% disadvantaged	Not many choice experiments in the context of priority setting have included an equity attribute (309, 481). Those that have done it, however, have defined 'equity' in terms of the socio-economic status or characteristics of the beneficiaries of the healthcare intervention, with two, three or four levels. McNamara et al (2021) in the UK, for instance, conducted a series of pairwise person-trade-off exercises to elicit participants' levels of health inequality aversion. The programmes from which participants have to choose can help "*people from the poorest fifth of society*" or "*people from the richest fifth of society*" (494). Lim et al (2012), aiming at eliciting public preference for health-care resource allocation in South Korea, added "patient's household income level" as one of the attributes and used four levels: High (the highest fourth income quartile); Mid-high (the third income quartile);

		Mid-low (the second income quartile) and Low (the lowest income quartile) (491). Recently, the NHS England and NHS Improvement launched a policy strategy or approach called Core20PLUS5 aiming at tackling socioeconomic health inequalities. This approach defines the most deprived 20% of the national population (as identified by the national Index of Multiple Deprivation - IMD), as the target population cohort (294). Following the studies mentioned above and the Core20PLUS5 approach adopted by the NHS England and NHS Improvement (294), the levels of this health equity attribute are framed in terms of the target population. As it is desirable to treat this attribute as continuous, three equidistance attribute levels are used, together with a pictorial indicating the proportion of the target population that comes from disadvantaged or vulnerable backgrounds.
6. Additional budget required *Numerical*	▪ £20 additional tax / per year  ▪ £40 additional tax / per year  ▪ £60 additional tax / per year 	Multiple studies have framed the 'cost' attribute in terms of the total health and social care cost per beneficiary for the health system, with three or four levels. Rutten-van Mölken et al (2018), for instance, used £2,400, £ 3,000 and £3,600 when measuring relative preferences for outcomes of integrated care of patients with multi-morbidity in the UK (149). Bourke et al (2018) used a similar approach when studying societal preferences for funding orphan drugs in the UK (483). A study conducted in 2019 in the UK, to assess the preferences for cervical screening, also used this option (i.e., Cost of your test to the NHS: £8; £20; £25; £40) (496). Fewer authors have adopted a budget impact definition in percentage terms (441). Other studies have referred to the resources that must be mobilized to fund the intervention, and have expressed it in terms of co-payments or taxes increases (354, 473, 484, 485). As most of the DCE studies on priority-setting in healthcare have been conducted in high-income countries (253, 255), the majority of the studies in which members of the public have responded the DCE surveys, have used this approach to frame the "cost" attribute (253, 481). Following this literature, and given that healthcare programmes (including new models of care) are funded by the NHS, three equidistance attribute levels in pounds are used to refer to the tax increases, together with a pictorial and an arrow proportional to the size of the tax increase.

Appendix 5.C. BRC event Pre test - DCE task



BE THE NHS MANAGER FOR A DAY!

Which healthcare programme should the NHS fund?

Imagine that you are responsible for health and social care in your county. You have a limited budget and have to decide which care programme to support and fund. Which of these two care programmes would you choose?:

	Care Programme A	Care Programme B
Years of life The number of additional years of life that each patient is expected to have with the care programme.	3 years of extra life	0.5 year of extra life
Quality of life Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression.	No improvements	Moderate improvements
Patient experience How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, shared decision-making, continuity and coordination of care.	Fair	Good
Target population Proportion of the target population that comes from disadvantaged backgrounds.	25% disadvantaged	75% disadvantaged
Size of the target population The number of people who benefit directly from the care programme per 100,000 citizens.	50 /per 100,000 citizens	10,000 / per 100,000 citizens
Cost Additional income tax per year that every taxpayer needs to pay to cover the costs of the healthcare programme.	£50 additional tax / per year	£0 additional tax / per year
Which care programme would you choose, A or B?	A	B

See the back of the page →

Just a couple of questions about you...

How old are you?: _____

What is your gender?:
- Male
- Female
- I Prefer no to answer

In your opinion, were the definitions used in this task clear (i.e. years of life, quality of life, etc...)?

_____
_____
_____

Are there any further comments that you would like to make regarding this task?

_____
_____
_____

THANK YOU VERY MUCH FOR PARTICIPATING IN THIS TASK!

With your answers, you are helping us to design a survey that we will run across England with over 500 participants!

NIHR | Applied Research Collaboration Oxford and Thames Valley

OXFORD POPULATION HEALTH HERC

Appendix 5.D. Pilot study

Pilot 1:

For the experimental design it is assumed that people's preferences would be driven by "Additional years of life" and "Quality of life improvements". Hence, two D-efficient designs were created using a partial profile with 8 rows and priors close to zero, and used constraints to avoid choice tasks in which "Additional years of life", "Quality of life improvements" and "Size of target population" were together displaying the highest or medium levels in one of the alternatives (e.g., reject: progB.yol=3 and progB.qol=6). The D-error for the "AddTax" design was 0.005871 and for the "NoAddTax" design it was equal to 0.007394.

During this first pilot responses from 43 people across England were collected, between the 8th of August and the 5th of September 2022. As a quality control check, a minimum time threshold was used by Dynata to flag and exclude 'speeders' or respondents who completed the survey too quickly (i.e., less than 1/3 of the median time). Each respondent answered two practice questions (one of them being the dominant test), and nine choice tasks. The ninth choice task was used as a stability test, and as such, it was equal to the second choice task. To reduce the burden of complexity, colour coding was used, following Jonker et al (2019) (302). Choice questions from the "NoAddTax" Design, with the 5 attributes, were completed by 22 people, while the choice questions from "AddTax" Design (i.e., with 6 attributes) were completed by 21 people. On average, people took 21 minutes to complete the entire

questionnaire (median: 9.6; min:3.3; max: 337; SD: 51). All participants responded as expected to the dominant test (i.e., all chose the strongly dominant alternative), and 77% passed the stability test.

Results from the multinomial logit (MNL) models are shown in **Table 5.D.1.** Model 1 is the MNL model with 6 attributes (i.e., "AddTax" Design) and Model 2 is the MNL model with 5 attributes (i.e., "NoAddTax" Design). Although in Model 1 the direction of the estimated coefficients is in line with prior expectations, results suggest that "*patient experience*" is the only relevant (i.e. statistically significant) attribute for respondents. The direction of "*quality of life improvements*" in Model 2 is negative which is contrary to prior expectations. In both models the magnitude of "*patient experience*" coefficients (bexp_fair, bexp_good) are the largest. All in all, these results suggest that, contrary to the prior belief, "*additional years of life*" and "*quality of life improvements*" do not seem to be driving choices significantly. As can be observed in **Table 5.D.2.,** when analysing responses of choice sets in which the level of "*quality of life improvements*" differs between two alternatives or care programmes, in only two out of eight choice sets (i.e., choice task 1 for "AddTax" Design and choice task 5 in the "No AddTax" Design), the majority of participants chose the care programme with the relative higher level on "*quality of life improvements*". In light of this, it was not possible to use results from this pilot. Therefore, a second pilot with a more simplified experimental design that excludes the constraints was conducted.

Table 5.D.1. Multinomial logit (MNL) choice models

	Model 1. AddTax Design			**Model 2. NoAddTax Design**		
Variable	**Coefficient**	**Robust s.e.***	**t-ratio**	**Coefficient**	**Robust s.e.***	**t-ratio**
asc	0.077	0.207	0.372	0.513	0.212	2.417
Additional years of life - YoL	0.148	0.170	8.869	0.118	0.142	0.834
Quality of life improv. - QoL	0.030	0.115	0.264	-0.103	0.122	-0.841
Patient experience - Exp						
Fair	1.387	0.326	4.250	1.163	0.390	2.984
Good	1.939	0.378	5.134	1.841	0.471	3.908
Size of target population - Size	0.000	0.000	0.944	0.000	0.000	1.291
Target population - Equ	0.008	0.005	1.514	0.003	0.002	0.884
Add. budget - AddTax	-0.031	0.007	-4.486			
Log likelihood (final)	-97.87			-96.26		
AIC	211.74			206.52		
BIC	237.67			229.54		

*s.e., standard error.

Table 5.D.2. Analysis of responses – Choice tasks in which "Quality of life improvements" differ between Care Programme A and Care programme B

Block	Choice set	Care programme A (i.e. alternative 1)						Care programme B (i.e. alternative 2)						A	B
		YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax	%	%
1	1	0.5	2	0	5000	25	40	0.5	6	1	50	75	20	9.5%	90.5%
1	2	0.5	2	1	10000	50	60	0.5	2	2	5000	25	20	23.8%	76.2%
1	3	3	2	1	50	25	20	0.5	2	0	10000	75	60	90.5%	9.5%
1	4	0.5	2	1	5000	75	20	0.5	4	2	50	25	40	66.7%	33.3%
1	5	0.5	6	0	50	25	20	0.5	2	2	10000	25	60	28.6%	71.4%
1	6	1	2	2	50	75	40	0.5	2	1	5000	25	20	57.1%	42.9%
1	7	0.5	2	2	10000	25	20	1	2	1	50	75	60	85.7%	14.3%
1	8	0.5	4	1	50	25	60	3	2	0	50	75	20	38.1%	61.9%
1	9	0.5	2	1	10000	50	60	0.5	2	2	5000	25	20	19.0%	81.0%
2	10	0.5	6	0	50	75	0	0.5	2	1	10000	25	0	18.2%	81.8%
2	11	3	2	0	50	25	0	0.5	2	2	10000	75	0	9.1%	90.9%
2	12	0.5	2	0	10000	75	0	1	2	1	50	25	0	50.0%	50.0%
2	13	0.5	2	0	5000	25	0	3	2	2	50	75	0	18.2%	81.8%
2	14	0.5	4	1	50	75	0	0.5	2	0	5000	25	0	81.8%	18.2%
2	15	1	2	2	50	75	0	0.5	6	1	50	25	0	86.4%	13.6%
2	16	0.5	2	1	5000	75	0	0.5	4	2	50	25	0	59.1%	40.9%
2	17	0.5	2	2	10000	25	0	0.5	2	0	5000	75	0	90.9%	9.1%
2	18	3	2	0	50	25	0	0.5	2	2	10000	75	0	22.7%	77.3%

Pilot 2:

For this pilot, two D-efficient designs assuming a multinomial logit model were created in Ngene (303). As in Pilot 1, colour coding and attribute level overlap were used (302). For the later, a partial profile designs was created, with the candidate sets created in Stata. The first design included the five attributes ('NoAddTax' Design) and a second one six attributes ('AddTax' Design, in which the attribute "*additional budget required*" is included). For each design, 24 rows and 2 blocks were used. In 'AddTax' Design, all attributes but "*Patient experience*" were treated as numerical. In the 'NoAddTax' Design, "*Patient experience*", "*Quality of life improvements*" & "*Additional years of life*" were treated as categorical, and the other two attributes as numerical to guarantee more than one choice sets with overlap on "*Patient experience*"[64]. Attribute level constraints on "*Size of the target population*" and "*Target population*" were imposed to guarantee a degree of attribute level balance[65] (303).

By the time the pilot study was designed, no theoretically relevant interaction effects between attributes were considered to be relevant. Previous similar studies only included interactions with socio-demographics in the estimation phase (293, 364, 431, 345, 484, 490)[66].

Both designs had a low D-error (0.001044 and 0.021268, respectively), and a relatively even distribution of movements between levels for each attribute. Forty (40) respondents (i.e., 10% of the sample size[67]) were randomly allocated across the two pilot designs, with 20 respondents assigned to 'NoAddTax' Design (with 5 attributes) and the remaining 20 assigned to complete the choice sets from pilot 'AddTax' Design (with 6 attributes).

[64] While the design treating as numerical all attributes but "*patient experience*" was considered, only one choice set (out of 24) with overlap in "*patient experience*". After running more than 20 potential designs, using different restrictions when constructing the candidate set, only a design in which "*Patient experience*", "*Quality of life improvements*" & "*Additional years of life*" were treated as categorical would offer more than one choice set with an overlap on ""*Patient experience*". A design in which all attributes were treated as categorical was also considered. Although with such design a more evenly distribution of movements between levels for each attribute was observed, and with this, nine choice sets (out of 24) with overlap on "*Patient experience*", it was decided to use the design in which only three of the attributes (out of five) are treated as categorical; this design is closer to the one to be used in the modelling exercise. The D-error of the design that treats all attributes as categorical is higher (D-error=0.50961) than the one obtained when only three attributes are treated as categorical (D-error= 0.021268). Nevertheless, the D-error of a design where the model has categorical variables cannot be compared to the D-error of a design where the model has numerical variables. This is because the D-error is based on the covariance matrix of the parameter estimates. The covariance matrix with effects coding is entirely different than the covariance matrix with numerical attributes (e.g., with effects coding the number of parameters is much larger and the attribute values are -1, 0, or 1). The D-error is the determinant of the covariance matrix, and if the covariance matrices are entirely different, they should not be compared.

[65] Ngene does not use often the middle levels of numerical attributes since this is less efficient (making larger trade-offs between extreme levels provides more information).

[66] The few papers that add interactions or suggest that future studies should consider them, is because: (a) They include as attributes increase in 'quality of life' or 'years of life' with current care and with new treatment separately or (b) They included as attributes characteristics of the beneficiary population, and interact them with health gains.

[67] This corresponds to 10% of the preliminary sample size, calculated using a rule of thumb.

Table 5.D.3. "Add Tax" Pilot Design

		Care programme A (i.e. alternative 1)						**Care programme B (i.e. alternative 2)**						**Trading attributes & overlaps**					
Block	Choice set	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
1.1	1	1	20	2	50	25	60	0.5	60	1	50	25	20	1	1	1	0	0	1
1.1	4	3	20	0	10000	75	60	0.5	20	1	50	75	20	1	0	1	1	0	1
1.1	7	1	20	2	50	75	40	1	20	0	10000	50	20	0	0	1	1	1	1
1.1	8	0.5	60	2	50	25	20	3	20	2	10000	25	40	1	1	0	1	0	1
1.1	9	3	60	0	50	75	40	0.5	20	1	10000	75	40	1	1	1	1	0	0
1.1	12	3	40	1	50	75	40	3	60	0	5000	75	60	0	1	1	1	0	1
1.1	13	3	40	1	10000	50	40	1	60	1	10000	75	60	1	1	0	0	1	1
1.1	14	1	20	1	10000	75	20	1	40	2	5000	25	20	0	1	1	1	1	0
1.1	15	1	60	0	10000	75	40	3	20	2	10000	50	40	1	1	1	0	1	0
1.1	16	3	60	0	10000	25	20	3	20	1	10000	75	60	0	1	1	0	1	1
1.1	17	0.5	60	0	50	75	20	0.5	60	1	5000	25	60	0	0	1	1	1	1
1.1	23	0.5	60	2	10000	50	40	3	40	2	50	75	40	1	1	0	1	1	0
1.2	2	0.5	60	2	50	75	40	1	20	0	50	75	20	1	1	1	0	0	1
1.2	3	0.5	20	2	50	75	20	3	60	1	50	25	20	1	1	1	0	1	0
1.2	5	1	20	1	50	75	60	0.5	20	0	10000	75	20	1	0	1	1	0	1
1.2	6	0.5	40	2	10000	75	20	1	60	0	10000	75	60	1	1	1	0	0	1
1.2	10	3	20	1	50	25	40	0.5	60	1	50	75	60	1	1	0	0	1	1
1.2	11	3	40	0	50	25	20	1	40	2	10000	25	60	1	0	1	1	0	1
1.2	18	0.5	60	0	10000	50	60	3	40	0	50	75	60	1	1	0	1	1	0
1.2	19	0.5	40	2	50	25	40	3	20	2	50	75	60	1	1	0	0	1	1
1.2	20	1	60	0	10000	75	20	3	20	2	10000	25	20	1	1	1	0	1	0

1.2	21	3	20	1	10000	75	40	0.5	20	2	10000	50	20	1	0	1	0	1	1
1.2	22	1	60	1	10000	50	60	1	40	2	10000	25	40	0	1	1	0	1	1
1.2	24	3	60	0	50	75	20	3	20	2	5000	50	20	0	1	1	1	1	0
Number of times each participant is trading on the attribute --->														**17**	**18**	**18**	**12**	**15**	**16**
Number of times there is overlap on the attribute --->														**7**	**6**	**6**	**12**	**9**	**8**

D error	0.001044
A error	0.217871

Table 5.D.4. "NoAdd Tax" Pilot Design:

		Care programme A (i.e. alternative 1)						**Care programme B (i.e. alternative 2)**						**Trading attributes & overlaps**					
Block	Choice situation	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
2.1	1	1	0	0	50	25	.	0	0	0	10000	75	.	1	0	0	1	1	.
2.1	2	0	1	2	5000	25	.	0	0	2	10000	75	.	0	1	0	1	1	.
2.1	6	0	1	1	10000	50	.	1	1	2	5000	50	.	1	0	1	1	0	.
2.1	7	1	2	0	50	75	.	0	0	0	5000	75	.	1	1	0	1	0	.
2.1	8	2	0	1	5000	50	.	1	0	1	10000	25	.	1	0	0	1	1	.
2.1	9	0	0	0	5000	75	.	0	0	1	50	25	.	0	0	1	1	1	.
2.1	14	0	2	2	50	50	.	2	2	0	5000	50	.	1	0	1	1	0	.
2.1	17	2	1	1	10000	25	.	0	2	1	50	25	.	1	1	0	1	0	.
2.1	18	0	1	0	50	50	.	2	0	0	50	75	.	1	1	0	0	1	.
2.1	19	0	0	0	5000	75	.	0	0	1	50	50	.	0	0	1	1	1	.
2.1	23	2	0	2	10000	75	.	2	1	1	10000	50	.	0	1	1	0	1	.

2.1	24	2	2	0	10000	25	.	1	2	1	5000	25	.	1	0	1	1	0	.
2.2	3	0	2	1	5000	25	.	0	0	0	10000	25	.	0	1	1	1	0	.
2.2	4	1	0	1	50	25	.	0	0	1	10000	50	.	1	0	0	1	1	.
2.2	5	2	0	1	5000	25	.	1	0	0	5000	50	.	1	0	1	0	1	.
2.2	10	0	0	1	10000	75	.	0	1	2	50	75	.	0	1	1	1	0	.
2.2	11	2	2	0	10000	50	.	2	1	2	10000	75	.	0	1	1	0	1	.
2.2	12	0	2	0	5000	75	.	2	2	1	5000	50	.	1	0	1	0	1	.
2.2	13	0	1	0	50	50	.	0	2	0	5000	25	.	0	1	0	1	1	.
2.2	15	1	2	0	10000	75	.	2	0	0	50	75	.	1	1	0	1	0	.
2.2	16	1	0	2	10000	25	.	1	1	0	50	25	.	0	1	1	1	0	.
2.2	20	0	0	0	5000	50	.	0	2	0	50	75	.	0	1	0	1	1	.
2.2	21	1	0	2	50	50	.	0	2	2	5000	50	.	1	1	0	1	0	.
2.2	22	1	1	1	50	75	.	1	1	2	10000	25	.	0	0	1	1	1	.
Number of times each participant is trading on the attribute --->														**13**	**13**	**13**	**19**	**14**	
Number of times there is overlap on the attribute --->														**11**	**11**	**11**	**5**	**10**	

D error:	*0.021268*
A error:	*0.553724*

Before running this second pilot, some exploratory analyses were conducted, simulating different patterns of behaviour to test both models (e.g., complete random responses; Strong preferences towards 'Additional years of life' and 'Quality of life improvements'). When simulating the dependent variable (i.e., the chosen option) full compensatory behaviour was not assumed (i.e., 90% A and 10% B was assumed when the attribute is best for A and 90% B and 10% A when the attribute is best for A). The models were estimated treating variables as indicated initially (i.e., all as numerical, except for 'patient experience') and treating all them as categorical. In all models estimated expected results in terms of the direction of the beta estimates were obtained. An example of one of the models estimated is presented below.

Table 5.D.5. NoAddTax pilot Design: MNL results simulating complete random responses

Variable	Coefficient	Robust s.e.*	t-ratio
asc	-0.029	0.0433	-0.656
Additional years of life - YoL	-0.022	0.032	-0.680
Quality of life improv. - QoL	0.000	0.002	-0.304
Patient experience - Exp			
Fair	0.031	0.077	0.406
Good	-0.013	0.085	-0.156
Size of target population - Size	0.000	0.000	-1.638
Target population - Equ	0.002	0.002	1.445
Log likelihood (final)		-1660.8	
AIC		3335.59	
BIC		3376.08	

*s.e., standard error.

Results of the pilot are presented in the tables below. All coefficients have the expected sign and are statistically significant. Results from this pilot are used as priors for the full launch experimental designs.

Table 5.D.6. Respondents characteristics from pilot study (n=43)

		n	%	Quotas
Gender				
	Male	22	51.2%	49%
	Female	20	46.5%	51%
	Other	1	2.3%	
Age				
	Less than 18	0	0.0%	0%
	18-24 years	4	9.3%	11%
	25-34 years	7	16.3%	17%
	35-44 years	5	11.6%	16%
	45-54 years	11	25.6%	17%
	55-64 years	6	14.0%	16%
	65+ years	10	23.3%	24%
Region				
	South East	5	11.6%	16%

	South West	4	9.3%	10%
	London	6	14.0%	16%
	East of England	6	14.0%	11%
	West Midlands	6	14.0%	10%
	East Midlands	3	7.0%	9%
	Yorkshire and the Humber	6	14.0%	10%
	North West	5	11.6%	13%
	North East	2	4.7%	5%
Ethnic group				
	White	35	81.4%	84%
	Mixed or multiple ethnic groups	2	4.7%	2%
	Asian or Asian British	3	7.0%	8%
	Black, Black British, Caribbean or African	2	4.7%	4%
	Other ethnic group	1	2.3%	2%
Employment of main income earner in the household				
	Managerial roles, administrative or professional	12	27.9%	27%
	Supervisory or clerical and junior managerial roles, administrative or professional	10	23.3%	28%
	Skilled manual workers	8	18.6%	20%
	Semi-skilled, unskilled manual workers, state pensioners, casual and lowest grade workers, unemployed with state benefits only.	13	30.2%	25%

Table 5.D.7. "AddTax" and "NoAddTax" pilot Design: MNL results

	AddTax Design			**NoAddTax Design****		
Variable	**Coefficient**	**Robust s.e.***	**t-ratio**	**Coefficient**	**Robust s.e.***	**t-ratio**
asc	0.088	0.136	0.642	-0.115	0.116	-0.995
Additional years of life - YoL	0.326	0.146	2.233			
1 year of extra life				0.530	0.346	1.533
3 years of extra life				1.625	0.392	4.149
Quality of life improv. - QoL	0.012	0.008	1.471			
40 points improvement				1.169	0.343	3.409
60 points improvement				1.782	0.599	2.977
Patient experience - Exp						
Fair	0.608	0.235	2.585	0.749	0.412	1.818
Good	0.895	0.339	2.642	1.956	0.714	2.738
Size of target population - Size	0.000	0.000	3.214	0.000	0.000	2.527
Target population - Equ	0.004	0.005	0.815	0.002	0.006	0.277
Add. budget needed - AddTax	-0.013	0.008	-1.679			
Log likelihood (final)	-161.03			-135.21		
AIC	338.05			288.41		
BIC	366.29			320.6		

*s.e., standard error.; **YoL, QoL and Exp treated as categorical attributes.

Appendix 5.E. Experimental design (full launch)

AddTax Design: 6 attributes (i.e., 'Additional budget required' included)

D-efficient partial profile design with 28 rows (2 blocks; 14 choices per participant) and priors close to zero. The model averaging approach was used in this design: The first model (M1) was a Bayesian D-Efficient (using priors from pilot #2), and the second model (M2) a D-efficient design with interactions and priors close to zero. The second model included interaction between "Exp" and "QoL" and between "AddTax" and all other attributes (except for 'Equ').

A partial profile design with candidate set constructed in Stata was used to ensure overlap in two attributes for all choice sets (i.e., each choice set includes an overlap on two attributes). All attributes were treated as continuous, but "Exp". Candidate set excludes potentially dominated choice sets and for that, restrictions were used when constructing candidate set (size of candidate set: 65,610 rows). The following choice sets were rejected when constructing the candidate set:

```
//avoid only difference in yol
drop if qolA>=qolB & expA>=expB & sizeA>=sizeB & equA>=equB & addtaxA<=addtaxB
drop if qolA<=qolB & expA<=expB & sizeA<=sizeB & equA<=equB & addtaxA>=addtaxB

//avoid only difference in qol
drop if yolA>=yolB & expA>=expB & sizeA>=sizeB & equA>=equB & addtaxA<=addtaxB
drop if yolA<=yolB & expA<=expB & sizeA<=sizeB & equA<=equB & addtaxA>=addtaxB

//avoid only difference in exp
drop if yolA>=yolB & qolA>=qolB & sizeA>=sizeB & equA>=equB & addtaxA<=addtaxB
drop if yolA<=yolB & qolA<=qolB & sizeA<=sizeB & equA<=equB & addtaxA>=addtaxB

//avoid only difference in size
drop if yolA>=yolB & qolA>=qolB & expA>=expB & equA>=equB & addtaxA<=addtaxB
drop if yolA<=yolB & qolA<=qolB & expA<=expB & equA<=equB & addtaxA>=addtaxB

//avoid only difference in equ
drop if yolA>=yolB & qolA>=qolB & expA>=expB & sizeA>=sizeB & addtaxA<=addtaxB
drop if yolA<=yolB & qolA<=qolB & expA<=expB & sizeA<=sizeB & addtaxA>=addtaxB

//avoid only difference in cost
drop if yolA>=yolB & qolA>=qolB & expA>=expB & sizeA>=sizeB & equA>=equB
drop if yolA<=yolB & qolA<=qolB & expA<=expB & sizeA<=sizeB & equA<=equB
```

Table 5.E.1. Experimental AddTax Design

Block	Choice set	Care programme A (i.e. alternative 1)						Care programme B (i.e. alternative 2)						Trading attributes & overlaps					
		YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
1	1	0.5	40	0	10000	75	60	0.5	20	2	10000	50	40	0	1	1	0	1	1
1	2	3	40	0	5000	50	60	1	60	1	50	50	60	1	1	1	1	0	0
1	4	0.5	20	2	10000	50	60	1	40	2	5000	25	60	1	1	0	1	1	0
1	7	0.5	40	1	10000	50	40	1	40	2	5000	25	40	1	0	1	1	1	0
1	10	1	60	1	50	50	40	1	60	2	10000	25	60	0	0	1	1	1	1
1	12	1	60	1	5000	50	60	3	60	0	5000	25	20	1	0	1	0	1	1
1	14	1	60	0	50	50	40	0.5	40	2	50	75	40	1	1	1	0	1	0
1	15	3	20	0	50	75	60	0.5	20	1	50	25	20	1	0	1	0	1	1
1	17	3	60	2	50	75	60	3	20	0	10000	75	40	0	1	1	1	0	1
1	21	0.5	60	1	5000	50	40	1	60	0	50	75	40	1	0	1	1	1	0
1	23	3	20	0	10000	75	20	1	60	1	10000	75	60	1	1	1	0	0	1
1	25	0.5	20	1	10000	75	40	0.5	60	1	5000	50	20	0	1	0	1	1	1
1	26	1	60	0	50	50	40	0.5	40	0	5000	50	20	1	1	0	1	0	1
1	27	3	20	2	5000	25	40	3	60	0	10000	25	60	0	1	1	1	0	1
2	3	3	60	0	10000	75	60	3	60	2	5000	25	40	0	0	1	1	1	1
2	5	3	40	1	5000	50	60	0.5	60	0	5000	50	20	1	1	1	0	0	1
2	6	0.5	20	2	5000	50	40	0.5	60	0	10000	50	60	0	1	1	1	0	1
2	8	1	60	0	10000	75	60	1	40	1	50	75	40	0	1	1	1	0	1

2	9	0.5	60	0	5000	75	40	1	20	1	5000	25	40	1	1	1	0	1	0
2	11	0.5	40	1	50	75	60	1	20	0	50	75	20	1	1	1	0	0	1
2	13	1	60	2	5000	25	60	1	40	1	10000	25	20	0	1	1	1	0	1
2	16	0.5	20	1	10000	50	60	1	40	1	5000	25	60	1	1	0	1	1	0
2	18	3	20	2	50	25	40	1	40	2	50	75	60	1	1	0	0	1	1
2	19	0.5	60	0	50	75	20	0.5	60	2	10000	50	60	0	0	1	1	1	1
2	20	1	20	1	50	75	60	0.5	60	1	5000	50	60	1	1	0	1	1	0
2	22	3	20	2	10000	25	40	1	40	0	10000	25	20	1	1	1	0	0	1
2	24	3	40	0	10000	75	40	3	40	1	5000	50	20	0	0	1	1	1	1
2	28	3	60	0	50	25	20	1	40	2	10000	25	20	1	1	1	1	0	0
Number of times each participant is trading on the attribute --->														**17**	**20**	**22**	**18**	**16**	**19**
Number of times there is overlap on the attribute --->														**11**	**8**	**6**	**10**	**12**	**9**

YoL - levels		# choice sets
0.5	1	9
1	3	6
3	0.5	2
Overlap		11

QoL - levels		# choice sets
20	40	6
40	60	7
60	20	7
Overlap		8

Exp - levels		# choice sets
0	1	10
1	2	3
2	0	9
Overlap		6

Size - levels		# choice sets
50	10000	5
5000	50	4
10000	50000	9
Overlap		10

Equ - levels		# choice sets
25	75	4
50	25	5
75	50	7
Overlap		12

AddTax - levels		# choice sets
20	40	4
40	60	8
60	20	7
Overlap		9

	M1	M2
D error	0.001433	0.000249
A error	0.276693	1.377686

Scenario 1

	Care programme A	Care programme B
Additional years of life	0.5year of extra life	0.5year of extra life
Quality of life improvements	40points improvement	20points improvement
Patient experience	Poor	Good
Size of target population	10,000	10,000
Target population	75% disadvantage	50% disadvantage
Additional budget required	60 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 2

	Care programme A	Care programme B
Additional years of life	3year of extra life	1year of extra life
Quality of life improvements	40points improvement	60points improvement
Patient experience	Poor	Fair
Size of target population	5,000	50
Target population	50% disadvantage	50% disadvantage
Additional budget required	60 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 3

	Care programme A	Care programme B
Additional years of life	3year of extra life	3year of extra life
Quality of life improvements	60points improvement	60points improvement
Patient experience	Poor	Good
Size of target population	10,000	5,000
Target population	75% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	40 additional tax/per year

Which care programme would you choose, A or B?	○	○

Scenario 4

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	20points improvement	40points improvement
Patient experience	Good	Good
Size of target population	10,000	5,000
Target population	50% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 5

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	0.5year of extra life
Quality of life improvements	40points improvement	60points improvement
Patient experience	Fair	Poor
Size of target population	5,000	5,000
Target population	50% disadvantage	50% disadvantage
Additional budget required	60 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 6

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	0.5year of extra life
Quality of life improvements	20points improvement	60points improvement
Patient experience	Good	Poor

Size of target population	5,000	10,000
Target population	50% disadvantage	50% disadvantage
Additional budget required	40 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 7

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	40points improvement	40points improvement
Patient experience	Fair	Good
Size of target population	10,000	5,000
Target population	50% disadvantage	25% disadvantage
Additional budget required	40 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 8

	Care programme A	**Care programme B**
Additional years of life	1year of extra life	1year of extra life
Quality of life improvements	60points improvement	40points improvement
Patient experience	Poor	Fair
Size of target population	10,000	50
Target population	75% disadvantage	75% disadvantage
Additional budget required	60 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 9

	Care programme A	**Care programme B**

Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	60points improvement	20points improvement
Patient experience	Poor	Fair
Size of target population	5,000	5,000
Target population	75% disadvantage	25% disadvantage
Additional budget required	40 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 10

	Care programme A	**Care programme B**
Additional years of life	1year of extra life	1year of extra life
Quality of life improvements	60points improvement	60points improvement
Patient experience	Fair	Good
Size of target population	50	10,000
Target population	50% disadvantage	25% disadvantage
Additional budget required	40 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 11

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	40points improvement	20points improvement
Patient experience	Fair	Poor
Size of target population	50	50
Target population	75% disadvantage	75% disadvantage
Additional budget required	60 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 12

	Care programme A	Care programme B
Additional years of life	1year of extra life	3year of extra life
Quality of life improvements	60points improvement	60points improvement
Patient experience	Fair	Poor
Size of target population	5,000	5,000
Target population	50% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 13

	Care programme A	Care programme B
Additional years of life	1year of extra life	1year of extra life
Quality of life improvements	60points improvement	40points improvement
Patient experience	Good	Fair
Size of target population	5,000	10,000
Target population	25% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 14

	Care programme A	Care programme B
Additional years of life	1year of extra life	0.5year of extra life
Quality of life improvements	60points improvement	40points improvement
Patient experience	Poor	Good
Size of target population	50	50
Target population	50% disadvantage	75% disadvantage
Additional budget required	40 additional tax/per year	40 additional tax/per year

Which care programme would you choose, A or B?	○	○

Scenario 15

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	0.5year of extra life
Quality of life improvements	20points improvement	20points improvement
Patient experience	Poor	Fair
Size of target population	50	50
Target population	75% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 16

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	20points improvement	40points improvement
Patient experience	Fair	Fair
Size of target population	10,000	5,000
Target population	50% disadvantage	25% disadvantage
Additional budget required	60 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 17

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	3year of extra life
Quality of life improvements	60points improvement	20points improvement
Patient experience	Good	Poor

Size of target population	50	10,000
Target population	75% disadvantage	75% disadvantage
Additional budget required	60 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 18

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	1year of extra life
Quality of life improvements	20points improvement	40points improvement
Patient experience	Good	Good
Size of target population	50	50
Target population	25% disadvantage	75% disadvantage
Additional budget required	40 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 19

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	0.5year of extra life
Quality of life improvements	60points improvement	60points improvement
Patient experience	Poor	Good
Size of target population	50	10,000
Target population	75% disadvantage	50% disadvantage
Additional budget required	20 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 20

	Care programme A	**Care programme B**

Additional years of life	1year of extra life	0.5year of extra life
Quality of life improvements	20points improvement	60points improvement
Patient experience	Fair	Fair
Size of target population	50	5,000
Target population	75% disadvantage	50% disadvantage
Additional budget required	60 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 21

	Care programme A	**Care programme B**
Additional years of life	0.5year of extra life	1year of extra life
Quality of life improvements	60points improvement	60points improvement
Patient experience	Fair	Poor
Size of target population	5,000	50
Target population	50% disadvantage	75% disadvantage
Additional budget required	40 additional tax/per year	40 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 22

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	1year of extra life
Quality of life improvements	20points improvement	40points improvement
Patient experience	Good	Poor
Size of target population	10,000	10,000
Target population	25% disadvantage	25% disadvantage
Additional budget required	40 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 23

	Care programme A	Care programme B
Additional years of life	3year of extra life	1year of extra life
Quality of life improvements	20points improvement	60points improvement
Patient experience	Poor	Fair
Size of target population	10,000	10,000
Target population	75% disadvantage	75% disadvantage
Additional budget required	20 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 24

	Care programme A	Care programme B
Additional years of life	3year of extra life	3year of extra life
Quality of life improvements	40points improvement	40points improvement
Patient experience	Poor	Fair
Size of target population	10,000	5,000
Target population	75% disadvantage	50% disadvantage
Additional budget required	40 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?		

Scenario 25

	Care programme A	Care programme B
Additional years of life	0.5year of extra life	0.5year of extra life
Quality of life improvements	20points improvement	60points improvement
Patient experience	Fair	Fair
Size of target population	10,000	5,000
Target population	75% disadvantage	50% disadvantage
Additional budget required	40 additional tax/per year	20 additional tax/per year

Which care programme would you choose, A or B?	○	○

Scenario 26

	Care programme A	**Care programme B**
Additional years of life	1year of extra life	0.5year of extra life
Quality of life improvements	60points improvement	40points improvement
Patient experience	Poor	Poor
Size of target population	50	5,000
Target population	50% disadvantage	50% disadvantage
Additional budget required	40 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 27

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	3year of extra life
Quality of life improvements	20points improvement	60points improvement
Patient experience	Good	Poor
Size of target population	5,000	10,000
Target population	25% disadvantage	25% disadvantage
Additional budget required	40 additional tax/per year	60 additional tax/per year
Which care programme would you choose, A or B?	○	○

Scenario 28

	Care programme A	**Care programme B**
Additional years of life	3year of extra life	1year of extra life
Quality of life improvements	60points improvement	40points improvement
Patient experience	Poor	Good

Size of target population	50	10,000
Target population	25% disadvantage	25% disadvantage
Additional budget required	20 additional tax/per year	20 additional tax/per year
Which care programme would you choose, A or B?		

Some exploratory analyses were conducted, simulating different patterns of behaviour to test both models:

- Complete random responses
- 'Additional years of life' is the only important attribute
- 'Quality of life improvements' is the only important attribute
- Strong preferences towards 'Additional years of life' and 'Quality of life improvements', but not the other attributes
- 'Patient experience' is the only important attribute

When simulating the dependent variable (i.e., the chosen option) full compensatory behaviour was not assumed (i.e. assuming 90% A and 10% B when the attribute is best for A and 90% B and 10% A when the attribute is best for B). Models were estimated by treating variables as indicated initially (i.e., all as numerical, except for 'patient experience'). In all models estimated expected results in terms of the direction of the beta estimates were obtained. In addition, the possibility of estimating some interaction terms of interest was explored (i.e., interaction between 'QoL' and 'Exp' and interaction between 'AddTax' and 'YoL', 'AddTax' and 'QoL', 'AddTax' and 'Exp', and 'AddTax' and 'Size'). The results of two of these estimations, simulating different patterns of behaviour, are presented below (treating 'size' as categorical variable).

Table 5.E.2. Complete random responses & Pattern of behaviour where 'additional years of life' is the only important attribute

	Complete random responses			**If 'additional years of life' is the only important attribute**		
Variable	**Coefficient**	**Robust s.e.***	**t-ratio**	**Coefficient**	**Robust s.e.***	**t-ratio**
asc	-0.023	0.051	-0.454	-0.353	0.051	-2.978
Additional years of life - YoL	-0.040	0.100	-0.397	0.570	0.170	3.347
Quality of life improv. - QoL	-0.001	0.018	-0.065	-0.009	0.026	-0.348
Patient experience - Exp						
Fair	0.046	0.087	0.527	-1.242	0.185	-6.730
Good	-0.007	0.109	-0.060	-1.311	0.342	-7.893
Size of target population - Size						
5,000 /per 100,000 citizens	0.183	0.264	0.694	-0.519	0.339	-1.533
10,000 /per 100,000 citizens	-0.246	0.202	-1.217	-0.468	0.527	-0.888
Target population - Equ	-0.001	0.003	-0.276	-0.033	0.004	-7.674
Add. budget needed - AddTax	-0.002	0.008	-0.208	0.000	0.034	-0.017
Exp_Fair * QoL	-0.001	0.016	-0.072	-0.009	0.028	-0.320
Exp_Good * QoL	-0.002	0.008	-0.306	-0.019	0.008	-1.534
AddTax * YoL	0.002	0.002	0.889	0.008	0.012	3.188

AddTax * QoL	0.000	0.000	0.859	0.000	0.000	2.930
Size_5,000 * AddTax	-0.005	0.006	-0.821	-0.015	0.008	-1.865
Size_10,000 * AddTax	0.006	0.004	1.435	-0.022	0.012	-1.893
Exp_Fair * QoL	-0.002	0.019	-0.088	0.000	0.015	-0.038
Exp_Good * AddTax	-0.003	0.006	-0.532	-0.001	0.015	-0.077
Log likelihood (final)		-1935.78			-1351.3	
AIC		3905.56			2736.61	
BIC		4006.49			2837.54	

*s.e., standard error.

NoAddTax Design: **5 attributes (i.e., 'Additional budget required' included)**

A D-efficient partial profile design with 28 rows (2 blocks; 14 choices per participant) and priors close to zero was created. As before, the model averaging approach was used: first model (M1) was a Bayesian D-Efficient (using priors from pilot #2), and the second model (M2) was a D-eficcient design with interactions and priors close to zero. The second model added interaction between "Exp" and "QoL".

A partial profile design with candidate set constructed in Stata was used to ensure overlap in two attributes for all choice sets. The first three attributes ('YoL', 'QoL' and 'Exp') were treated as categorical, and the other two ('Size' and 'Equ') were treated as continuous. Candidate set excludes potentially dominated choice sets and for that, restrictions were applied when constructing candidate set (size of candidate set: 14,580 rows). The following (i.e., strongly dominated) choice sets were rejected when constructing the candidate set:

drop if yolA>=yolB & qolA>=qolB & expA>=expB & sizeA>=sizeB & equA>=equB
drop if yolA<=yolB & qolA<=qolB & expA<=expB & sizeA<=sizeB & equA<=equB

Table 5.E.3. Experimental NoAddTax Design

Block	Choice set	Care programme A (i.e. alternative 1)						Care programme B (i.e. alternative 2)						Trading attributes & overlaps					
		YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
1	3	0	2	2	50	75		1	1	2	50	25		1	1	0	0	1	.
1	5	0	1	0	5000	25		2	0	0	5000	75		1	1	0	0	1	.
1	6	0	2	1	5000	25		1	2	0	5000	75		1	0	1	0	1	.
1	7	1	2	1	50	75		2	1	1	50	25		1	1	0	0	1	.
1	8	1	2	2	10000	25		0	1	2	10000	75		1	1	0	0	1	.
1	9	1	2	1	5000	75		0	2	2	5000	25		1	0	1	0	1	.
1	10	1	1	2	50	75		1	0	2	10000	25		0	1	0	1	1	.
1	13	1	0	1	5000	25		0	2	1	10000	25		1	1	0	1	0	.
1	14	2	0	1	5000	75		1	0	2	50	75		1	0	1	1	0	.
1	18	2	1	0	5000	50		1	2	0	5000	75		1	1	0	0	1	.
1	19	1	0	1	50	50		0	0	0	50	75		1	0	1	0	1	.
1	23	2	0	0	10000	25		2	1	0	5000	50		0	1	0	1	1	.
1	25	1	2	1	10000	50		2	2	1	50	25		1	0	0	1	1	.
1	26	2	1	2	5000	50		2	2	0	5000	25		0	1	1	0	1	.
2	1	2	1	0	50	75		0	1	2	50	25		1	0	1	0	1	.
2	2	2	1	2	50	25		0	1	1	10000	25		1	0	1	1	0	.
2	4	1	0	0	5000	25		0	0	1	5000	75		1	0	1	0	1	.
2	11	0	2	2	50	25		2	2	1	10000	25		1	0	1	1	0	.
2	12	1	1	2	5000	25		2	0	2	50	25		1	1	0	1	0	.
2	15	2	0	1	10000	25		2	2	1	50	50		0	1	0	1	1	.
2	16	0	2	0	5000	75		1	2	1	50	75		1	0	1	1	0	.
2	17	1	0	2	10000	50		1	2	1	50	50		0	1	1	1	0	.
2	20	2	1	2	50	50		0	1	2	10000	75		1	0	0	1	1	.
2	21	0	2	0	5000	50		1	1	1	5000	50		1	1	1	0	0	.
2	22	1	2	2	10000	75		2	2	0	50	75		1	0	1	1	0	.

2	24	0	1	0	10000	75		1	1	0	5000	25		1	0	0	1	1	.
2	27	2	0	1	50	75		2	2	0	10000	75		0	1	1	1	0	.
2	28	0	2	0	5000	50		0	0	0	50	75		0	1	0	1	1	.
Number of times each participant is trading on the attribute --->														21	15	14	16	18	
Number of times there is overlap on the attribute --->														7	13	14	12	10	

YoL - levels		# choice sets
0	1	10
1	2	6
2	0	5
Overlap		7

QoL - levels		# choice sets
0	1	4
1	2	6
2	0	5
Overlap		13

Exp - levels		# choice sets
0	1	6
1	2	5
2	0	3
Overlap		14

Size - levels		# choice sets
50	10000	9
5000	50	4
10000	50000	3
Overlap		12

Equ - levels		# choice sets
25	75	10
50	25	4
75	50	4
Overlap		10

	M1	M2
D error	0.02194	0.03921
A error	0.752303	0.75318

Scenario 1

	Care programme A	Care programme B
Additional years of life	3 year of extra life	0.5 year of extra life
Quality of life improvements	40 points improvement	40 points improvement
Patient experience	Poor	Good
Size of target population	50	50
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 2

	Care programme A	Care programme B
Additional years of life	3 year of extra life	0.5 year of extra life
Quality of life improvements	40 points improvement	40 points improvement
Patient experience	Good	Fair
Size of target population	50	10,000
Target population	25% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 3

	Care programme A	Care programme B
Additional years of life	0.5 year of extra life	1 year of extra life
Quality of life improvements	60 points improvement	40 points improvement
Patient experience	Good	Good
Size of target population	50	50
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 4

	Care programme A	Care programme B
Additional years of life	1 year of extra life	0.5 year of extra life
Quality of life improvements	20 points improvement	20 points improvement
Patient experience	Poor	Fair
Size of target population	5,000	5,000
Target population	25% disadvantage	75% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 5

	Care programme A	Care programme B
Additional years of life	0.5 year of extra life	3 year of extra life
Quality of life improvements	40 points improvement	20 points improvement
Patient experience	Poor	Poor
Size of target population	5,000	5,000
Target population	25% disadvantage	75% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 6

	Care programme A	Care programme B
Additional years of life	0.5 year of extra life	1 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Fair	Poor
Size of target population	5,000	5,000
Target population	25% disadvantage	75% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 7

	Care programme A	Care programme B

Additional years of life	1 year of extra life	3 year of extra life
Quality of life improvements	60 points improvement	40 points improvement
Patient experience	Fair	Fair
Size of target population	50	50
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 8

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	0.5 year of extra life
Quality of life improvements	60 points improvement	40 points improvement
Patient experience	Good	Good
Size of target population	10,000	10,000
Target population	25% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 9

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	0.5 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Fair	Good
Size of target population	5,000	5,000
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 10

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	1 year of extra life

Quality of life improvements	40 points improvement	20 points improvement
Patient experience	Good	Good
Size of target population	50	10,000
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 11

	Care programme A	**Care programme B**
Additional years of life	0.5 year of extra life	3 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Good	Fair
Size of target population	50	10,000
Target population	25% disadvantage	25% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 12

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	3 year of extra life
Quality of life improvements	40 points improvement	20 points improvement
Patient experience	Good	Good
Size of target population	5,000	50
Target population	25% disadvantage	25% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 13

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	0.5 year of extra life
Quality of life improvements	20 points improvement	60 points improvement

Patient experience	Fair	Fair
Size of target population	5,000	10,000
Target population	25% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 14

	Care programme A	**Care programme B**
Additional years of life	3 year of extra life	1 year of extra life
Quality of life improvements	20 points improvement	20 points improvement
Patient experience	Fair	Good
Size of target population	5,000	50
Target population	75% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 15

	Care programme A	**Care programme B**
Additional years of life	3 year of extra life	3 year of extra life
Quality of life improvements	20 points improvement	60 points improvement
Patient experience	Fair	Fair
Size of target population	10,000	50
Target population	25% disadvantage	50% disadvantage
Which care programme would you choose, A or B?		

Scenario 16

	Care programme A	**Care programme B**
Additional years of life	0.5 year of extra life	1 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Poor	Fair

Size of target population	5,000	50
Target population	75% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 17

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	1 year of extra life
Quality of life improvements	20 points improvement	60 points improvement
Patient experience	Good	Fair
Size of target population	10,000	50
Target population	50% disadvantage	50% disadvantage
Which care programme would you choose, A or B?		

Scenario 18

	Care programme A	**Care programme B**
Additional years of life	3 year of extra life	1 year of extra life
Quality of life improvements	40 points improvement	60 points improvement
Patient experience	Poor	Poor
Size of target population	5,000	5,000
Target population	50% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 19

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	0.5 year of extra life
Quality of life improvements	20 points improvement	20 points improvement
Patient experience	Fair	Poor
Size of target population	50	50

Target population	50% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 20

	Care programme A	**Care programme B**
Additional years of life	3 year of extra life	0.5 year of extra life
Quality of life improvements	40 points improvement	40 points improvement
Patient experience	Good	Good
Size of target population	50	10,000
Target population	50% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 21

	Care programme A	**Care programme B**
Additional years of life	0.5 year of extra life	1 year of extra life
Quality of life improvements	60 points improvement	40 points improvement
Patient experience	Poor	Fair
Size of target population	5,000	5,000
Target population	50% disadvantage	50% disadvantage
Which care programme would you choose, A or B?		

Scenario 22

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	3 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Good	Poor
Size of target population	10,000	50
Target population	75% disadvantage	75% disadvantage

Which care programme would you choose, A or B?	○	○

Scenario 23

	Care programme A	**Care programme B**
Additional years of life	3 year of extra life	3 year of extra life
Quality of life improvements	20 points improvement	40 points improvement
Patient experience	Poor	Poor
Size of target population	10,000	5,000
Target population	25% disadvantage	50% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 24

	Care programme A	**Care programme B**
Additional years of life	0.5 year of extra life	1 year of extra life
Quality of life improvements	40 points improvement	40 points improvement
Patient experience	Poor	Poor
Size of target population	10,000	5,000
Target population	75% disadvantage	25% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 25

	Care programme A	**Care programme B**
Additional years of life	1 year of extra life	3 year of extra life
Quality of life improvements	60 points improvement	60 points improvement
Patient experience	Fair	Fair
Size of target population	10,000	50
Target population	50% disadvantage	25% disadvantage
Which care programme would you choose, A or B?	○	○

Scenario 26

	Care programme A	Care programme B
Additional years of life	3 year of extra life	3 year of extra life
Quality of life improvements	40 points improvement	60 points improvement
Patient experience	Good	Poor
Size of target population	5,000	5,000
Target population	50% disadvantage	25% disadvantage
Which care programme would you choose, A or B?		

Scenario 27

	Care programme A	Care programme B
Additional years of life	3 year of extra life	3 year of extra life
Quality of life improvements	20 points improvement	60 points improvement
Patient experience	Fair	Poor
Size of target population	50	10,000
Target population	75% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Scenario 28

	Care programme A	Care programme B
Additional years of life	0.5 year of extra life	0.5 year of extra life
Quality of life improvements	60 points improvement	20 points improvement
Patient experience	Poor	Poor
Size of target population	5,000	50
Target population	50% disadvantage	75% disadvantage
Which care programme would you choose, A or B?		

Some exploratory analyses were conducted, simulating different patterns of behaviour to test both models:

- Complete random responses (chosen_option_r)
- 'Additional years of life' is the only important attribute (chosen_option_y)
- 'Quality of life improvements' is the only important attribute (chosen_option_q)

- Strong preferences towards 'Additional years of life' and 'Quality of life improvements', but not the other attributes (chosen_option_yq)
- 'Patient experience' is the only important attribute (chosen_option_exp)

When simulating the dependent variable (i.e., the chosen option) full compensatory behaviour was not assumed. The models were estimated by treating variables as indicated initially for this design (i.e., 'YoL', 'QoL' and 'Exp' as categorical, and 'Size' and 'Equ' as continuous). In all models estimated expected results in terms of the direction of the beta estimates were obtained. The possibility of estimating the interactions terms of interest was also tested (i.e., interaction between 'QoL' and 'Exp'). Some of the results of such estimation, simulating different patterns of behaviour, are presented below.

Table 5.E.4. Complete random responses & a pattern of behaviour where 'Additional years of life' is the only important attribute

	Complete random responses			**Pattern of behaviour where 'additional years of life' is the only important attribute**		
Variable	**Coefficient**	**Robust s.e.***	**t-ratio**	**Coefficient**	**Robust s.e.***	**t-ratio**
asc	-0.034	0.039	-0.874	0.027	0.036	0.746
Additional years of life - YoL	0.025	0.032	0.773	1.316	0.233	5.644
Quality of life improv. - QoL	0.000	0.010	-0.009	-0.002	0.021	-0.071
Patient experience - Exp						
Fair	-0.075	0.073	-1.026	0.152	0.040	3.830
Good	0.030	0.100	0.302	-0.436	0.107	-4.089
Size of target population - Size	0.000	0.000	1.479	0.000	0.000	-2.837
Target population - Equ	0.000	0.001	0.282	-0.008	0.001	-9.457
Exp_Fair * QoL	0.000	0.003	-0.029	-0.002	0.021	-0.073
Exp_Good * QoL	0.000	0.005	-0.035	-0.003	0.021	-0.142
Log likelihood (final)	-1938.13			-1241.02		
AIC	3894.27			2500.04		
BIC	3947.7			2553.47		

*s.e., standard error.

Table 5.E.5. "AddTax" Design with 'mirror blocks'

Block_d	Choice task	**Care programme A (i.e. alternative 1)**						**Care programme B (i.e. alternative 2)**					
		YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
1	1	0.5	40	0	10000	75	60	0.5	20	2	10000	50	40
1	2	3	40	0	5000	50	60	1	60	1	50	50	60
1	4	0.5	20	2	10000	50	60	1	40	2	5000	25	60
1	7	0.5	40	1	10000	50	40	1	40	2	5000	25	40
1	10	1	60	1	50	50	40	1	60	2	10000	25	60
1	12	1	60	1	5000	50	60	3	60	0	5000	25	20
1	14	1	60	0	50	50	40	0.5	40	2	50	75	40

1	15	3	20	0	50	75	60	0.5	20	1	50	25	20
1	17	3	60	2	50	75	60	3	20	0	10000	75	40
1	21	0.5	60	1	5000	50	40	1	60	0	50	75	40
1	23	3	20	0	10000	75	20	1	60	1	10000	75	60
1	25	0.5	20	1	10000	75	40	0.5	60	1	5000	50	20
1	26	1	60	0	50	50	40	0.5	40	0	5000	50	20
1	27	3	20	2	5000	25	40	3	60	0	10000	25	60
2	3	3	60	0	10000	75	60	3	60	2	5000	25	40
2	5	3	40	1	5000	50	60	0.5	60	0	5000	50	20
2	6	0.5	20	2	5000	50	40	0.5	60	0	10000	50	60
2	8	1	60	0	10000	75	60	1	40	1	50	75	40
2	9	0.5	60	0	5000	75	40	1	20	1	5000	25	40
2	11	0.5	40	1	50	75	60	1	20	0	50	75	20
2	13	1	60	2	5000	25	60	1	40	1	10000	25	20
2	16	0.5	20	1	10000	50	60	1	40	1	5000	25	60
2	18	3	20	2	50	25	40	1	40	2	50	75	60
2	19	0.5	60	0	50	75	20	0.5	60	2	10000	50	60
2	20	1	20	1	50	75	60	0.5	60	1	5000	50	60
2	22	3	20	2	10000	25	40	1	40	0	10000	25	20
2	24	3	40	0	10000	75	40	3	40	1	5000	50	20
2	28	3	60	0	50	25	20	1	40	2	10000	25	20
3	1	0.5	20	2	10000	50	40	0.5	40	0	10000	75	60
3	2	1	60	1	50	50	60	3	40	0	5000	50	60
3	4	1	40	2	5000	25	60	0.5	20	2	10000	50	60
3	7	1	40	2	5000	25	40	0.5	40	1	10000	50	40
3	10	1	60	2	10000	25	60	1	60	1	50	50	40
3	12	3	60	0	5000	25	20	1	60	1	5000	50	60
3	14	0.5	40	2	50	75	40	1	60	0	50	50	40
3	15	0.5	20	1	50	25	20	3	20	0	50	75	60
3	17	3	20	0	10000	75	40	3	60	2	50	75	60
3	21	1	60	0	50	75	40	0.5	60	1	5000	50	40
3	23	1	60	1	10000	75	60	3	20	0	10000	75	20
3	25	0.5	60	1	5000	50	20	0.5	20	1	10000	75	40
3	26	0.5	40	0	5000	50	20	1	60	0	50	50	40
3	27	3	60	0	10000	25	60	3	20	2	5000	25	40
4	3	3	60	2	5000	25	40	3	60	0	10000	75	60
4	5	0.5	60	0	5000	50	20	3	40	1	5000	50	60
4	6	0.5	60	0	10000	50	60	0.5	20	2	5000	50	40
4	8	1	40	1	50	75	40	1	60	0	10000	75	60
4	9	1	20	1	5000	25	40	0.5	60	0	5000	75	40
4	11	1	20	0	50	75	20	0.5	40	1	50	75	60
4	13	1	40	1	10000	25	20	1	60	2	5000	25	60
4	16	1	40	1	5000	25	60	0.5	20	1	10000	50	60
4	18	1	40	2	50	75	60	3	20	2	50	25	40

4	19	0.5	60	2	10000	50	60	0.5	60	0	50	75	20
4	20	0.5	60	1	5000	50	60	1	20	1	50	75	60
4	22	1	40	0	10000	25	20	3	20	2	10000	25	40
4	24	3	40	1	5000	50	20	3	40	0	10000	75	40
4	28	1	40	2	10000	25	20	3	60	0	50	25	20

Table 5.E.6. "NoAddTax" Design with 'mirror blocks'

Block_d	Choice task	**Care programme A (i.e. alternative 1)**						**Care programme B (i.e. alternative 2)**					
		YoL	QoL	Exp	Size	Equ	Add Tax	YoL	QoL	Exp	Size	Equ	Add Tax
5	3	0	2	2	50	75	.	1	1	2	50	25	.
5	5	0	1	0	5000	25	.	2	0	0	5000	75	.
5	6	0	2	1	5000	25	.	1	2	0	5000	75	.
5	7	1	2	1	50	75	.	2	1	1	50	25	.
5	8	1	2	2	10000	25	.	0	1	2	10000	75	.
5	9	1	2	1	5000	75	.	0	2	2	5000	25	.
5	10	1	1	2	50	75	.	1	0	2	10000	25	.
5	13	1	0	1	5000	25	.	0	2	1	10000	25	.
5	14	2	0	1	5000	75	.	1	0	2	50	75	.
5	18	2	1	0	5000	50	.	1	2	0	5000	75	.
5	19	1	0	1	50	50	.	0	0	0	50	75	.
5	23	2	0	0	10000	25	.	2	1	0	5000	50	.
5	25	1	2	1	10000	50	.	2	2	1	50	25	.
5	26	2	1	2	5000	50	.	2	2	0	5000	25	.
6	1	2	1	0	50	75	.	0	1	2	50	25	.
6	2	2	1	2	50	25	.	0	1	1	10000	25	.
6	4	1	0	0	5000	25	.	0	0	1	5000	75	.
6	11	0	2	2	50	25	.	2	2	1	10000	25	.
6	12	1	1	2	5000	25	.	2	0	2	50	25	.
6	15	2	0	1	10000	25	.	2	2	1	50	50	.
6	16	0	2	0	5000	75	.	1	2	1	50	75	.
6	17	1	0	2	10000	50	.	1	2	1	50	50	.
6	20	2	1	2	50	50	.	0	1	2	10000	75	.
6	21	0	2	0	5000	50	.	1	1	1	5000	50	.
6	22	1	2	2	10000	75	.	2	2	0	50	75	.
6	24	0	1	0	10000	75	.	1	1	0	5000	25	.
6	27	2	0	1	50	75	.	2	2	0	10000	75	.
6	28	0	2	0	5000	50	.	0	0	0	50	75	.
7	3	1	1	2	50	25	.	0	2	2	50	75	.
7	5	2	0	0	5000	75	.	0	1	0	5000	25	.
7	6	1	2	0	5000	75	.	0	2	1	5000	25	.
7	7	2	1	1	50	25	.	1	2	1	50	75	.
7	8	0	1	2	10000	75	.	1	2	2	10000	25	.
7	9	0	2	2	5000	25	.	1	2	1	5000	75	.

7	10	1	0	2	10000	25	.	1	1	2	50	75	.
7	13	0	2	1	10000	25	.	1	0	1	5000	25	.
7	14	1	0	2	50	75	.	2	0	1	5000	75	.
7	18	1	2	0	5000	75	.	2	1	0	5000	50	.
7	19	0	0	0	50	75	.	1	0	1	50	50	.
7	23	2	1	0	5000	50	.	2	0	0	10000	25	.
7	25	2	2	1	50	25	.	1	2	1	10000	50	.
7	26	2	2	0	5000	25	.	2	1	2	5000	50	.
8	1	0	1	2	50	25	.	2	1	0	50	75	.
8	2	0	1	1	10000	25	.	2	1	2	50	25	.
8	4	0	0	1	5000	75	.	1	0	0	5000	25	.
8	11	2	2	1	10000	25	.	0	2	2	50	25	.
8	12	2	0	2	50	25	.	1	1	2	5000	25	.
8	15	2	2	1	50	50	.	2	0	1	10000	25	.
8	16	1	2	1	50	75	.	0	2	0	5000	75	.
8	17	1	2	1	50	50	.	1	0	2	10000	50	.
8	20	0	1	2	10000	75	.	2	1	2	50	50	.
8	21	1	1	1	5000	50	.	0	2	0	5000	50	.
8	22	2	2	0	50	75	.	1	2	2	10000	75	.
8	24	1	1	0	5000	25	.	0	1	0	10000	75	.
8	27	2	2	0	10000	75	.	2	0	1	50	75	.
8	28	0	0	0	50	75	.	0	2	0	5000	50	.

Appendix 5.F. Survey instrument

LANDING PAGE

What type of care programme would people prefer In England?: An online survey of people's preferences

Thank you for accepting our invitation to take this survey.
Your responses will help us to inform local healthcare investment decisions in England.

If you are completing this survey using a tablet or smartphone, we recommend you use landscape mode.

In total, this survey will take about 15-20 minutes to complete. It will take approximately 10 minutes to work through the instructions and warm-up questions, and about 5-10 minutes to complete the survey questions themselves.

Please click next to start

SCREENING QUESTIONS

S1. How old are you?: ___xxx years_ [PN: Enter a number, minimum=0; maximum=99]

S1DUMMY. Please select the age category you belong to [one choice]:

- Less than 18 [PN: close]
- 18-24 years
- 25-34 years
- 35-44 years
- 45-54 years
- 55-64 years
- 65+ years

S2. What is your gender? [one choice]

- Male
- Female
- Other

S3. Where do you live? [one choice]

- South East
- South West
- London
- East of England
- West Midlands
- East Midlands
- Yorkshire and the Humber
- North West
- North East

S4. What is your ethnic group? [one choice]

- White (English, Welsh, Scottish, Northern Irish, British, Irish, Gypsy or Irish traveller, Roma, Any other white background)
- Mixed or multiple ethnic groups (White & Black Caribbean, White and Black African, White and Asian, Any other Mixed or multiple ethnic background)
- Asian or Asian British (Indian, Pakistani, Bangladesh, Chinese, any other Asian background)
- Black, Black British, Caribbean or African
- Other ethnic group (Arab, or any other group)

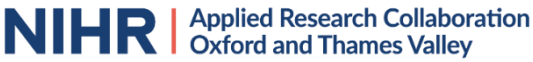
S5. Which one of the following categories best describes the employment of the main income earner in your household? [one choice]

- Work as a professional such as lawyer, doctor, accountant, or other occupation that requires extensive education and training
- Very senior manager in business or commerce
- Top-level civil servant
- Self-employed with 25+ employees
- Self-employed with 5-24 employees
- Self-employed with 1-4 employees (in addition to you)

- Middle management executive in large organization
- Junior management
- Principal officer in local government or civil service
- Other public services (police, health, education, fire service)
- Other non-manual work, such as clerical, secretarial, or sales
- Skilled manual worker (working in a trade or craft occupation)
- Other manual worker with responsibility for other people
- Semi-skilled worker or worker in manual or service occupation
- Apprentice or trainee to skilled worker
- Student
- Still at school
- Unemployed and seeking work
- Retired
- Not in paid work for other reason
- Never worked

WELCOME PAGE





Welcome to this survey of the University of Oxford!

Local healthcare decision-makers in England are responsible for commissioning (or purchasing) the best health care services for their communities. The budget to do so, however, is limited and consequently, local decision-makers are forced to choose among numerous care programmes that can be offered.

As a member of the public, you are being invited to participate in this survey. You will see a series of descriptions of different *hypothetical* care programmes, in pairs, and we will ask you to choose which of the pair you would prefer local healthcare decision-makers to fund and support. The results will help us understand public views and preferences, and with this inform local healthcare investment decisions in England. Before the end of the survey, we will also ask you a couple of socio-demographic questions. Some of these questions will ask about your health status, which may be perceived as sensitive. For each one of these questions, you will have as one of the options, "*I prefer not to answer*".

This survey is being conducted by researchers at the Nuffield Department of Population Health, University of Oxford. We are interested in your views about how local healthcare budgets should be spent across England. Your participation is entirely voluntary.

Please try to answer all of the questions in this survey. Notice that there are no right or wrong answers, as we are primarily interested in your views. At the start of the survey, you will see some background information to help you answer the questions. Please read this carefully before you complete the survey.

This survey should take between 15 and 20 minutes to complete. You may withdraw at any point during the survey for any reason, before submitting your answers, by closing your browser window.

The principal researcher is Pamela Gongora, MSc. Colleagues at the University may assist with the statistical analysis of the data. The survey responses will be collected by Dynata. Your answers will be completely anonymous, and we will use all reasonable endeavours to keep them confidential. Responses to the survey will be collected into a single data file which does not identify any individual, transferred to the University researcher, and stored for a minimum of three years after publication as a password-protected file. Your IP address will not be stored by the University. The data will only be used for research purposes, and the results may be used in academic publications. Your information will not be shared with any other institutions.

For more information on how your information will be processed and protected, please review The University of Oxford's privacy policy here: https://compliance.admin.ox.ac.uk/individual-rights & https://www.ox.ac.uk/privacy-policy

This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee [*R74765/RE001*].

What if there is a problem?

Please contact Dynata. If they cannot resolve your concern, they will contact the lead researcher and work with her to deal with your concern. If you remain unhappy or wish to make a formal complaint, Dynata will provide you with contact details for the Chair of the Research Ethics Committee at the University of Oxford.
Having read the information on the previous page, do you agree to participate in this study?
☐ Yes, I agree to take part
☐ No, I do not consent [PN: TERMINATE]

THANK YOU FOR YOUR HELP

ABOUT THIS SURVEY

As mentioned before, with this survey we aim to understand which care programmes members of the public in England would prefer to see funded.

To respond to this survey we would like you to imagine that you are responsible for health and social care in your county. **You have a limited budget and have to decide which care programme to support and fund.** We are going to ask you to make a series of choices for 14 different scenarios offering two care programmes. The care programmes are described by six different features. In each scenario, we will show you a side-by-side description of the two care programmes, and we would like you to choose one of them. Even if they look alike, these 14 scenarios are different. Repeating these scenarios will helps us understand the importance you place on different features of health care.

We would like you to answer based on your views, as a member of the public.

The following pages provide more information about the **six** [PN: if Design 1] / **five** [PN: if Design 2] **features or characteristics** used to describe the care programmes in each scenario and the choices that we will ask you to make. Afterwards, a practice question will be presented to familiarise you with the paired comparisons that you will see in the 14 scenarios. You will then see the 14 scenarios, and in each one of them, we will ask you to select your most preferred option. The survey will finish with a couple of socio-demographic questions.

THANK YOU FOR YOUR HELP

Click next for the first feature

THE SIX FEATURES *[PN: if Design 1]*

THE FIVE FEATURES *[PN: if Design 2]*

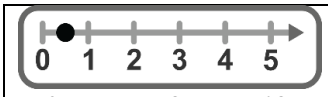
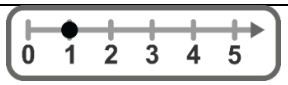
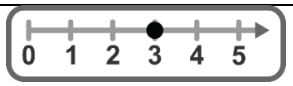
Additional years of life

The first feature is **Additional years of life**. This is, the number of additional years of life that each patient is expected to have with the care programme. In the scenarios, for each care programme, the additional years of life that patients gain due to the care programme will be described using any of the following three options:

- 0.5 years of extra life
- 1 year of extra life
- 3 years of extra life

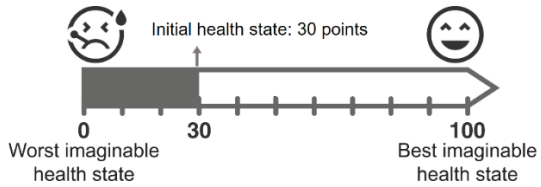
Please answer the next question about **additional years of life**. There are no right or wrong answers – we are just interested in what you think.

P1. How many additional years of life, do you think, a patient with a chronic condition under a care programme is expected to live? (one choice)

 **0.5 year of extra life**	 **1 year of extra life**	 **3 years of extra life**
O	**O**	**O**

Improvements in quality of life

The second feature relates to **improvements in patients' quality of life** due to the care programme. This could include improvements in terms of mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression. Improvements in quality of life can be described by using a scale numbered from 0 to 100, in which 100 means the best imaginable health state and 0 means the worst imaginable health state. The initial current heath state indicated in grey is 30.



The black area in the scale indicates the improvements in terms of quality of life. In the scenarios, for each care programme, improvements in quality of life due to the care programme will be described using any of the following three options:

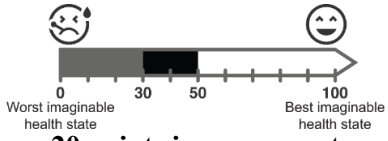
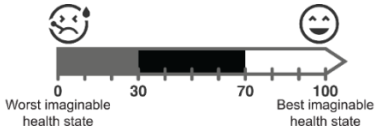
- 20 points improvements
- 40 points improvements
- 60 points improvements

Please answer the next question about **improvements in quality of life**. There are no right or wrong answers – we are just interested in what you think.

P2. Please read this short story about an imaginary man named Rob and answer the question that follows:

Rob is a 50 years old man who had a stroke. He relies on someone to take a shower or get dressed and is severely depressed. After being in a care programme for one year, Rob is now able to wash himself but still experiences depressive episodes.

Which option below, do you think, best describes the improvements in terms of quality of life that Rob has experienced due to the care programme? (one choice)

 **20 points improvement**	 **40 points improvement**	 **60 points improvement**
O	**O**	**O**

Patient experience

The third feature is **Patient experience**, which refers to how patients, their families and carers experience health and social care while taking part in the care programme. Patient experience includes ease of access, quality of communication, shared decision-making, continuity and/or coordination of care. In the scenarios, for each care programme, we will describe the patient experience using any of the following three options:

- Poor
- Fair
- Good

Please answer the next question about **patient experience**. There are no right or wrong answers – we are just interested in what you think.

P3. How do you think, patients with chronic kidney disease have on average rated their experience with their GP surgery in the last year? (one choice)

 **Poor**	 **Fair**	 **Good**
O	**O**	**O**

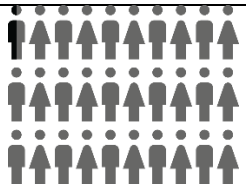
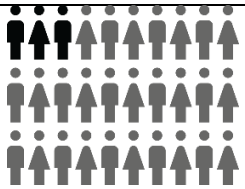
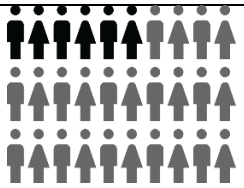
Size of the target population

The fourth feature is **Size of the target population.** This is, the number of people who benefit directly from the care programme per 100,000 citizens. In the scenarios, for each care programme, we will describe the size of the target population using any of the following three options:

- 50 / per 100,000 citizens
- 5,000 / per 100,000 citizens
- 10,000 / per 100,000 citizens

Please answer the next question about **size of target population**. There are no right or wrong answers – we are just interested in what you think.

P4. How many people, do you think, receive health and social care from a programme run in London for patients with Long Covid? (one choice)

		
50 / per 100,000 citizens	**5,000 / per 100,000 citizens**	**10,000 / per 100,000 citizens**
O	**O**	**O**

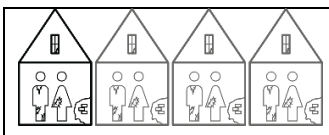
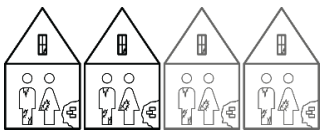
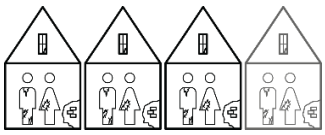
Target population

The fifth feature is the **Target population** of the care programme**.** This feature refers to the proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, the proportion of people covered by the care programme that comes from low-income households. In the scenarios, for each care programme, we will describe the target population by any of the following three options:

- 25% disadvantaged
- 50% disadvantaged
- 75% disadvantaged

Please answer the next question about **target population**. There are no right or wrong answers – we are just interested in what you think.

P5. A care programme in Manchester offers health and social care for patients with cancer. Which proportion of these patients, do you think, come from disadvantaged or vulnerable backgrounds? (one choice)

		
25% disadvantaged	**50% disadvantaged**	**75% disadvantaged**

O	**O**	**O**

Additional budget required [PN: To be presented to participants that will respond the survey with this attribute as one of the options in the choice task (i.e. Design 1)]

The final feature is **Additional budget required**. This feature refers to the additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme. In the scenarios, for each care programme, we will describe the additional budget required to fund the care programme using any of the following three options:

- £20 additional tax / per year
- £40 additional tax / per year
- £60 additional tax / per year

Please answer the next question about **Additional budget required**. There are no right or wrong answers – we are just interested in what you think.

P6. A care programme for children with rare diseases is about to be implemented in Derby. To cover the total cost of this programme, how much additional income tax, do you think, would be required? (one choice)

 **£20 additional tax / per year**	 **£40 additional tax / per year**	 **£60 additional tax / per year**
O	**O**	**O**

The care programmes are hypothetical and as such, we will refer to them as Care programme A and Care programme B. This is because we are not interested in any particular care programme, but in which type of programme you would prefer to be funded based on the features listed above.

You will see how Care Programme A performs relative to Care Programme B in each one of the features, with the darkest purple used to denote the worst, and lighter purple used to denote better performance. If both care programmes perform equally, they will be presented in the same colour.

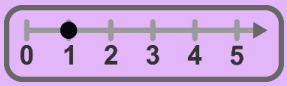
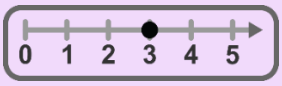
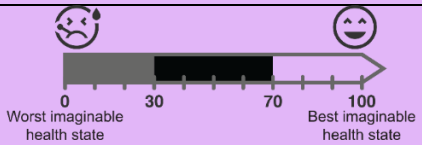
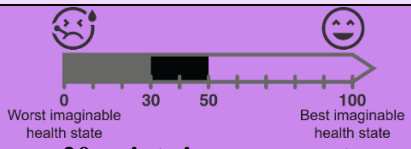
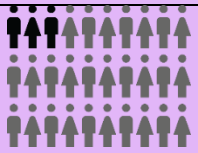
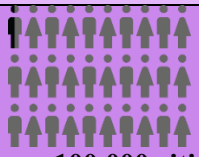
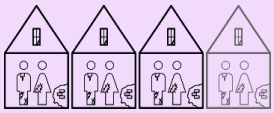
In each of the 14 hypothetical scenarios, you will have to choose between Care Programme A and Care programme B. You will not have the options "*I can't decide*" or "*I'm indifferent*". This is because in real life, local healthcare decision-makers are forced to make investment decisions on behalf of their local populations, based on a limited budget. Even if the decision is difficult, local decision-makers have to decide which care programmes to support and fund, at the expense of other care programmes. Regardless of how difficult it may look, we would like you to imagine that you are in the shoes of local healthcare decision-makers and make a choice.

PRACTICE QUESTIONS

Before we start responding the 14 scenario questions of the main study, we would like you to complete two practice choice questions.

P7. *Imagine that you are responsible for health and social care in your county. You have a limited budget and have to decide which care programme to support and fund. Which of these two care programmes would you choose?*

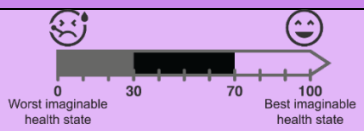
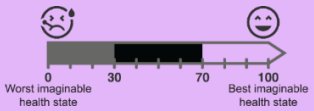
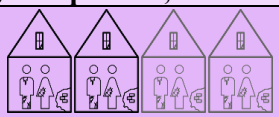
The darkest purple denotes the worst, and lighter purple denotes better performance. If both care programmes perform equally, they will be presented in the same colour.

	Care programme A	***Care programme B***
Additional years of life The number of additional years of life that each patient is expected to have with the care programme.	 **1 year of extra life**	 **3 years of extra life**
Quality of life improvements Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression, due to the care programme	 **40 points improvement**	 **20 points improvement**
Patient experience How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, etc…	 **Fair**	 **Fair**
Size of target population The number of people who benefit directly from the care programme per 100,000 citizens.	 **5,000 / per 100,000 citizens**	 **50 / per 100,000 citizens**
Target population Proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, people from low-income households.	 **75% disadvantaged**	 **25% disadvantaged**
Additional budget required [PN: only to be displayed for Block 1] Additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme.	 **£20 additional tax / per year**	 **£60 additional tax / per year**
Which care programme would you choose, A or B? [one choice]	***A*** ○	***B*** ○

P8. *Imagine that you are responsible for health and social care in your county. You have a limited budget and have to decide which care programme to support and fund. Which of these two care programmes would you choose?*

The darkest purple denotes the worst, and lighter purple denotes better performance. If both care programmes perform equally, they will be presented in the same colour.

[PN: If respondents choose B, the survey TERMINATES for them. We are treating this question as a quality filter]

	Care programme A	*Care programme B*
Additional years of life The number of additional years of life that each patient is expected to have with the care programme.	 **0.5 years of extra life**	 **0.5 years of extra life**
Quality of life improvements Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression, due to the care programme	 **40 points improvements**	 **40 points improvements**
Patient experience How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, etc…	 **Good**	 **Fair**
Size of target population The number of people who benefit directly from the care programme per 100,000 citizens.	 **10,000 / per 100,000 citizens**	 **5,000 / per 100,000 citizens**
Target population Proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, people from low-income households.	 **50% disadvantaged**	 **50% disadvantaged**
Additional budget required [PN: only to be displayed for Block 1] Additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme.	 **£20 additional tax / per year**	 **£40 additional tax / per year**
Which care programme would you choose, A or B? [one choice]	***A*** ○	***B*** ○

MAIN QUESTIONS

Thank you for completing the practice questions.

Now we would like you to complete the main part of the survey.

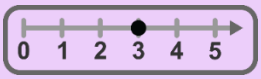
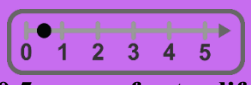
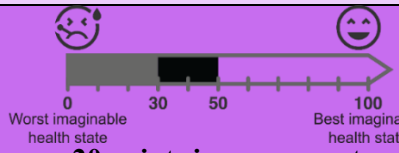
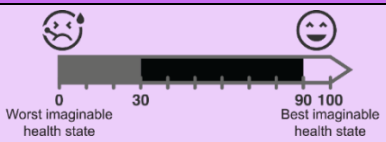
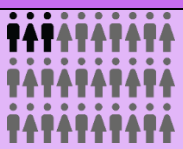
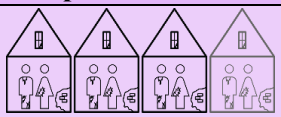
We are going to present you with 14 scenarios that are similar to the practice question. Each scenario asks you to make the same choice about which care programme should be funded and supported by local healthcare decision-makers in England. However, the exact features of the care programmes will be different in each scenario.

Regardless of how difficult it may look, please indicate which care programme you would choose if you were the local healthcare decision-maker in your county.

QUESTION 1

QT1. *Imagine that you are responsible for health and social care in your county. You have a limited budget and have to decide which care programme to support and fund. Which of these two care programmes would you choose?*

The darkest purple denotes the worst, and lighter purple denotes better performance. If both care programmes perform equally, they will be presented in the same colour.

	Care programme A	*Care programme B*
Additional years of life The number of additional years of life that each patient is expected to have with the care programme.	 **3 years of extra life**	 **0.5 years of extra life**
Quality of life improvements Improvements in patients' mobility, self-care, usual activities, pain, discomfort, anxiety, and/or depression, due to the care programme	 **20 points improvement**	 **60 points improvement**
Patient experience How patients, their families and carers experience health and social care. For example, ease of access, quality of communication, etc…	 **Good**	 **Poor**
Size of target population The number of people who benefit directly from the care programme per 100,000 citizens.	 **5,000 / per 100,000 citizens**	 **5,000 / per 100,000 citizens**
Target population Proportion of the target population that comes from disadvantaged or vulnerable backgrounds. For example, people from low-income households.	 **25% disadvantaged**	 **75% disadvantaged**
Additional budget required Additional income tax per year that every taxpayer needs to pay to cover the costs of the care programme.	 **£20 additional tax/ per year**	 **£20 additional tax/ per year**
Which care programme would you choose, A or B? [one choice]	***A*** ○	***B*** ○

[PN: etc… to QUESTION QT14]

DEBRIEFING QUESTIONS

Q1. How would you rate the level of difficulty of the questionnaire? [one choice]

a. Not difficult
b. Somehow difficult
c. Difficult
d. Very difficult

Q2. Are there any further comments that you would like to make regarding this survey? [PN: free text but add a note warning participants NOT to add any personal information]

_____

_____

_____

_____

RESPONDENT CHARACTERISTICS

Now we would like to ask some questions about you. All of the information that you provide will help us in our analysis, and all of your details will remain confidential.

If you do not want to answer some of these questions, you do not have to. All questions will have the option "*I prefer not to answer*".

Q3. Which one is the highest educational or school qualification that you have obtained? [one choice]

a. Primary education
b. Secondary education
c. Short cycle tertiary education
d. Bachelor's or equivalent level
e. Postgraduate education (e.g. Master, PhD, or equivalent level)
f. I prefer not to answer

Q4. In which sector do you work? [one choice]

a. Agriculture, forestry & fishing
b. Mining, energy and water supply
c. Manufacturing
d. Construction
e. Wholesale, retail & repair of motor vehicles
f. Transport & storage
g. Accommodation & food services
h. Information & communication
i. Financial & insurance activities
j. Real estate activities
k. Professional, scientific & technical activities
l. Administrative & support services
m. Public admin & defence; social security
n. Education
o. Human health & social work activities
p. Other services
q. I am unemployed
r. I am a student
s. I prefer not to answer

Q5. Under each heading, please tick the ONE box that best describes your health TODAY:

a. MOBILITY [one choice]:
    - have no problems in walking about
    - I have slight problems in walking about
    - I have moderate problems in walking about
    - I have severe problems in walking about
    - I am unable to walk about

b. SELF-CARE [one choice]:
    - I have no problems washing or dressing myself
    - I have slight problems washing or dressing myself
    - I have moderate problems washing or dressing myself
    - I have severe problems washing or dressing myself
    - I am unable to wash or dress myself

c. USUAL ACTIVITIES (e.g. work, study, housework, family or leisure activities) [one choice]:
    - I have no problems doing my usual activities
    - I have slight problems doing my usual activities
    - I have moderate problems doing my usual activities
    - I have severe problems doing my usual activities
    - I am unable to do my usual activities

d. PAIN / DISCOMFORT [one choice]:
    - I have no pain or discomfort
    - I have slight pain or discomfort
    - I have moderate pain or discomfort
    - I have severe pain or discomfort
    - I have extreme pain or discomfort

e. ANXIETY / DEPRESSION [one choice]:
    - I am not anxious or depressed
    - I am slightly anxious or depressed
    - I am moderately anxious or depressed
    - I am severely anxious or depressed
    - I am extremely anxious or depressed

    - I prefer not to answer

Q6. We would like to know how good or bad is your health TODAY.

This scale is numbered from 0 to 100.

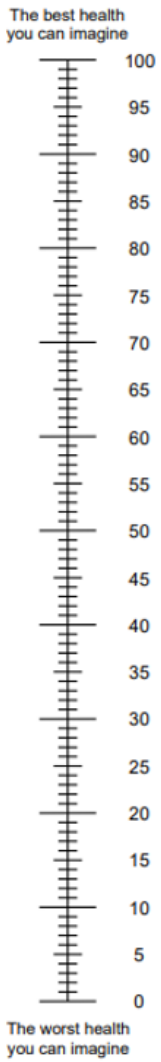
100 means the best health you can imagine.

0 means the worst health you can imagine.

Please use this scale to indicate how your health is TODAY.

Now write the number you marked on the scale in the box below.

YOUR HEALTH TODAY = ☐



- ○ I prefer not to answer

Q7. Do you have any of the following health problems? Please select as many as apply [PN: multiple choice]

Health problems [PN: This is the header of the table]	Select
Asthma	
Cancer during the past five years	
Chronic back pain or sciatica	
Chronic bronchitis COPD or emphysema	
Chronic kidney disease	
Cognitive problems, memory loss, dementia, Alzheimer's	
Colon problem, irritable bowel or colitis	
Congestive heart failure	
Depression, anxiety or emotional difficulties	
Diabetes	
Hard of hearing	
Heart disease, angina (chest pain from heart problem), heart attack, bypass surgery or angioplasty	
High blood pressure	
High cholesterol	
Osteoarthritis (not rheumatoid arthritis)	
Osteoporosis (brittle bones)	

Overweight	
Poor circulation in your legs	
Rheumatoid arthritis	
Rheumatic disease, fibromyalgia or lupus	
Stomach problem, ulcer, gastritis or reflux	
Stroke	
Thyroid disorder	
Vision problem	
No health problems	
Other health problems	
I prefer not to answer [PN: Exclusive. Code 99]	

Q8. Have you experienced any of the above health problems?

	Yes	No	I prefer not to answer
In your family [one choice]	**O**	**O**	**O**
As a carer for a family member [one choice]	**O**	**O**	**O**

Appendix 5.G. Census balanced quotas

Population (by gender)	Total	Male (%)	Female (%)
North East	5%	49%	51%
North West	13%	49%	51%
Yorkshire and The Humber	10%	49%	51%
East Midlands	9%	49%	51%
West Midlands	10%	49%	51%
East	11%	49%	51%
London	16%	50%	50%
South East	16%	49%	51%
South West	10%	49%	51%
Total (England)	100%	49%	51%

Source: ONS (2021) mid-year population estimates (England population +18). Office for National Statistics: Experimental statistics

Population (by age group)	18–24	25–34	35–44	45–54	55–64	65+
North East	11%	16%	14%	16%	17%	25%
North West	11%	17%	15%	17%	16%	24%
Yorkshire and The Humber	11%	17%	15%	17%	16%	24%
East Midlands	11%	16%	15%	17%	16%	25%
West Midlands	11%	17%	15%	17%	15%	24%
East	9%	16%	16%	17%	16%	25%
London	11%	23%	21%	17%	13%	16%
South East	10%	15%	16%	18%	16%	25%
South West	10%	15%	14%	16%	17%	28%
Total (England)	11%	17%	16%	17%	16%	24%

Source: ONS (2021) mid-year population estimates (England population +18). Office for National Statistics: Experimental statistics

Ethnic group	%
Asian or Asian British	8.3%
Black, African, Caribbean or Black British	3.7%
Mixed or multiple ethnic groups	1.9%
White	84.2%
Other ethnic group	1.9%
Total	100.0%

Source: Based on Experimental statistics: Population estimates by ethnic group, England and Wales: 2019. Population estimates by ethnic group, England and Wales: 2019 (England all population). Office for National Statistics: Experimental statistics

Appendix 5.H. Multinomial logit estimates

Table 5.H.1 Parameter coefficients from the multinomail logit (dummy-coded)

Variable	MNL DCE - NoAddTax Coefficient (s.e)	t-ratio	MNL DCE - AddTax Coefficient (s.e)	t-ratio
asc	0.076(0.042)	1.807	-0.013(0.041)	-0.312
Additional years of life				
1 year of extra life	0.532(0.067)	7.943	0.424(0.099)	4.269
3 years of extra life	1.347(0.105)	12.793	1.467(0.171)	8.598
Quality of life improvements				
40 points improvement	0.488(0.101)	4.840	0.380(0.11)	3.450
60 points improvement	0.985(0.103)	9.576	0.765(0.137)	5.591
Patient experience				
Fair	1.259(0.095)	13.287	1.044(0.148)	7.049
Good	1.472(0.130)	11.334	1.224(0.114)	10.724
Size of target population				
5,000 /per 100,000 citizens	0.737(0.103)	7.180	1.006(0.155)	6.506
10,000 /per 100,000 citizens	1.161(0.081)	14.292	1.285(0.154)	8.356
Target population				
50% disadvantaged	0.555(0.100)	5.576	0.153(0.100)	1.521
75% disadvantaged	0.505(0.066)	7.703	0.249(0.169)	1.475
Additional budget needed				
£40 additional tax / per year			-0.506(0.107)	-4.741
£60 additional tax / per year			-0.963(0.140)	-6.866
Log likelihood (final)	-1658.12		-1728.42	
AIC	3338.23		3482.84	
BIC	3403.6		3560.1	

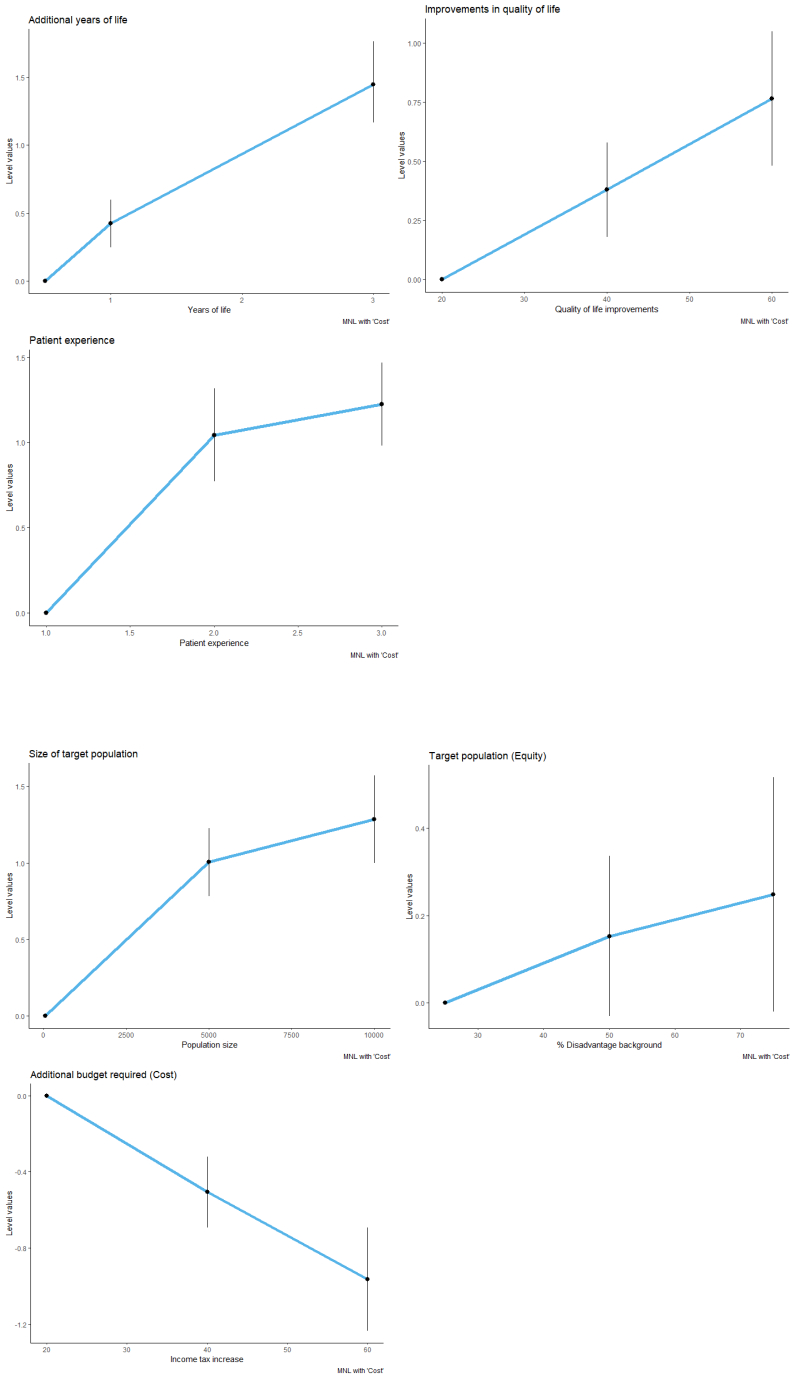
Note: ASC,Alternative-specific constant;s.e., standard error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion

The part-worth utilities (i.e. preference weights[68]) for the DCE-AddTax and DCE-NoAddTax are presented below. Estimates are based on dummy-coded multinomial logit models (see Appendix 5.8). Both figures indicate that higher levels of 'YoL', 'QoL', 'Exp' and 'Size' are preferred to lower levels. The preference weights for the 'Equ' attribute suggest the same in the "AddTax" – DCE, but the most preferred level (i.e. 75% disadvantaged population) has a lower weight than the medium level (50% disadvantaged population) in the "NoAddTax" – DCE. As expected, the 'AddTax' attribute displays a negative relationship between levels, indicating preference for lower taxes. A linear continuous specification could be used for 'YoL', 'QoL' and 'AddTax', but not for 'Size' or 'Equ' as the relative marginal utility of a 1-unit change in any of these two attributes is more important when the change is between level 1 and level 2, than when the change is between level 2 and level 3.

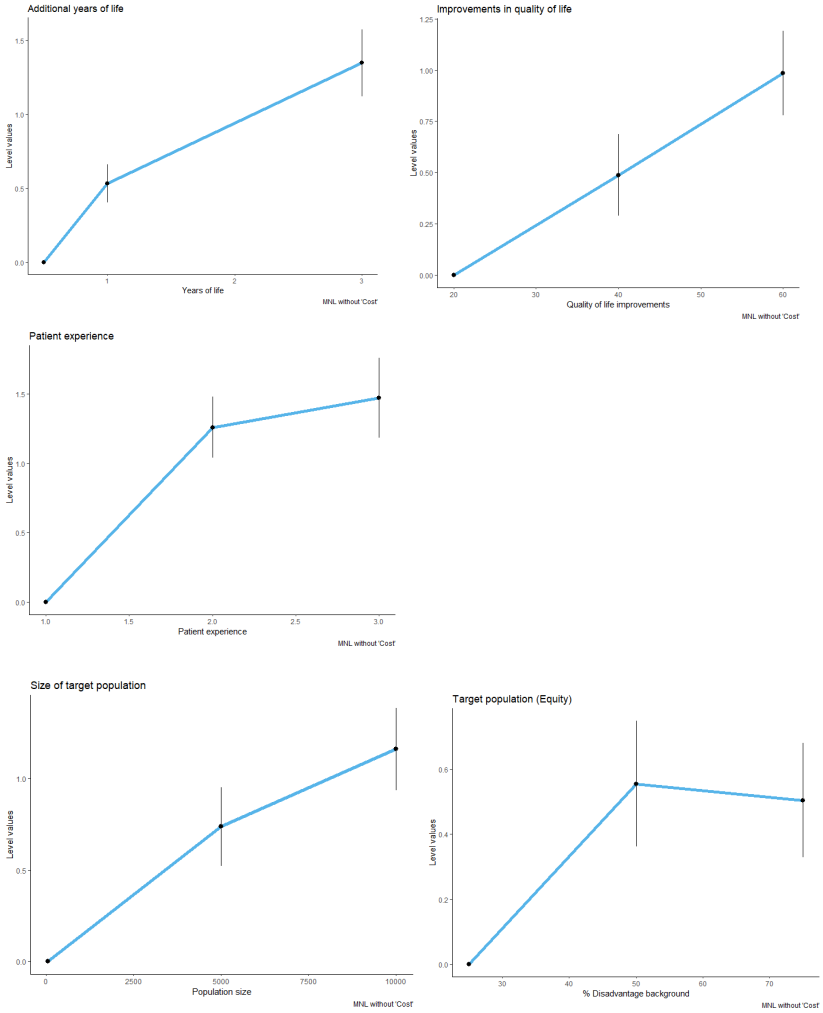
[68] Preference weights measure relative preference, implying that only variations in attribute-level evaluations and the proportional magnitudes of those changes among attributes hold significant meanings (298).

Figure 5.H.1. Part-worth utilities

a. DCE-AddTax



b. DCE-NoAddTax



Appendix 5.I. Comparison between MNL and MMNL & MMNL with QALY gains

Table 5.I.1. Comparison between MNL and MMNL

	MNL DCE - NoAddTax		MNL DCE - AddTax		MMNL DCE - NoAddTax				MMNL DCE - AddTax			
Variable	**Coefficient (s.e)**	**t-ratio**	**Coefficient (s.e)**	**t-ratio**	**Coefficient (s.e)**	**t-ratio**	**Standard dev. (s.e)**	**t-ratio**	**Coefficient (s.e)**	**t-ratio**	**Standard dev. (s.e)**	**t-ratio**
asc	0.076(0.042)	1.807	-0.013(0.041)	-0.312	0.094(0.054)	1.766	0.190(0.139)	1.330	-0.003(0.058)	-0.049	0.236(0.137)	1.754
Additional years of life												
1 year of extra life	0.532(0.067)	7.943	0.424(0.099)	4.269	0.711(0.085)	7.990	-0.025(0.396)	-0.106	0.731(0.137)	5.411	-0.194(0.310)	-0.678
3 years of extra life	1.347(0.105)	12.793	1.467(0.171)	8.598	1.891(0.174)	9.979	1.100(0.158)	6.542	2.288(0.265)	9.023	1.330(0.218)	6.058
Quality of life improvements												
40 points improvement	0.488(0.101)	4.840	0.380(0.11)	3.450	0.690(0.134)	4.681	0.015(0.207)	0.265	0.742(0.160)	4.706	0.008(0.176)	0.133
60 points improvement	0.985(0.103)	9.576	0.765(0.137)	5.591	1.313(0.14)	8.416	-0.482(0.224)	-1.784	1.096(0.185)	5.241	-0.603(0.172)	-3.659
Patient experience												
Fair	1.259(0.095)	13.287	1.044(0.148)	7.049	1.610(0.137)	9.419	0.615(0.198)	2.832	1.403(0.218)	6.790	-0.911(0.161)	-5.576
Good	1.472(0.130)	11.334	1.224(0.114)	10.724	1.880(0.192)	8.665	-1.147(0.209)	-4.686	2.02(0.206)	9.213	1.430(0.174)	7.885
Size of target population												
5,000 /per 100,000 citizens	0.737(0.103)	7.180	1.006(0.155)	6.506	1.068(0.140)	6.637	-0.335(0.315)	-1.243	1.204(0.207)	6.982	-0.033(0.161)	-0.932
10,000 /per 100,000 citizens	1.161(0.081)	14.292	1.285(0.154)	8.356	1.672(0.151)	9.040	-1.159(0.145)	-7.424	1.879(0.232)	7.471	-1.207(0.136)	-8.576
Target population												
50% disadvantaged	0.555(0.100)	5.576	0.153(0.100)	1.521	0.768(0.139)	5.520	-0.565(0.243)	-2.373	0.311(0.140)	2.129	-0.115(0.316)	-0.535
75% disadvantaged	0.505(0.066)	7.703	0.249(0.169)	1.475	0.714(0.111)	5.674	-0.959(0.127)	-6.633	0.256(0.222)	1.282	0.329(0.265)	1.450
Additional budget needed												
£40 additional tax / per year			-0.506(0.107)	-4.741					-0.737(0.144)	-5.381	0.047(0.231)	0.564
£60 additional tax / per year			-0.963(0.140)	-6.866					-1.491(0.214)	-6.754	1.265(0.143)	8.319
Log likelihood (final)	-1658.12		-1728.42		-1587.49				-1611.79			
AIC	3338.23		3482.84		3218.98				3275.59			
BIC	3403.6		3560.1		3349.71				3430.09			

Note: ASC,Alternative-specific constant; s.d.,standard deviation; s.e., standard error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion

Table 5.I.2. Parameter coefficients from the mixed logit model (MMNL) - QALY variable

	MMNL DCE - NoAddTax					MMNL DCE - AddTax				
Variable	**Coefficient (s.e)**	**Confidence interval**	**t-ratio**	**Standard dev. (s.e)**	**t-ratio**	**Coefficient (s.e)**	**Confidence interval**	**t-ratio**	**Standard dev. (s.e)**	**t-ratio**
asc	0.085 (0.05)	-0.013 to 0.184	1.692	-0.098 (0.389)	-0.151	-0.007 (0.053)	-0.113 to 0.099	-0.137	-0.249 (0.127)	-1.777
QALY gains	1.220 (0.138)	0.925 to 1.515	8.097	0.952 (0.133)	7.237	0.986 (0.15)	0.707 to 1.266	6.914	-0.706 (0.206)	-3.773
Patient experience-Exp										
Fair	1.479 (0.129)	1.163 to 1.795	9.173	-0.455 (0.193)	-2.299	0.849 (0.135)	0.569 to 1.128	5.948	1.079 (0.13)	8.612
Good	1.527 (0.169)	1.167 to 1.887	8.320	-0.824 (0.195)	-3.835	1.296 (0.158)	0.979 to 1.614	8.012	-1.326 (0.161)	-8.004
Size of target population-Size										
5,000 /per 100,000 citizens	0.858 (0.117)	0.610 to 1.106	6.783	0.129 (0.338)	0.587	1.268 (0.139)	0.999 to 1.537	9.241	-0.014 (0.161)	-0.459
10,000 /per 100,000 citizens	1.519 (0.145)	1.172 to 1.866	8.578	1.137 (0.137)	7.701	1.354 (0.159)	1.000 to 1.708	7.500	-1.149 (0.134)	-8.319
Target population-Equ										
50% disadvantaged	0.958 (0.121)	0.722 to 1.194	7.955	-0.393 (0.247)	-1.686	0.092 (0.108)	-0.128 to 0.312	0.817	-0.078 (0.277)	-0.448
75% disadvantaged	0.624 (0.102)	0.398 to 0.850	5.414	0.866 (0.119)	5.931	-0.109 (0.145)	-0.391 to 0.173	-0.757	0.095 (0.966)	0.069
Additional budget needed-AddTax										
£40 additional tax / per year						-0.196 (0.114)	-0.415 to 0.023	-1.75438	-0.154 (0.229)	-1.1429
£60 additional tax / per year						-0.738 (0.144)	-1.029 to -0.447	-4.97527	-1.117 (0.129)	-8.3961
Log likelihood (final)	-1637.26					-1657.93				
AIC	3306.53					3355.85				
BIC	3401.61					3474.7				

Note: ASC,Alternative-specific constant; s.d.,standard deviation; s.e., standard error; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion. The Root-Log likelihood (RLH) test reflects the fit of the choice model on the respondents' choices. The probability of an RLH value being <= 1/k was computed, with k denoting the number of alternatives per choice tasks. In all models the Pr (RLH ≤ 0.5) is close to zero.

Table 5.I.3. Comparison between DCE-NoAdd Tax & DCE-AddTax of MRS

Additional Years of Life (YoL) in the denominator	**t-ratio**
Ratio of QoL and YoL	-4.74
Ratio of (Exp_fair - Exp_poor) and YoL	-3.96
Ratio of (Size_5000 - Size_50) and YoL	-0.61
Ratio of (Equ_50 - Equ_25) and YoL	-42.37

Improvements in quality of life (QoL) in the denominator	**t-ratio**
Ratio of YoL and QoL	-4.69
Ratio of (Exp_fair - Exp_poor) and QoL	-0.20
Ratio of (Size_5000 - Size_50) and QoL	13.39
Ratio of (Equ_50 - Equ_25) and QoL	-38.96

Note: One-tailed t-test for two normally distributed but independent samples, with equal means as the null-hypothesis.

Appendix 5.J.Correlated mixed logit models & comparison of RI scores

Table. 5.J.1. Results of mixed logit models with correlated random parameters

	MNL DCE - AddTax				MNL DCE - NoAddTax			
Variable	**Estimate**	**s.e.**	**t-ratio**	**p-value**	**Estimate**	**s.e.**	**t-ratio**	**p-value**
asc_mu	0.00	0.06	0.07	0.47	0.09	0.05	1.79	0.04
asc_sig	0.33	0.17	1.98	0.02	-0.15	0.19	-0.77	0.22
Additional years of life (YoL)								
Mean-1 year of extra life	0.70	0.14	5.15	0.00	0.67	0.09	7.74	0.00
Standard dev-1 year of extra life	0.47	0.17	2.69	0.00	0.13	0.35	0.37	0.36
Mean-3 year of extra life	2.08	0.25	8.35	0.00	1.87	0.19	10.02	0.00
Standard dev-3 year of extra life	-1.35	0.21	-6.52	0.00	-1.03	0.17	-6.09	0.00
Quality of life improvements (QoL)								
Mean-40 points improv.	0.73	0.16	4.69	0.00	0.75	0.16	4.79	0.00
Standard dev-40 points improv.	0.04	0.07	0.60	0.27	-0.05	0.07	-0.70	0.24
Mean-60 points improv.	0.98	0.23	4.29	0.00	1.31	0.16	7.96	0.00
Standard dev-60 points improv.	-0.61	0.19	-3.13	0.00	0.46	0.31	1.47	0.07
Patient experience (Exp)								
Mean-Fair	1.29	0.22	5.87	0.00	1.62	0.16	10.19	0.00
Standard dev - Fair	-0.79	0.17	-4.56	0.00	0.56	0.20	2.85	0.00
Mean-Good	1.96	0.24	8.23	0.00	1.85	0.21	8.69	0.00
Standard dev - Good	0.77	0.27	2.87	0.00	0.96	0.26	3.67	0.00
Size of target population (Size)								
Mean-5,000/per 100,000	1.17	0.19	6.04	0.00	1.03	0.16	6.48	0.00
Standard dev-5,000/per 100,000	-0.04	0.05	-0.88	0.19	0.37	0.39	0.93	0.18
Mean-10,000/per 100,000	1.75	0.27	6.50	0.00	1.64	0.18	9.16	0.00
Standard dev-10,000/per100,000	-0.80	0.22	-3.71	0.00	-0.33	0.25	-1.32	0.09
Target population (Equ)								
Mean-50% disadvantaged	0.31	0.15	2.07	0.02	0.73	0.14	5.08	0.00
Standard dev-50% disadvantaged	-0.12	0.47	-0.25	0.40	0.53	0.25	2.11	0.02
Mean-75% disadvantaged	0.10	0.21	0.49	0.31	0.66	0.13	5.10	0.00
Standard dev-75% disadvantaged	0.14	0.47	0.29	0.39	0.45	0.51	0.88	0.19
Additional budget needed (AddTax)								
Mean-£40 add tax	-0.67	0.15	-4.54	0.00				
Standard dev-£40 add tax	0.19	0.55	0.35	0.36				
Mean-£60 add tax	-1.39	0.24	-5.68	0.00				
Standard dev-£60 add tax	0.59	0.39	1.51	0.07				
Covariance of random parameters								
QoL_6 & YoL_3	-0.22	0.24	-0.91	0.18	0.19	0.17	1.11	0.13
Exp_Good & YoL_3	0.88	0.28	3.19	0.00	0.45	0.26	1.71	0.04
Exp_Good & QoL_6	0.43	0.30	1.41	0.08	-0.19	0.40	-0.47	0.32

Size_10,000 & YoL_3	-0.43	0.19	-2.30	0.01	0.22	0.22	1.01	0.16
Size_10,000 & QoL_6	0.24	0.31	0.79	0.22	0.02	0.50	0.05	0.48
Size_10000 & Exp_Good	-0.72	0.21	-3.39	0.00	-1.05	0.16	-6.68	0.00
Equ_75 & YoL_3	-0.23	0.21	-1.11	0.13	0.55	0.14	3.81	0.00
Equ_75 & QoL_6	-0.60	0.16	-3.63	0.00	0.51	0.32	1.60	0.05
Equ_75 & Exp_Good	-0.10	0.21	-0.50	0.31	-0.19	0.27	-0.73	0.23
Equ_75 & Size_10,000	-0.02	0.24	-0.09	0.46	0.40	0.36	1.12	0.13
AddTax_60 & YoL_3	-0.07	0.23	-0.29	0.38				
AddTax_60 & QoL_6	-0.07	0.24	-0.31	0.38				
AddTax_60 & Exp_Good	0.33	0.21	1.56	0.06				
AddTax_60 & Size_10,000	-0.80	0.24	-3.35	0.00				
AddTax_60 & Equ_75	-0.54	0.27	-2.01	0.02				
Log likelihood (final)		-1584.6				-1566.18		
AIC		3251.2				3196.36		
BIC		3494.83				3386.52		

Appendices – Chapter 6

Appendix 6.A. Dataset structure & transformation

Figure 6.A.1. Example of original structure of the dataset

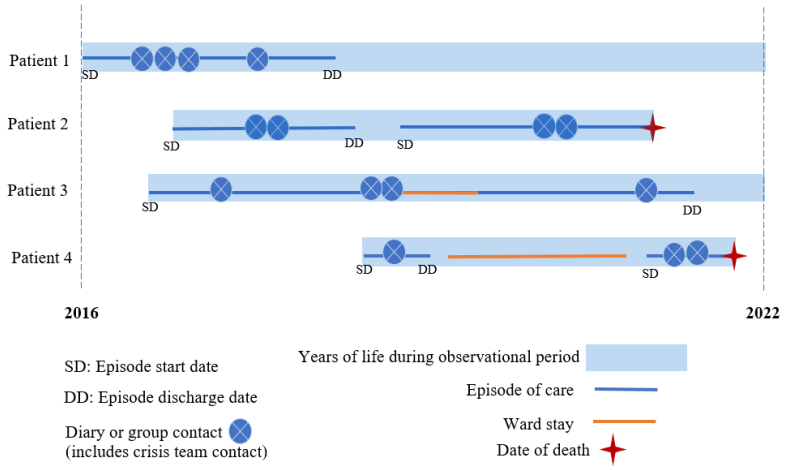
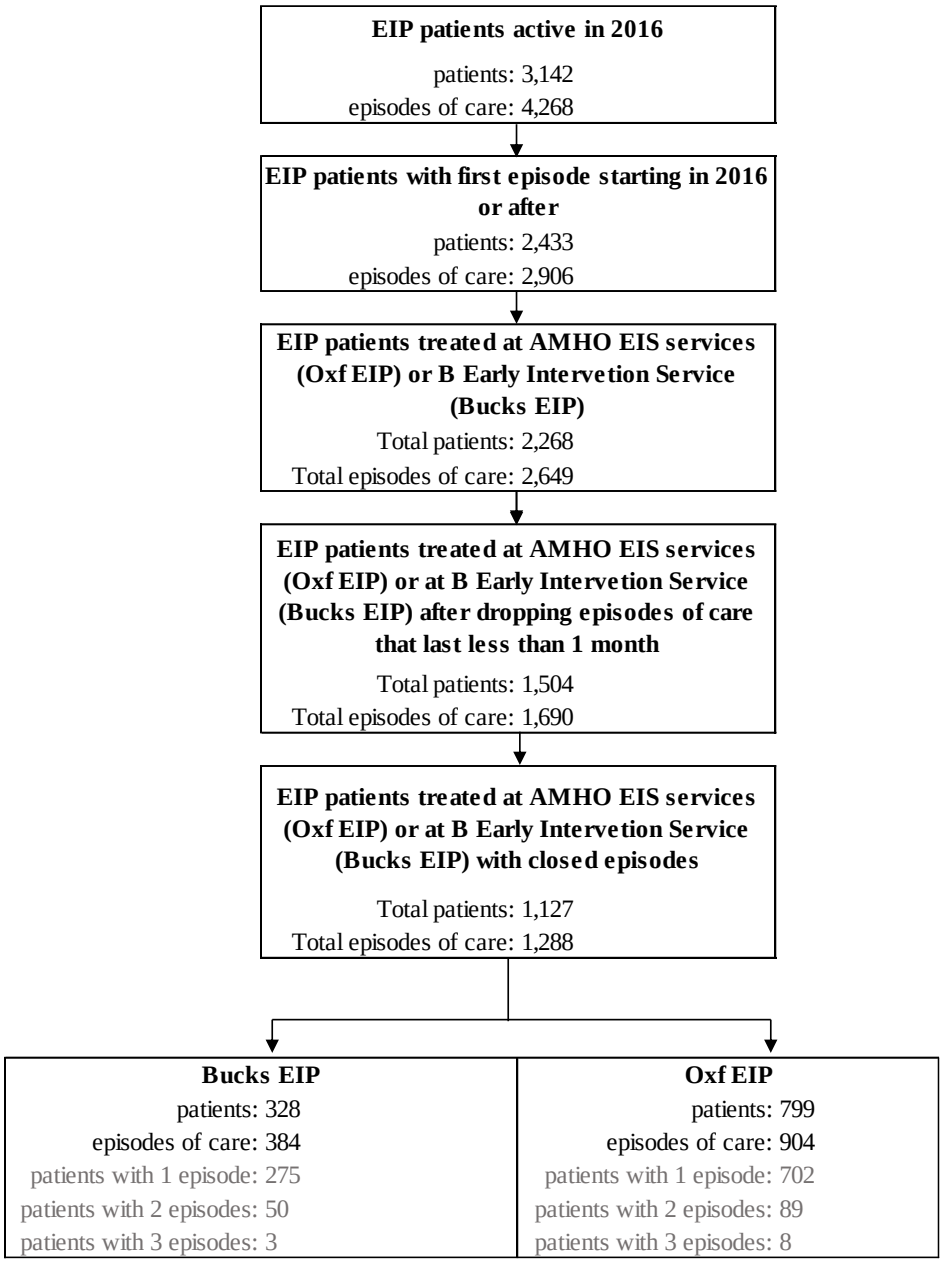


Figure 6.A.2. Steps followed for sample identification



Appendix 6.B. Selection of performance indicators

Framework's assessment criteria	Performance indicators considered at first	Assessment in light of data availability & Feedback from OHFT psychiatrists	Performance indicators selected
1. Final or intermediate health outcomes	Years of life during the observational period.	This indicator was favoured over intermediate health outcomes due to data availability reasons. To estimate scores on this indicator, the dataset includes the date of birth, date of death, start and discharge dates from EIP services. In discussions with clinicians, it was agreed that although the dataset only included information for six years, and patients could not be observed along their lifespans, this variable could be considered a good proxy of patient longevity.	Years of life during the observational period
2. Quality of life and well-being considerations	the Health of the Nation Outcome Scales (HoNOS) score	HoNOS is a clinical instrument used to assess the health status and social functioning of people with SMI (241). This instrument consists of 12 scales designed to assess behavioural problems, impairment, symptomatic problems, and social problems. The total score is obtained by summing the scores from each of the 12 individual scales. Each scale ranges from 0 (no problem) to 4 (severe to very severe problem), making the total HoNOS score range between 0 and 48. Although the CRIS dataset included HoNOS variables, several patients had no HoNOS score (% 51 of the sample). Similarly, DIALOG data was also explored, as it is a service user-rated outcome and experience measure, that focuses on quality of life, care needs and treatment satisfaction (270). There were however few patients with DIALOG records in the dataset.	Average time-to-relapse (days) across all relapses per patient

	Time to relapse after an EIS episode	This indicator refers to the time to relapse immediately after discharge from the first episode (i.e. days between discharge from the first episode and contact with the crisis team or admission to the hospital). This indicator was not chosen as it captured the legacy effect of the EIP services rather than patients' quality of life over the observational period.	
	Average time-to-relapse across all relapses per patient.	Reducing relapses aligns with the goals of EIP services, as outlined in the NHS England Guidelines (238, 381). The use of this metric is also supported by previous research using it when assessing the effectiveness of early interventions or cognitive–behavioural therapy to treat patients with psychosis (377, 383, 384, 465). According to clinicians, this approach was reasonable. Previous studies have also used other indicators, such as reducing symptom severity or functionality/social functioning (242) (383) (270), but they were not considered due to the lack of data.	
3. Patient experience	Waiting time to first or second contact with EIP services	The "waiting time from referral to treatment" indicator was considered due to related standards defined in NHS England Guidelines (238, 270, 381). However, clinicians working with CRIS datasets did not recommend it due to concerns about data quality. For example, the first "contact" registered in the dataset after the start of the referral or the start of the episode of care could indicate a second assessment rather than treatment.	Disengagement with the service
	Treatment discontinuation/ Disengagement from the service.	This indicator has been used by previous studies in the field (377) and was validated with the clinicians. Estimating performance scores on this indicator was possible as the variables needed were available (i.e., start date of episode of care, contact with EIP services). Following the literature and clinical advice, a patient was classified as having disengaged from EIP services if there had been	

		no contact for three consecutive months following the episode's start date.	
4. Equity	Ratio: Average time-to-relapse (deprived decile)/ Average time-to-relapse (richer decile Ratio: Average time-to-relapse (Black&Mixed)/ Average time-to-relapse (White)	Studies have mainly explored the heterogeneity of results on health-related outcomes across different socio-economic groups or have analysed inequalities graphically (381, 650). Health disparities are typically expressed as relative measures (i.e. ratios) obtained by comparing the groups with the lowest and highest levels of health (388, 389). A review showed that using relative measures of the effect remains a common practice: almost 90% of studies reporting health disparities used comparable measures (388). Since ethnicity data was unavailable for 259 patients (23% of the sample), the deprivation deciles ratio was favoured over the one based on ethnicity groups.	Ratio of the average time-to-relapse between the most deprived deciles and the richest decile.
5. Cost	Annual average cost across the observational period of EIP patients	In light of data available and taking into account the NHS National Cost Collection for mental health services 2022 (391), annual healthcare costs per patient included: (i) cost of initial assessment, which is done at the start of any EIP episode of care; (ii) cost of each contact with mental health services, based on Patient Level Information and Costing System (PLICS); and (iii) cost per bed day (hospitalisation). To calculate this variable for each patient, the cost of each year is first calculated and then summed up across the observational period. Afterwards, this value is divided by the years of life during the observational period. Following clinical advice, all attended contacts in addition to contacts that were cancelled by patients were taken into account. A similar approach has been used in previous studies (242) (384).	Annual average cost across the observational period of EIP patients

Appendix 6.C. Performance indicators

Table 6.C.1. Outcome variables before PSM

CRITERIA		Descriptive statistics before PSM	
Framework's assessment criteria	**Performance indicator**	**Bucks EIP N=328**	**Oxf EIP N=799**
1. Final health outcomes	Years of life during the observational period, mean (sd)	4.08 (1.55)	3.96 (1.57)
2. Quality of life and well-being considerations	Average time-to-relapse (days) across all relapses per patient, mean (sd)	389.47 (452.02)	321.76 (385.71)
3. Quality of care or features of the process of care delivery	Disengagement with the service, % (n)	9.45% (31)	12.89% (103)
4.Equity	Ratio of the average time-to-relapse between the most deprived deciles and the richest decile, (se)*	0.50 (0.18)	0.83 (0.19)
5. Cost or budget considerations	Annual average cost across the observational period of EIP patients	£ 5,909.28	£ 4,509.86

*bootstrapped standard errors (s.e.) with 1,000 repetitions.

Figure 6.C.1. Years of life during the observational period: Histogram



Figure 6.C.2. Average time-to-relapse (days) across all relapses per patient, mean: Histogram

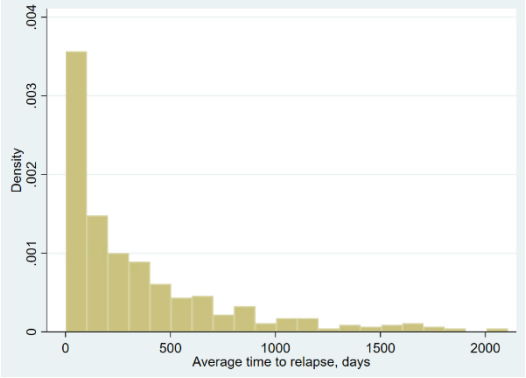


Figure 6.C.3. Ratio of the average time-to-relapse between the most deprived deciles and the richest decile

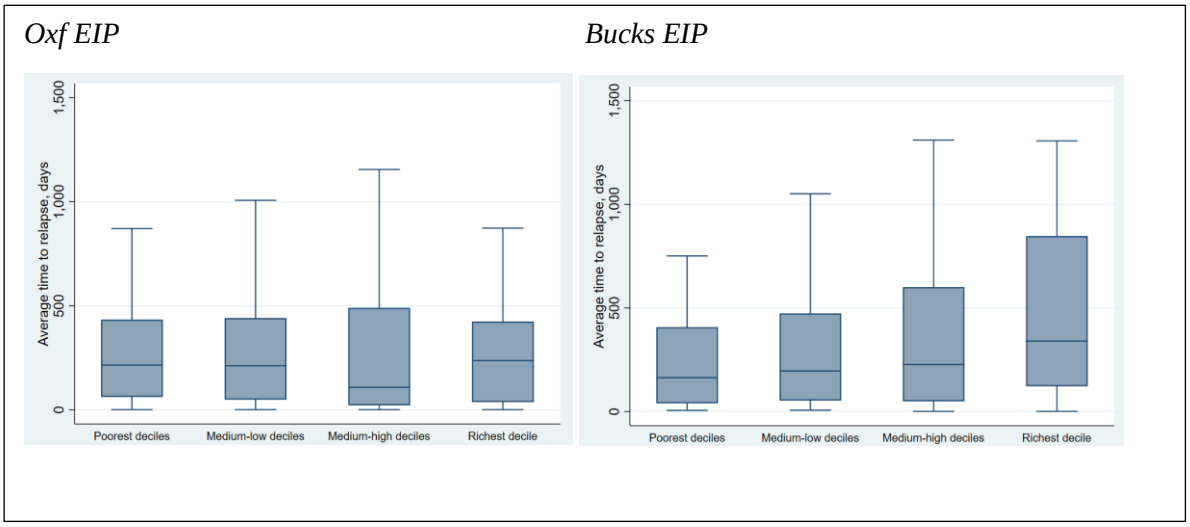
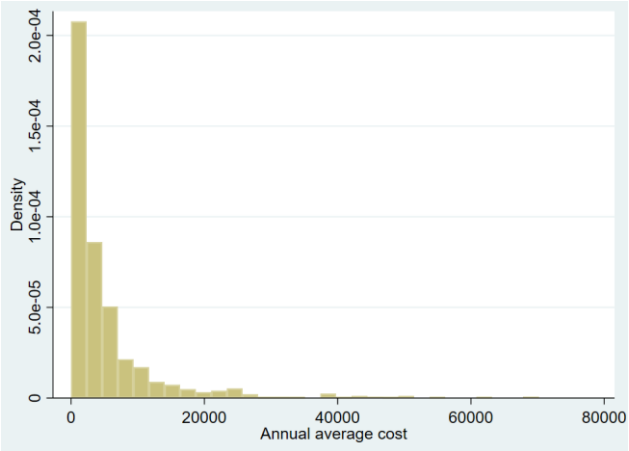


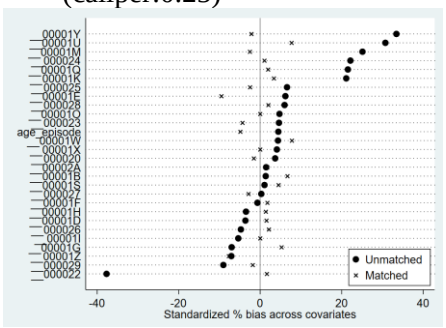
Figure 6.C.4. Annual average cost across the observational period of EIP patients: Histogram



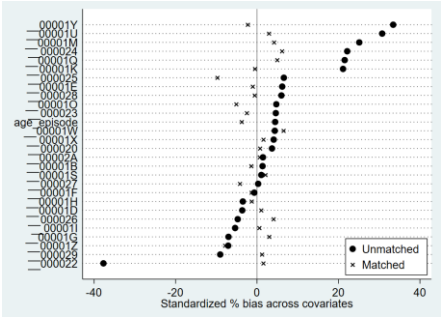
Appendix 6.D. Selection of Propensity Score Matching (PSM) technique

Figure 6.D.1. Standardised mean difference before & after PSM, for all techniques explored

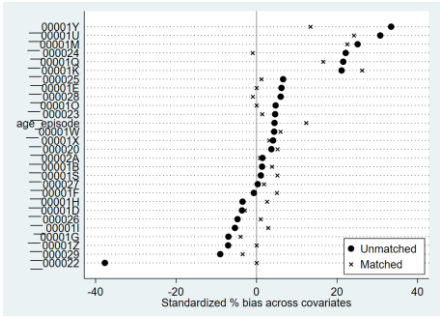
1. Nearest Neighbour without replacement (caliper:0.25)



2. Nearest Neighbor with replacement (caliper:0.25)



3. Mehalanobis



4. Kernel

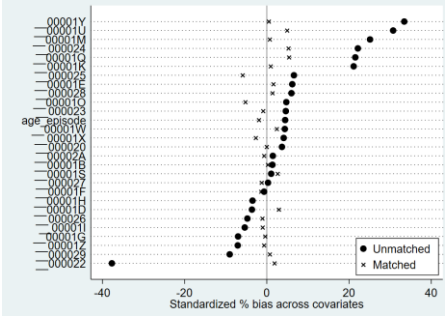
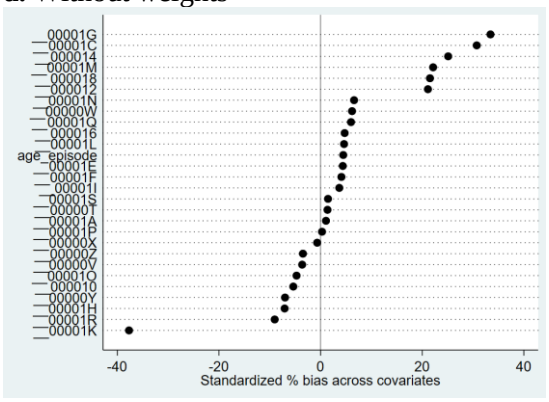


Figure 6.D.2. Standardised mean difference before & after Stabilized Inverse Probability Weighting

a. Without weights



b. With weights

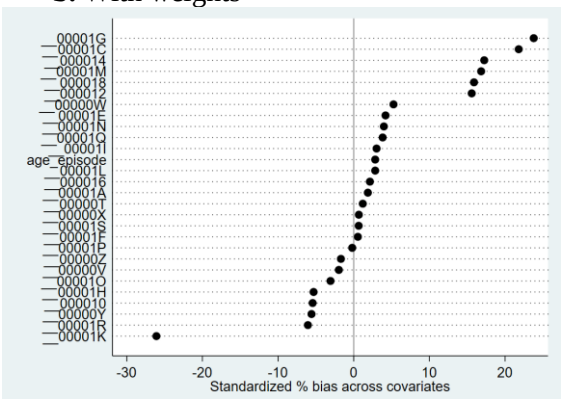


Figure 6.D.3. Results of PSM with kernel

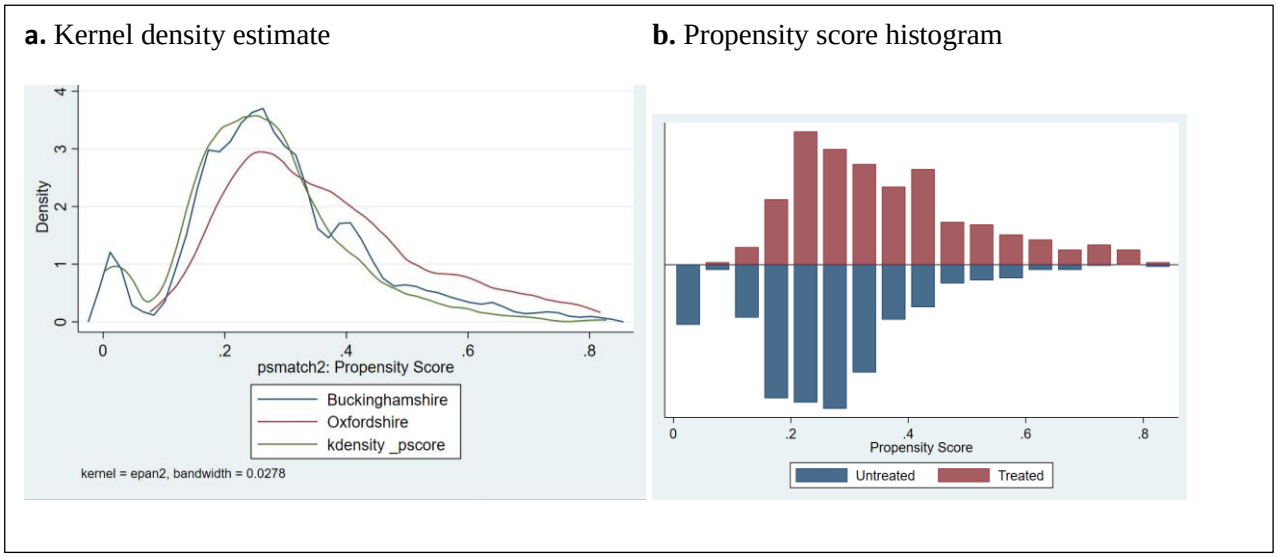


Table 6.D.1. Descriptive statistics of the weights of matched controls

	Percentiles	Smallest		
1%	0.0017	0.0005		
5%	0.1221	0.0006		
10%	0.1668	0.0006	Obs.	740
25%	0.2595	0.0009	Sum of weights.	740
50%	0.3795		Mean.	0.4297
		Largest	Standard Dev.	0.36
75%	0.3989	2.3846		
90%	0.8108	2.6134	Variance.	0.1296
95%	0.8749	3.1633	Skewness.	5.1822
99%	1.8829	5.1814	Kurtosis.	52.237

Appendix 6.E. GMLs performance

Table 6.E.1. Selection of GLM models

CRITERIA		Distribution family	Link function	AIC	BIC	Selected
Framework's assessment criteria	**Outcome measure**					
Final or intermediate health outcomes	Years of life during observational period	Gamma	Log	2.838822	-7134.968	
		Gaussian	**Log**	**1.125479**	**-6970.337**	**X**
		Gaussian	Identity	1.130227	-6968.84	
		Gamma	Identity	2.838577	-7135.227	
Quality of life and well-being considerations	Average time-to-relapse (days) across all relapses per patient	**Gamma**	**Log**	**9.381122**	**-2129.967**	**X**
		Gaussian	Log	10.2263	4.06e+07	
		Gamma	Identity	9.360192	-2143.703	
		Gaussian	Identity	10.28112	4.42e+07	
Quality of care or features of the process of care delivery	Disengagement with the service	**Binomial**	**Logit**	**0.373120**	**-6824.371**	**X**
		Binomial	Probit	0.373613	-6823.849	
Cost or budget considerations	Annual average cost across the observational period of EIP patients (£)	**Gamma**	**Log**	**11.29026**	**-6437.993**	**X**
		Gaussian	Log	12.53163	3.28e+10	
		Gamma	Identity	Convergence not achieved		
		Gaussian	Identity	Convergence not achieved		

Akaike Information Criterion (AIC); Bayesian Information Criterion (BIC); (a) GLM on a subsample with patients that started the episode of care before 2022. The log link function is typically used when the dependent variable is continuous and strictly positive, as it ensures that predictions will also be positive. For 'Disengagement with the service', which is a binary variable, the logit and probit link functions were used, which are the most commonly used link functions for binary outcome variables in a GLM. All GLM models included the following covariates: age at the start of the episode of care, gender, year of start of the episode of care, mental health diagnoses variables, ethnicity and deprivation deciles.

Table 6.E.2. GLM regression outputs

Performance indicator: Years of life during observational period

Variable	Coefficient*	s.e.	p-value	95% Confidence interval
Bucks_EIS	0.999	0.012	0.931	0.976 - 1.022
Age	0.999	0.000	0.187	0.998 - 1.000
Female	1.012	0.011	0.269	0.991 - 1.033
Year				-
2017	0.852	0.012	0.000	0.830 - 0.875
2018	0.691	0.010	0.000	0.671 - 0.710
2019	0.527	0.009	0.000	0.510 - 0.545
2020	0.366	0.008	0.000	0.350 - 0.383
2021	0.192	0.007	0.000	0.178 - 0.207
2022	0.071	0.006	0.000	0.059 - 0.085
Mental health diagnosis				

Schizophrenia	0.978	0.012	0.067	0.956	-	1.002
Affective D.	0.996	0.015	0.796	0.968	-	1.025
Organic D.	0.989	0.030	0.713	0.931	-	1.050
D. substance use	0.997	0.021	0.906	0.957	-	1.040
Other D.	1.020	0.009	0.028	1.002	-	1.038
Other non-mental D.	1.009	0.016	0.573	0.978	-	1.040
Ethnicity						
Black	1.003	0.017	0.852	0.970	-	1.038
Mixed	0.998	0.015	0.897	0.970	-	1.027
Asian	1.001	0.022	0.976	0.958	-	1.045
Other	1.025	0.021	0.232	0.985	-	1.066
Not Reported	0.969	0.023	0.197	0.925	-	1.016
Multi deprivation index decile						
2	1.049	0.037	0.177	0.979	-	1.125
3	0.994	0.025	0.823	0.947	-	1.044
4	0.977	0.029	0.435	0.923	-	1.035
5	0.968	0.030	0.290	0.912	-	1.028
6	0.987	0.025	0.601	0.939	-	1.037
7	0.973	0.025	0.297	0.925	-	1.024
8	0.948	0.031	0.103	0.888	-	1.011
9	0.987	0.024	0.580	0.941	-	1.034
10	0.983	0.024	0.480	0.938	-	1.030
Log pseudolikelihood:	-565.379		AIC:	1.125479		
N. of observations:	1,058		BIC:	-6970.337		

GLM with Gaussian distribution and log link function.*Exponentiation of the coefficient. Age refers to age at the start of the episode of care; Year to the year the episode of care started; D indicates mental health disorders; s.e., robust standard error.

Performance indicator: Average days-to-relapse across all relapses per patient

Variable	Coefficient*	s.e.	p-value	95% Confidence interval		
Bucks_EIS	1.203	0.145	0.125	0.950	-	1.522
Age	1.007	0.005	0.199	0.996	-	1.018
Female	1.148	0.128	0.216	0.923	-	1.429
Year					-	
2017	0.957	0.156	0.788	0.695	-	1.318
2018	0.769	0.128	0.116	0.554	-	1.067
2019	0.414	0.084	0.000	0.278	-	0.617
2020	0.412	0.889	0.000	0.271	-	0.629
2021	0.293	0.116	0.002	0.135	-	0.636
2022	0.038	0.008	0.000	0.025	-	0.057
Mental health diagnosis						
Schizophrenia	0.793	0.116	0.113	0.595	-	1.057
Affective D.	0.738	0.097	0.018	0.563	-	0.947
Organic D.	1.007	0.324	0.982	0.537	-	1.889
D. substance use	0.997	0.199	0.989	0.675	-	1.474

Other D.	1.855	0.146	0.697	0.885	-	1.384
Other non-mental D.	0.857	0.141	0.346	0.621	-	1.183
Ethnicity						
Black	0.685	0.147	0.879	0.449	-	1.044
Mixed	1.966	0.385	0.001	1.340	-	2.886
Asian	0.913	0.180	0.643	0.620	-	1.343
Other	1.674	0.425	0.043	1.017	-	2.755
Not Reported	0.856	0.169	0.430	0.581	-	1.260
Multi deprivation index decile						
2	1.323	0.727	0.611	0.458	-	3.886
3	1.167	0.636	0.776	0.481	-	3.399
4	1.396	0.748	0.533	0.489	-	3.987
5	1.877	0.616	0.896	0.351	-	3.305
6	1.315	0.725	0.619	0.447	-	3.874
7	1.814	0.548	0.979	0.357	-	2.882
8	1.085	0.593	0.881	0.372	-	3.167
9	1.243	0.662	0.683	0.437	-	3.531
10	1.338	0.695	0.575	0.483	-	3.706
Log pseudolikelihood:	-2122.968		AIC:	9.381122		
N. of observations:	459		BIC:	-2129.967		

GLM with Gamma distribution and log link function. *Exponentiation of the coefficient. Age refers to age at the start of the episode of care; Year to the year the episode of care started; D indicates mental health disorders; s.e., robust standard error.

Performance indicator: Disengagement with the service

Variable	Coefficient*	s.e.	p-value	95% Confidence interval		
Bucks_EIS	0.966	0.232	0.884	0.693	-	1.547
Age	0.980	0.012	0.085	0.957	-	1.003
Female	0.899	0.227	9.648	0.541	-	1.466
Year					-	
2017	1.483	0.652	0.370	0.627	-	3.519
2018	1.440	0.640	0.412	0.693	-	3.439
2019	1.493	0.614	0.330	0.667	-	3.341
2020	0.564	0.295	0.274	0.293	-	1.572
2021	1.024	0.579	0.967	0.344	-	3.952
2022	2.979	2.663	0.222	0.516	-	17.192
Mental health diagnosis						
Schizophrenia	0.214	0.768	0.000	0.106	-	0.432
Affective D.	0.429	0.191	0.058	0.179	-	1.028
Organic D.	13.812	11.513	0.002	2.696	-	70.976
D. substance use	1.187	0.833	0.807	0.300	-	4.792
Other D.	1.757	0.669	0.134	0.841	-	3.670
Other non-mental D.	0.664	0.387	0.482	0.212	-	2.079
Ethnicity						

Black	1.553	1.014	0.501	0.431	-	5.586
Mixed	3.933	1.980	0.006	1.473	-	10.561
Asian	0.426	0.394	0.232	0.105	-	1.724
Other	2.999	2.614	0.235	0.500	-	16.933
Not Reported	3.337	1.036	0.000	1.816	-	6.134
Multi deprivation index decile						
2	0.077	0.102	0.052	0.006	-	1.026
3	0.780	0.819	0.804	0.098	-	6.050
4	0.623	0.597	0.622	0.052	-	4.079
5	1.976	1.065	0.941	0.155	-	7.486
6	0.423	0.412	0.377	0.063	-	2.853
7	0.842	0.779	0.852	0.137	-	5.153
8	1.448	1.343	0.690	0.235	-	8.919
9	0.569	0.550	0.559	0.085	-	3.791
10	0.636	0.595	0.628	0.102	-	3.978
Log pseudolikelihood:	-167.380		AIC:	0.3731198		
N. of observations:	1,058		BIC:	-6824.371		

GLM with Binomial distribution and logit link function. *The exponentiated value of the log odds (i.e., odds ratio). Age refers to age at the start of the episode of care; Year to the year the episode of care started; D indicates mental health disorders; s.e., robust standard error.

Performance indicator: Annual average cost £

Variable	Coefficient *	s.e.	p-value	95% Confidence interval		
Bucks_EIS	1.195	0.093	0.022	1.026	-	1.392
Age	0.990	0.003	0.006	0.984	-	0.997
Female	0.924	0.084	0.383	0.774	-	1.103
Year					-	
2017	1.085	0.131	0.502	0.856	-	1.375
2018	1.642	0.198	0.000	1.296	-	2.080
2019	1.743	0.235	0.000	1.337	-	2.271
2020	2.781	0.403	0.000	2.094	-	3.694
2021	2.757	0.508	0.000	1.921	-	3.955
2022	1.276	0.458	0.498	0.631	-	2.579
Mental health diagnosis						
Schizophrenia	2.873	0.230	0.000	2.456	-	3.361
Affective D.	2.009	0.198	0.000	1.656	-	2.436
Organic D.	2.288	0.976	0.052	0.992	-	5.277
D. substance use	1.266	0.201	0.136	0.928	-	1.728
Other D.	1.334	0.143	0.007	1.082	-	1.645
Other non-mental D.	1.425	0.168	0.003	1.130	-	1.796
Ethnicity						
Black	1.111	0.174	0.502	0.817	-	1.511

Mixed	0.784	0.135	0.156	0.559	-	1.098
Asian	1.072	0.130	0.568	0.845	-	1.359
Other	1.092	0.292	0.743	0.646	-	1.844
Not Reported	0.419	0.059	0.000	0.319	-	0.552
Multi deprivation index decile						
2	0.643	0.334	0.395	0.232	-	1.779
3	0.736	0.287	0.432	0.343	-	1.580
4	0.567	0.196	0.101	0.288	-	1.118
5	0.571	0.199	0.109	0.288	-	1.132
6	0.592	0.201	0.123	0.304	-	1.152
7	0.662	0.227	0.228	0.338	-	1.295
8	0.522	0.185	0.067	0.260	-	1.047
9	0.499	0.168	0.039	0.258	-	0.965
10	0.562	0.188	0.084	0.292	-	1.081
Log pseudolikelihood:	-5942.549		AIC:	11.29026		
N. of observations:	1,058		BIC:	-6437.993		

GLM with Gamma distribution and log link function.*Exponentiation of the coefficient. Age refers to age at the start of the episode of care; Year to the year the episode of care started; D indicates mental health disorders; s.e., robust standard error.

References

1 OECD. Health at a Glance 2023: OECD Indicators, OECD Publishing, Paris; 2023. Avaiable from: https://doi.org/10.1787/7a7afb35-en

2 Robinson R. The global challenge of health care rationing. In: Coulter A, Ham C, editors. Buckingham: Open University Press; 2000. 267 p. ISBN 0 335 20463 5. Health Economics. 2001;10:86.

3 Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.

4 Norheim OF. The elusive challenge of priority setting in health and health care. Glob Chall. 2017;1(1):28-9.

5 Greer SL, Wismar M, Figueras J. Strengthening Health System Governance European Observatory on Health Systems and Policies Series. New York, USA: The European Observatory on Health Systems and Policies. World Health Organization. Open University Press; 2016.

6 Baker R, Mason H, McHugh N, Donaldson C. Public values and plurality in health priority setting: What to do when people disagree and why we should care about reasons as well as choices. Social Science & Medicine. 2021;277:113892.

7 NHS England and NHS Improvement. Building strong integrated care systems everywhere. ICS implementation guidance on working with people and communities. London; 2021 September 2, 2021.

8 Charles A. Integrated care systems explained [Internet]. 2021 [updated 2021 May 11; cited 2021 Oct 12]. Available from: https://www.kingsfund.org.uk/publications/integrated-care-systems-explained.

9 Department of Health and Social Care. Health and Care Act 2022 Core Measures. 19 July 2022.

10 UK Government. Health and Care Act 2022, c. 32. London: The Stationery Office; 2022.

11 UK Government. Health and Social Care Act 2012: fact sheets [Internet]. 2012 [updated 2012 June 15; cited 2021 Oct 12]. Available from: https://www.gov.uk/government/publications/health-and-social-care-act-2012-fact-sheets.

12 NHS England and NHS Improvement. Integrated Care Systems: design framework. London; June 2021.

13 NHS England and NHS Improvement. Working together at scale: guidance on provider collaboratives. London; 2021 Aug. Report No.: PAR754.

14 Appleton G, Holmes L, Rickatson F. Care England report. From Inception to Implementation: A Year of Integrated Care Systems. Care England; 2023. [cited 2023 Sep 20]. Available from: https://www.careengland.org.uk/from-inception-to-implementation-a-year-of-integrated-care-systems/.

15 Gröne O, Garcia-Barbero M. Integrated care: a position paper of the WHO European Office for Integrated Health Care Services. Int J Integr Care. 2001;1:e21.

16 Ham C, Curry N. Integrated care: What is it? Does it work? What does it mean for the NHS? [Internet]. King's Fund. 2010 [cited 2021 Jun 6]. Available from: https://www.kingsfund.org.uk/publications/articles/integrated-care-summary-what-it-does-it-work-what-does-it-mean-nhs.

17 Shaw S, Rosen R, Rumbold B. What is integrated care? . Research report. Nuffield Trust. London; 2011 June, 2011.

18 Tsiachristas A, Vrangbæk K, Gongora-Salazar P, Kristensen SR. Integrated care in a Beveridge system: experiences from England and Denmark. Health Econ Policy Law. 2023;18(4):345-61.

19 NHS England. Five Year Forward View. Published: 22 October 2014. Updated: 8 March 2019.

20 Gray M. Value based healthcare. BMJ. 2017;356:j437.

21 Hurst L, Mahtani K, luddemann A, Lewis S, Harvey K, Briggs A, et al. Defining Value-based Healthcare in the NHS. Centre for Evidence-Based Medicine Report. 2019;04:1–13.

22 WHO. Integrated care models: an overview. World Health Organization Regional Office for Europe, Health Services Delivery Programme DoHSaPH; 2016 October 2016. Report No.: 0035-9211.

23 Tsiachristas A, Stein KV, Evers S, Rutten-van Mölken M. Performing economic evaluation of integrated care: Highway to hell or stairway to heaven? Int J Integr Care. 2016;16(4)

24 NHS England. The NHS Long Term Plan. London; 2019 January, 2019.

25 Lewis S. Value-based healthcare: is it the way forward? Future Healthc J. 2022;9(3):211-5.

26 Williams I, Bryan S. Understanding the limited impact of economic evaluation in health care resource allocation: A conceptual framework. Health Policy. 2007;80(1):135-43.

27 Marks L, Hunter DJ, Scalabrini S, Gray J, McCafferty S, Payne N, et al. The return of public health to local government in England: Changing the parameters of the public health prioritization debate? Public Health. 2015;129(9):1194-203.

28 Addicott R. Challenges of commissioning and contracting for integrated care in the National Health Service (NHS) in England. Australian Journal of Primary Health. 2016;22:50-4.

29 Currie G, Croft C, Chen Y, Kiefer T, Staniszewska S, Lilford RJ. The capacity of health service commissioners to use evidence: a case study. Health Services and Delivery Research. 2018;6(12):1-198.

30 Rumbold B, Weale A, Rid A, Wilson J, Littlejohns P. Public Reasoning and Health-Care Priority Setting: The Case of NICE. Kennedy Institute of Ethics Journal 2017;27(1):107-34.

31 Mauskopf JA, Paul JE, Grant DM, Stergachis A. The Role of Cost—Consequence Analysis in Healthcare Decision—Making. PharmacoEconomics. 1998;13(3):277-88.

32 Peacock SJ, Richardson JRJ, Carter R, Edwards D. Priority setting in health care using multi-attribute utility theory and programme budgeting and marginal analysis (PBMA). Social Science & Medicine. 2007;64(4):897-910.

33 Mitton C, Peacock S, Donaldson C, Bate A. Using PBMA in health care priority setting: description, challenges and experience. Appl Health Econ Health Policy. 2003;2(3):121-7.

34 Nicholls J, Lawlor E, Neitzert E, Goodspeed T. A Guide to Social Return on Investment 2012. The SROI Network; 2012.

35 Arvidson M, Lyon F, McKay S, Moro D. Valuing the Social? The Nature and Controversies of Measuring Social Return on Investment (SROI). Voluntary Sector Review. 2013;4:3-18.

36 van den Bogaart EHA, Kroese MEAL, Spreeuwenberg MD, Ruwaard D, Tsiachristas A. Economic Evaluation of New Models of Care: Does the Decision Change Between Cost-Utility Analysis and Multi-Criteria Decision Analysis? Value in Health. 2021;24(6):795-803.

37 Hoedemakers M, Tsiachristas A, Rutten-van Mölken M. Moving Beyond Quality-Adjusted Life-Years in Elderly Care: How Can Multicriteria Decision Analysis Complement Cost-Effectiveness Analysis in Local-Level Decision Making. Value in Health. 2022;25(10):1717-25.

38 Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. Health Technol Assess. 2021;25:57.

39 Marsh K, Lanitis T Fau - Neasham D, Neasham D Fau - Orfanos P, Orfanos P Fau - Caro J, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics. 2014;32(4):345-65.

40 Frazão TDC, Camilo DGG, Cabral ELS, Souza RP. Multicriteria decision analysis (MCDA) in health care: A systematic review of the main characteristics and methodological steps. BMC Medical Informatics and Decision Making. 2018;18(1):1-16.

41 Adunlin G, Diaby V, Xiao H. Application of multicriteria decision analysis in health care: A systematic review and bibliometric analysis. Health Expectations. 2015;18(6):1894-905.

42 Khan I, Pintelon L, Martin H. The Application of Multicriteria Decision Analysis Methods in Health Care: A Literature Review. Medical Decision Making. 2021:0272989X211019040-0272989X.

43 Department of Health and Social Care. The Prioritisation Framework: making the most of your budget Public health2018 [Available from: https://www.gov.uk/government/publications/the-prioritisation-framework-making-the-most-of-your-budget.

44 Hunter DJ, Gray J, Maniatopoulos G, . Shifting the Gravity of Spending – Mark 2: Final Evaluation Report. Newcastle University and Northumbria University; 2019 February.

45 Health Economics Unit. Evaluating robotic surgery in the NHS [Internet]. 2023 [cited 2023 Sep 7]. Available from: https://healtheconomicsunit.nhs.uk/evaluating-robotic-surgery-in-the-nhs/.

46 Health Economics Unit. HEU's Jack Ettinger to speak at the HIMSS European Health Conference 2022 [Internet]. 2022 [updated 2022 Jun 16; cited 2023 Sep 7]. Available from: https://healtheconomicsunit.nhs.uk/jack-ettinger-at-himss-conference/.

47 Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making - An introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value in Health. 2016;19(1):1-13.

48 Marsh K, Ijzerman M, Thokala P, Baltussen R, Boysen M, Kaló Z, et al. Multiple criteria decision analysis for health care decision making - Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value in Health. 2016;19(2):125-37.

49 The King's Fund. How far has the government gone towards integrating care? [Internet]. 2015 [cited 2022 May 16]. Available from: https://www.kingsfund.org.uk/projects/verdict/how-far-has-government-gone-towards-integrating-care.

50 Miller R, Glasby J, Dickinson H. Integrated Health and Social Care in England: Ten Years On. Int J Integr Care. 2021 Oct 29;21(4):6. doi: 10.5334/ijic.5666. eCollection 2021 Oct-Dec.

51 Wenzel L, Robertson R. What is commissioning and how is it changing? [Internet]. King's Fund. 2019 [updated 2019 Sep 19; cited 2022 May 16]. Available from: https://www.kingsfund.org.uk/publications/what-commissioning-and-how-it-changing.

52 NHS England. NHS commissioning. What is commissioning? [Internet]. [cited 2021 May 16]. Available from: https://www.england.nhs.uk/commissioning/what-is-commissioning/.

53 Powell T, Gheera M, Foster D, Balogun B, Conway L. The Health and Care Bill [Bill 140 of 2021-22]. London: House of Commons Library; 2021 Jul 12 [cited 2021 Nov 4]. Available from: https://researchbriefings.files.parliament.uk/documents/CBP-9232/CBP-9232.pdf

54 World Health Assembly. World Health Assembly (69): Framework on integrated, people-centred health services: report by the Secretariat Geneva: World Health Organization; 2016.

55 Baxter S, Johnson M, Chambers D, Sutton A, Goyder E, Booth A. The effects of integrated care: a systematic review of UK and international evidence. BMC Health Services Research. 2018;18(1):350.

56 Coultas C, Kieslich K, Littlejohns P. Patient and public involvement in priority-setting decisions in England's Transforming NHS: An interview study with Clinical Commissioning Groups in South London sustainability transformation partnerships. Health expectations : an international journal of public participation in health care and health policy. 2019;22(6):1223-30.

57 Corbacho B, Pinto-Prades JL. Health economic decision-making: A comparison between UK and Spain. British Medical Bulletin. 2012;103(1):5-20.

58 Cossio-Gil Y, Omara M, Watson C, Casey J, Chakhunashvili A, Gutiérrez-San Miguel M, et al. The Roadmap for Implementing Value-Based Healthcare in European University Hospitals—Consensus Report and Recommendations. Value in Health. 2022;25(7):1148-56.

59 Green J, Thorogood N. Qualitative Methods for Health Research | SAGE Publications Ltd2018. 440- p.

60 Microsoft Corporation. Microsoft Teams (version 2017) [software]. Redmond (WA): Microsoft Corporation; 2017.

61 Bowling A. Research Methods In Health: Investigating Health And Health Services: Open University Press; 2014.

62 International QSR. NVivo Qualitative Data Analysis Software [Software]. 1999.

63 Jackson K, Bazeley P. Qualitative Data Analysis with NVivo. 3rd ed. London: SAGE Publications Ltd; 2019.

64 Shaw SE, Smith JA, Porter A, Rosen R, Mays N. The work of commissioning: a multisite case study of healthcare commissioning in England's NHS. BMJ Open. 2013;3(9):e003341.

65 Greener I, Harrington BE, Hunter DJ, Mannion R, Powell M. Reforming healthcare: What's the evidence? 1 ed: Bristol University Press; 2014.

66 UK Government. National Health Service and Community Care Act 1990, c. 19. London: The Stationery Office; 1990.

67 Nuffield Trust. NHS reform timeline. Health & Social Care Explained. 2020 [cited 2021 May 16]. Available from: https://www.nuffieldtrust.org.uk/health-and-social-care-explained/nhs-reform-timeline.

68 Glasby J, Dickinson H, Miller R. Partnership working in England: where we are now and where we've come from. Int J Integr Care. 2011;11(Spec Ed 10th Anniversary Edition):e002.

69 The Health Foundation. Policy Navigator. The history of the NHS, social care, and public health policy. London; 2020 [cited 2021 May 16]. Available from: https://navigator.health.org.uk/.

70 Smith J, Regen E, Shapiro J, Baines D. National evaluation of general practitioner commissioning pilots: lessons for primary care groups. Br J Gen Pract. 2000;50(455):469-72.

71 UK Government. Health Act 1999, c. 8. London: The Stationery Office; 1999.

72 Glendinning C, Hudson B, Hardy B, Young R. The Health Act 1999 section 31 partnership 'flexibilities'. In: Glasby J, Peck E, editors. Care Trusts: Partnership Working in Action. 1st Edition. London: CRC Press; 2002.

73 UK Government. Health and Social Care Act 2001, Chapter 15 London; 2001.

74 UK Government. National Health Service Reform and Health Care Professions Act 2002. Chapter 17. 2002.

75 UK Government. Health and Social Care (Community Health and Standards) Act 2003. Chapter 43. 2003.

76 UK Government. National Health Service Act 2006, Chapter 41 (2006).

77 UK Government. Local Government and Public Involvement in Health Act 2007, Chapter 28 (2007).

78 UK Government. Health and Social Care Act 2008, Chapter 14 London; 2008.

79 UK Government. Health Act 2009. Explanatory notes. Chapter 21. [Internet]. 2009 [cited 2021 Feb 23]. Available from: https://www.legislation.gov.uk/ukpga/2009/21/notes/contents

80 Glasby J, Le Grand J, Duffy S. A healthy choice? Direct payments and healthcare in the English NHS. Policy & Politics. 2009;37.

81 Ling T, Brereton L, Conklin A, Newbould J, Roland M. Barriers and facilitators to integrating care: experiences from the English Integrated Care Pilots. International journal of integrated care. 2012;12(5):1-12.

82 RAND Europe and Ernst&Young LLP. National Evaluation of the Department of Health's Integrated Care Pilots. RAND Corporation; 2012.

83 Powell Davies, G., Williams, A.M., Larsen, K., Perkins, D., Roland, M. and Harris, M.F. (2008), Coordinating primary health care: an analysis of the outcomes of a systematic review. Medical Journal of Australia, 188: S65-S68. https://doi.org/10.5694/j.1326-5377.2008.tb01748.x

84 Anderson M, Pitchforth E, Edwards N, Alderwick H, McGuire A, Mossialos E. The United Kingdom: Health system review. London: European Observatory on Health Systems and Policies; 2022.

85 Briggs ADM, Göpfert A, Thorlby R, Allwood D, Alderwick H, Adam D, et al. Integrated health and care systems in England: can they help prevent disease? Health J. 2020;2:13.

86 National Collaboration for Integrated Care and Support. Integrated Care and Support: Our Shared Commitment. May 2013 [cited 2021 May 16]. Available from: https://assets.publishing.service.gov.uk/media/5a7b082d40f0b66a2fc04a37/DEFINITIVE_FINAL_VERSION_Integrated_Care_and_Support_-_Our_Shared_Commitment_2013-05-13.pdf

87 Department of Health and Social Care. Care Act 2014 Part 1: factsheets. 2016 [updated 19 April 2016; cited 14 Sep 2022]. Available from: https://www.gov.uk/government/publications/care-act-2014-part-1-factsheets/care-act-factsheets.

88 Mattke S, Seid M, Ma S. Evidence for the effect of disease management: is $1 billion a year a good investment? Am J Manag Care. 2007 Dec;13(12):670-6. PMID: 18069910.

89 NHS England. New Care Models: Vanguards - developing a blueprint for the future of NHS and care services. [Sep 2016; cited 2021 Apr 14]. Available from: https://www.england.nhs.uk/wp-content/uploads/2015/11/new_care_models.pdf.

90 NHS England. Delivering the Forward View: NHS planning guidance 2016/17 – 2020/21. [Dec 2015 ; cited 2021 Apr 14]. Available from: https://www.england.nhs.uk/wp-content/uploads/2015/12/planning-guid-16-17-20-21.pdf.

91 The King's Fund. Sustainability and transformation plans (STPs) explained. 2017 [updated 21 February 2017; cited 2021 May 16]. Available from: https://www.kingsfund.org.uk/topics/integrated-care/sustainability-transformation-plans-explained.

92 NHS England. London Health and Social Care Devolution. Memorandum of Understanding. Published: November 2017. Available from: https://www.england.nhs.uk/london/wp-content/uploads/sites/8/2017/11/171115_Signed-Memorandum-of-Understanding-Report-VF.pdf.

93 NHS England and Improvement. NHS England and NHS Improvement: working closer together. 2018 [updated 27 March 2018]. Available from: https://www.england.nhs.uk/2018/03/nhs-england-and-nhs-improvement-working-closer-together/.

94 Charles A, Ewbank L, McKenna H, Wenzel L. The NHS long-term plan explained. The King's Fund; 2019 [updated 23 January 2019; cited 2021 May 17]. Available from: https://www.kingsfund.org.uk/publications/nhs-long-term-plan-explained.

95 Department of Health and Social Care. Health and Care Bill: Integrated Care Boards and local health and care systems 2021 [updated July 19, 2021; cited 2021 May 17] Available from: https://www.gov.uk/government/publications/health-and-care-bill-factsheets/health-and-care-bill-integrated-care-boards-and-local-health-and-care-systems.

96 NHS Confederation. The health and care bill. Briefing. London; 2021 Jul 7. Available from: https://www.nhsconfed.org/system/files/2021-07/The-health-and-care-bill_0.pdf.

97 Department of Health & Social Care. Guidance on the preparation of integrated care strategies [Internet]. 2022 [updated 2022 July 29; cited 2022 Nov 12]. Available from: https://www.gov.uk/government/publications/guidance-on-the-preparation-of-integrated-care-strategies/guidance-on-the-preparation-of-integrated-care-strategies.

98 Department of Health & Social Care. Better Care Fund framework 2022-23 published 2022 [updated 19 July 2022; cited 2022 Nov 12] Available from: https://www.gov.uk/government/news/better-care-fund-framework-2022-23-published.

99 NHS England. Next steps on the NHS Five Year Forward View. London; 2017 March, 2017.

100 Charles A, Wenzel L, Kershaw M, Ham C, Walsh N. A year of integrated care systems. Reviewing the journey so far. London: The King's Fund; 2018.

101 The King's Fund. The Health and Care Bill: six key questions 2021 [updated Sep 6, 2021; cited 14 Apr 2022]. Available from: https://www.kingsfund.org.uk/publications/health-and-care-bill-key-questions.

102 NHS England and NHS Improvement. Integrated Care Systems: Guidance. [Updated Sep 2, 2021; ; cited Apr 14 2022]. Available from: https://www.england.nhs.uk/publication/integrated-care-systems-guidance/.

103 Department of Health & Social Care. Annex B: glossary. Definitions of terms found in 'Guidance on the preparation of integrated care strategies' [Internet]. 2022 [updated 2022 July 29; cited 2023 Nov 22]. Available from: https://www.gov.uk/government/publications/guidance-on-the-preparation-of-integrated-care-strategies/annex-b-glossary.

104 Department of Health. Statutory Guidance on Joint Strategic Needs Assessments and Joint Health and Wellbeing Strategies. London: Department of Health; 2013 Mar 26. Available from: https://assets.publishing.service.gov.uk/media/5a7b88cced915d131105fdff/Statutory-Guidance-on-Joint-Strategic-Needs-Assessments-and-Joint-Health-and-Wellbeing-Strategies-March-2013.pdf

105 Department of Health & Social Care. Health and wellbeing boards – guidance 2022 [updated 22 November 2022]. Available from: https://www.gov.uk/government/publications/health-and-wellbeing-boards-guidance/health-and-wellbeing-boards-guidance.

106 NHS Commissioning Board. Commissioning support: Key facts. [Internet]. 2014 Jul 22 [cited 2022 May 16]. Available from: https://www.england.nhs.uk/wp-content/uploads/2014/07/nhs-comm-board-ann-rep-1314.pdf

107 Goddard M. Integrated care systems and equity: prospects and plans. J Integr Care. 2023;31(5):29-42.

108 NHS England and Improvement. Technical Guide to Allocation Formulae and Convergence: For 2022/23 to 2024/25 revenue allocations. [Internet]. 2021 Dec [cited 2022 May 11]. Available from: https://www.england.nhs.uk/wp-content/uploads/2023/01/allocations-2023-24-to-2024-25-technical-guide-to-formulae-v5.pdf

109 NHS England. About the Better Care Fund. [Internet]. [cited 2023 Nov 11]. Available from: https://www.england.nhs.uk/ourwork/part-rel/transformation-fund/better-care-fund/about-the-better-care-fund/

110 Department of Health & Social Care. 2023 to 2025 Better Care Fund policy framework [Internet]. [updated 2023 Apr 5; cited 2023 Nov 11]. Available from: https://www.gov.uk/government/publications/better-care-fund-policy-framework-2023-to-2025/2023-to-2025-better-care-fund-policy-framework

111 The King's Fund. Health and wellbeing boards (HWBs) explained. [Internet]. 2016 [cited 2021 May 15]. Available from: https://www.kingsfund.org.uk/publications/health-wellbeing-boards-explained

112 NHS England and NHS Improvement. Roadmap for integrating specialised services within Integrated Care Systems [Internet]. 2022 May 31 [cited 2023 Nov 11]; Contract No.: PAR1440. Available from: https://www.england.nhs.uk/wp-content/uploads/2022/05/PAR1440-specialised-commissioning-roadmap-addendum-may-2022.pdf

113 Charles A, Ewbank L, Naylor C, Walsh N, Murray R. Developing place-based partnerships: The foundation of effective integrated care systems The King's Fund 2021 [updated 20 April, 2021]. Available from: https://www.kingsfund.org.uk/publications/place-based-partnerships-integrated-care-systems.

114 Dunn P, Fraser C, Williamson S, Alderwick H. Integrated care systems: what do they look like? The Health Foundation. [Internet]. 2022 Jun 15 [cited 2023 Sep 2]. Available from: https://www.health.org.uk/publications/long-reads/integrated-care-systems-what-do-they-look-like

115 NHS England. Allocations 2022. [Internet]. [cited 2023 Sept 12]. Available from: https://www.england.nhs.uk/allocations/

116 Charles A, Murray R. Are ICSs getting off to the right start? The King's Fund. [Internet]. 2022 Jul 12 [cited 2023 Sep 2]. Available from: https://www.kingsfund.org.uk/blog/2022/07/are-ics-getting-right-start

117 NHS England. Advice on how to establish a remote 'total triage' model in general practice using online consultations [Internet]. 2020 Sep 15 [cited 2022 June 2]. Available from: https://www.england.nhs.uk/coronavirus/documents/advice-on-how-to-establish-a-remote-total-triage-model-in-general-practice-using-online-consultations/

118 Humphries R. Health and wellbeing boards and integrated care systems. The King's Fund. [Internet]. 2019 Nov 13 [cited 2022 May 18]. Available from: https://www.kingsfund.org.uk/publications/articles/health-wellbeing-boards-integrated-care-systems

119 NHS England. Equality, diversity and health inequalities [Internet]. [cited 2022 May 20]. Available from: https://www.england.nhs.uk/about/equality/

120 NHS England and NHS Improvement. Guidance on the preparation of integrated care board constitutions. 2022 May 13. Contract No.: PAR1551. Available from: https://www.england.nhs.uk/publication/integrated-care-systems-guidance/

121 NHS England. About NHS RightCare Intelligence [Internet]. [cited 2021 May 19]. Available from: https://www.england.nhs.uk/rightcare/products/about-nhs-rightcare-intelligence/.

122 Berwick DM, Nolan TW, Whittington J. The triple aim: Care, health, and cost. Health Affairs. 2008;27(3):759-69.

123 NHS Digital. Diabetes Transformation Metrics (DIAB) 2020. [Internet]. 2020 Mar 31 [cited 2021 May 19]. Available from: https://digital.nhs.uk/data-and-information/data-collections-and-data-sets/data-collections/diabetes-transformation-metrics-diab

124 NHS England, Department of Health, Department for Communities and Local Government, Local Government Association. The Better Care Fund. How to... understand and

measure impact. [Internet]. 2015 May. Report No.: Issue 04. Available from: https://www.england.nhs.uk/wp-content/uploads/2015/06/bcf-user-guide-04.pdf.pdf

125 NHS Digital. Quality Outcomes Framework (QOF): Disease prevalence and care quality achievement rates [Internet]. 2022 Jan 4 [cited 2022 Jun 10]. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/general-practice-data-hub/quality-outcomes-framework-qof

126 NHS Digital. Clinical audits and registries [Internet]. [cited 2021 May 20]. Available from: https://digital.nhs.uk/data-and-information/clinical-audits-and-registries

127 Foster J. Integrated Performance Report. Oxfordshire Clinical Commissioning Group. Board Meeting. Oxford; 2020 Jun 9.

128 Leeder S, Russell L, Beaton A. Towards More Meaningful Measures in Healthcare. Asia Pacific Journal of Health Management. 2016;11(3):2-25.

129 National Audit Office. A review of the role and costs of clinical commissioning groups. Report by the Comptroller and Auditor General. London: NHS England; 2018 December 14, 2018.

130 NHS England and NHS Improvement. Fair Shares - a guide to NHS allocations. 2020 [Internet]. [cited 2023 Sep 20]. Available from: https://www.england.nhs.uk/wp-content/uploads/2020/02/nhs-allocations-infographics-v3-23-24.pdf

131 NHS England. CCG allocations 2019/20 to 2023/24 (all funding streams, spreadsheet) [Internet]. London: 2019 [updated 2019 Jul 17; cited 2021 May 20]. Available from: https://www.england.nhs.uk/publication/ccg-allocations-2019-20-to-2023-24-all-funding-streams-spreadsheet/

132 Goddard M. Access to health care services – an English policy perspective. Health Economics, Policy and Law. 2009;4(2):195-208.

133 Murage P, Crawford SM, Bachmann M, Jones A. Geographical disparities in access to cancer management and treatment services in England. Health & Place. 2016;42:11-8.

134 Gatrell AC, Wood DJ. Variation in geographic access to specialist inpatient hospices in England and Wales. Health & Place. 2012;18(4):832-40.

135 Bray BD, Paley L, Hoffman A, James M, Gompertz P, Wolfe CDA, et al. Socioeconomic disparities in first stroke incidence, quality of care, and survival: a nationwide registry-based cohort study of 44 million adults in England. The Lancet Public Health. 2018;3(4):e185-e93.

136 Saunders CA-O, Flynn S, Massou E, Lyratzopoulos G, Abel G, Burt J. Sociodemographic inequalities in patients' experiences of primary care: an analysis of the General Practice Patient Survey in England between 2011 and 2017. (1758-1060 (Electronic)).

137 Elwell-Sutton T, Tinson A, Greszczuk C, Finch D, Holt-White E, Everest G, et al. Creating healthy lives. A whole-government approach to long-term investment in the nation’s health. London: The Health Foundation; 2019 Sep [Internet]. Available from: https://www.health.org.uk/publications/reports/creating-healthy-lives

138 QualityWatch. How have inequalities in the quality of care changed over the last 10 years? London: The Health Foundation & Nuffield Trust; 2020 Jan [Internet]. Available from: https://www.nuffieldtrust.org.uk/public/files/2020-01/quality_inequality/v2/#exploreMoreSection

139 Department of Health and Social Care. Data saves lives: reshaping health and social care with data (draft). 2021 Sep 8 [Internet]. Available from: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data-draft/data-saves-lives-reshaping-health-and-social-care-with-data-draft

140 Propper C. Competition in health care: lessons from the English experience. Health Economics, Policy and Law. 2018;13(3-4):492-508.

141 Crocker H, Kelly L, Harlock J, Fitzpatrick R, Peters M. Measuring the benefits of the integration of health and social care: qualitative interviews with professional stakeholders and patient representatives. BMC Health Services Research. 2020;20(1):515.

142 Billings J, Mikelyte R, Coleman A, MacInnes J, Allen P, Croke S, et al. Supporting integrated care in practice: Perceptions of a national support programme for the development of new care models in England. Journal of Integrated Care. 2020;28(1):27-40.

143 Checkland K, Coleman A, Billings JR, MacInnes J, Mikelyte R, Laverty L, et al. National evaluation of the Vanguard new care models programme interim report: understanding the national support programme. 2019. [Internet]. Available from: https://pure.manchester.ac.uk/ws/portalfiles/portal/103375904/Interim_report_of_the_NCM_external_evaluation_final_v1.pdf

144 Nolte E. Evidence Supporting Integrated Care. In: Amelung V, Stein V, Suter E, Goodwin N, Nolte E, Balicer R, editors. Handbook Integrated Care. Cham: Springer International Publishing; 2021. p. 39-52.

145 Rocks S, Berntson D, Gil-Salmerón A, Kadu M, Ehrenberg N, Stein V, et al. Cost and effects of integrated care: a systematic literature review and meta-analysis. European Journal of Health Economics. 2020;21(8):1211-21.

146 Kadu M, Ehrenberg N, Stein V, Tsiachristas A. Methodological Quality of Economic Evaluations in Integrated Care: Evidence from a Systematic Review. Int J Integr Care. 2019;19(3):17.

147 Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337(7676):979-83.

148 Kumpunen S, Edwards N, Georghiou T, Hughes G. Why do evaluations of integrated care not produce the results we expect? International Journal of Care Coordination. 2020;23(1):9-13.

149 Rutten-van Mölken M, Leijten F, Hoedemakers M, Tsiachristas A, Verbeek N, Karimi M, et al. Strengthening the evidence-base of integrated care for people with multi-morbidity in Europe using Multi-Criteria Decision Analysis (MCDA). BMC Health Serv Res. 2018;18(1):576.

150 Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061.

151 Storey J, Holti R, Hartley J, Marshall M, Matharu T. Health services and delivery research: Clinical leadership in service redesign using Clinical Commissioning Groups: a mixed-methods study. 2018;6(2):2050-4349.

152 Starling A. Implementing new models of care: Lessons from the new care models programme in England. International Journal of Care Coordination. 2018;21(1-2):50-4.

153 Petsoulas C, Allen P, Checkland K, Coleman A, Segar J, Peckham S, et al. Views of NHS commissioners on commissioning support provision. Evidence from a qualitative study examining the early development of clinical commissioning groups in England. BMJ Open. 2014;4(10):e005970.

154 National Institute for Health and Care Excellence. Haematological cancers: improving outcomes (NG47). [Internet]. London: National Institute for Health and Care Excellence; 2016 May 25. [cited 2022 May 19]. Available from: https://www.nice.org.uk/guidance/ng47

155 National Institute for Health and Care Excellence. Rehabilitation for adults with complex psychosis (NG181). [Internet]. London: National Institute for Health and Care Excellence; 2020 August 19. [cited 2022 July 13]. Available from: https://www.nice.org.uk/guidance/ng181

156 National Institute for Health and Care Excellence. Emergency and acute medical care in over 16s: service delivery and organisation (NG94). [Internet]. London: National Institute for Health and Care Excellence; 2018 March 28. [cited 2022 May 19]. Available from: https://www.nice.org.uk/guidance/ng94

157 Wye L, Brangan E, Cameron A, Gabbay J, Klein JH, Pope C. Evidence based policy making and the 'art' of commissioning – how English healthcare commissioners access and use information and academic research in 'real life' decision-making: an empirical qualitative study. BMC Health Services Research. 2015;15(1):430.

158 Addicott R. Commissioning and contracting for integrated care. [Internet]. London: The King's Fund; 2014 November 11. [cited 2021 May 19]. .https://www.kingsfund.org.uk/sites/default/files/kings-fund-commissioning-contracting-integrated-care-nov14.pdf

159 Hewitt P. The Hewitt Review. An independent review of integrated care systems. 2023 4 April 2023.

160 The Integrated Care Systems Network in London. The state of integrated care systems 2022/23. Riding the storm. [cited 2023 Dec 21]. Available from: https://www.nhsconfed.org/system/files/2023-08/the-state-of-integrated-care-systems-2022-23-%285969%29.pdf.

161 Keeney RL, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge: Cambridge University Press; 1993.

162 Baltussen R, Marsh K, Thokala P, Diaby V, Castro H, Cleemput I, et al. Multicriteria Decision Analysis to Support Health Technology Assessment Agencies: Benefits, Limitations, and the Way Forward. Value Health. 2019;22(11):1283-8.

163 European Medicines Agency (EMA). Benefit-risk methodology project. [Internet]. [cited 2021 April 17]. Available from: https://www.ema.europa.eu/en/about-us/what-we-do/regulatory-science-research/benefit-risk-methodology#:~:text=The%20benefit-risk%20methodology%20project%20The%20Agency%20strives%20towards,decisions%20on%20the%20benefits%20and%20risks%20of%20medicines.

164 Chachoua L, Dabbous M, François C, Dussart C, Aballéa S, Toumi M. Use of Patient Preference Information in Benefit-Risk Assessment, Health Technology Assessment, and Pricing and Reimbursement Decisions: A Systematic Literature Review of Attempts and Initiatives. Front Med (Lausanne). 2020 Oct 26;7:543046. doi: 10.3389/fmed.2020.543046. PMID: 33195294; PMCID: PMC7649266.

165 Angelis A, Kanavos P. Value-Based Assessment of New Medical Technologies: Towards a Robust Methodological Framework for the Application of Multiple Criteria Decision Analysis in the Context of Health Technology Assessment. PharmacoEconomics. 2016;34(5):435-46.

166 Angelis A, Linch M, Montibeller G, Molina-Lopez T, Zawada A, Orzel K, et al. Multiple Criteria Decision Analysis for HTA across four EU Member States: Piloting the Advance Value Framework. Social Science & Medicine. 2020;246:112595-.

167 Wagner, M., Samaha, D., Khoury, H., O'Neil, W. M., Lavoie, L., Bennetts, L., ... & Goetghebeur, M. (2018). Development of a framework based on reflective MCDA to support patient–clinician shared decision-making: the case of the management of Gastroenteropancreatic Neuroendocrine Tumors (GEP-NET) in the United States. Advances in therapy, 35, 81-99.

168 Wagner, M., Samaha, D., Cuervo, J., Patel, H., Martinez, M., O'Neil, W. M., & Jimenez-Fonseca, P. (2018). Applying reflective multicriteria decision analysis (MCDA) to patient–clinician shared decision-making on the management of gastroenteropancreatic neuroendocrine tumors (GEP-NET) in the Spanish context. Advances in Therapy, 35, 1215-1231.

169 Cleemput I, Devriese S, Christiaens W, Kohn L. Multi-criteria decision analysis for the appraisal of medical needs: a pilot study. Health Services Research (HSR). Brussels: Belgian Health Care Knowledge Centre (KCE); 2016. Report No.: 272.

170 Cleemput I, Devriese S, Kohn L, Westhovens R. A multi-criteria decision approach for ranking unmet needs in healthcare. Health Policy. 2018;122(8):878-84.

171 Youngkong S, Baltussen R, Tantivess S, Mohara A, Teerawattananon Y. Multicriteria Decision Analysis for Including Health Interventions in the Universal Health Coverage Benefit Package in Thailand. Value in Health. 2012;15(6):961-70.

172 Jehu-Appiah C, Baltussen R, Acquah C, Aikins M, Amah D'Almeida S, Bosu WK, et al. Balancing Equity and Efficiency in Health Priorities in Ghana: The Use of Multicriteria Decision Analysis. Value in Health. 2008;11(7):1081-7.

173 Samanlioglu F, Kaya BE. Evaluation of the COVID-19 Pandemic Intervention Strategies with Hesitant F-AHP. Journal of Healthcare Engineering. 2020;2020:8835258.

174 De Nardo P, Gentilotti E, Mazzaferri F, Cremonini E, Hansen P, Goossens H, et al. Multi-Criteria Decision Analysis to prioritize hospital admission of patients affected by COVID-19 in low-resource settings with hospital-bed shortage. International Journal of Infectious Diseases. 2020;98:494-500.

175 Forestal RL, Pi S-M. A hybrid approach based on ELECTRE III-genetic algorithm and TOPSIS method for selection of optimal COVID-19 vaccines. Journal of Multi-Criteria Decision Analysis. 2021;n/a(n/a).

176 Rahman MR, Islam AHMH, Islam MN. Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. Model Earth Syst Environ. 2020:1-29.

177 Ruggeri M, Cadeddu C, Roazzi P, Mandolini D, Grigioni M, Marchetti M. Multi–Criteria–Decision–Analysis (MCDA) for the Horizon Scanning of Health Innovations an Application to COVID 19 Emergency. International Journal of Environmental Research and Public Health. 2020;17(21):7823.

178 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

179 Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Systematic Reviews. 2021;10(1):39.

180 Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, et al. When and how to update systematic reviews: consensus and checklist. BMJ. 2016;354:i3507.

181 Alberani V, Director BAA, De P, Pietrangeli C, Maria A, Mazza R. The use of grey literature in health sciences: a preliminary survey. Bull Med Libr Assoc. 1990;78(4):358-63.

182 Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: An efficient and complete method to develop literature searches. Journal of the Medical Library Association. 2018;106(4):531-41.

183 Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Research Synthesis Methods. 2020;11(2):181-217.

184 Haddaway NR, Collins AM, Coughlin D, Kirk S. The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. PLOS ONE. 2015;10(9):e0138237.

185 Boeker M, Vach W, Motschall E. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. BMC Medical Research Methodology. 2013;13(1):131.

186 Hupe M. EndNote X9. Journal of Electronic Resources in Medical Libraries. 2019;16(3-4).

187 Tsiachristas A, Cramm JM, Nieboer A, Rutten- van Mölken M. Broader economic evaluation of disease management programs using multi-criteria decision analysis. International Journal of Technology Assessment in Health Care. 2013;29(3):301-8.

188 Broekhuizen H, Groothuis-Oudshoorn CGM, Van Til JA, Hummel JM, Maarten, Ijzerman J. A Review and Classification of Approaches for Dealing with Uncertainty in Multi-Criteria Decision Analysis for Healthcare Decisions. 2015.

189 Microsoft Corporation. Microsoft Excel. Version 2018. [Software]. Redmond, WA: Microsoft Corporation; 2018.

190 Taylor KS, Mahtani KR, Aronson JK. Summarising good practice guidelines for data extraction for systematic reviews and meta-analysis. BMJ Evidence-Based Medicine. 2021;26(3):88.

191 Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Evidence and Value: Impact on DEcisionMaking--the EVIDEM framework and potential applications. BMC health services research. 2008;8:270-.

192 Belton V, Theodor J. S. Multiple Criteria Decision Analysis : An Integrated Approach. Boston [Mass.]; London: Kluwer Academic2002.

193 Saaty TL. Axiomatic Foundation of the Analytic Hierarchy Process. Management Science. 1986;32(7):841-55.

194 Brans J-P, Mareschal B. Promethee Methods. In: Figueira J, Greco S, Ehrogott M, editors. Multiple Criteria Decision Analysis: State of the Art Surveys. New York, NY: Springer New York; 2005. p. 163-86.

195 Hwang C-LY, Kwangsun. Multiple attribute decision making: methods and applications. New York: Springer-Verlag; 1981.

196 Bana e Costa CA, Vansnick J-C. A critical analysis of the eigenvalue method used to derive priorities in AHP. European Journal of Operational Research. 2008;187(3):1422-8.

197 Tervonen T, Gelhorn H, Sri Bhashyam S, Poon JL, Gries KS, Rentz A, et al. MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. Pharmacoepidemiol Drug Saf. 2017;26(12):1483-91.

198 Hansen P, Ombler F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. Journal of Multi-Criteria Decision Analysis. 2008;15(3-4):87-107.

199 Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. Quality in Health Care. 2001;10(suppl 1):i55.

200 Babigumira JB, Jenny AM, Bartlein R, Stergachis A, Garrison LP. Health technology assessment in low- and middle-income countries: a landscape assessment. Journal of Pharmaceutical Health Services Research. 2016;7(1):37-42.

201 Löblová O. Three worlds of health technology assessment: explaining patterns of diffusion of HTA agencies in Europe. Health Economics, Policy and Law. 2016;11(3):253-73.

202 Benoit C, Gorry P. Health technology assessment: the scientific career of a policy concept. International Journal of Technology Assessment in Health Care. 2017;33(1):128-34.

203 Devlin N, Sussex J. Incorporating Multiple Criteria in HTA: Methods and Processes. OHE Monograph [Internet]. 2011. Available from: https://www.ohe.org/publications/incorporating-multiple-criteria-hta-methods-and-processes.

204 European Medicines Agency (EMA). Benefit-risk methodology project. Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. London, United Kingdom: European Medicines Agency; 2010. Report No.: EMA/549682/2010.

205 Liberti L, McAuslane JN, Walker S. Standardizing the Benefit-Risk Assessment of New Medicines. Pharmaceutical Medicine. 2011;25(3):139-46.

206 Drake JI, de Hart JCT, Monleón C, Toro W, Valentim J. Utilization of multiple-criteria decision analysis (MCDA) to support healthcare decision-making FIFARMA, 2016. Journal of Market Access & Health Policy. 2017;5(1):1360545.

207 Angelis A, Phillips LD. Advancing structured decision-making in drug regulation at the FDA and EMA. British Journal of Clinical Pharmacology. 2021;87(2):395-405.

208 Angelis A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. Social Science & Medicine. 2017;188:137-56.

209 Marttunen M, Lienert J, Belton V. Structuring problems for Multi-Criteria Decision Analysis in practice: A literature review of method combinations. European Journal of Operational Research. 2017;263(1):1-17.

210 Marsh KD, Sculpher M, Caro JJ, Tervonen T. The Use of MCDA in HTA: Great Potential, but More Effort Needed. Value in Health. 2018;21(4):394-7.

211 Dodgson J, Spackman M, Pearman A, Phillips. Multi-criteria analysis: a manual. London; 2009.

212 Oliveira MD, Mataloto I, Kanavos P. Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art. The European Journal of Health Economics. 2019;20(6):891-918.

213 Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force working group-6. Medical Decision Making. 2012;32(5):722-32.

214 Durbach IN, Stewart TJ. Modeling uncertainty in multi-criteria decision analysis. Eur J Oper Res. 2012;223(1):1-14.

215 Groothuis-Oudshoorn CGM, Broekhuizen H, van Til J. Dealing with Uncertainty in the Analysis and Reporting of MCDA. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, editors. Cham: Springer International Publishing; 2017. p. 67-85.

216 Gray AM, Clarke P, Wolstenholme J, Wordsworth S, Clarke P, Frew EJ, Louviere JJ. Measuring, valuing, and analysing health outcomes. In: Applied Methods of Cost-Effectiveness Analysis in Health Care. 2011:86-116.

217 Marsh K, Goetghebeur M, Thokala P, Baltussen R. Multi-Criteria Decision Analysis to Support Healthcare Decisions. 1 ed: Springer, Cham; 2017. VI, 329 p.

218 Dolan JG. Patient priorities in colorectal cancer screening decisions. Health expectations : an international journal of public participation in health care and health policy. 2005;8(4):334-44.

219 Wagner M, Khoury H, Bennetts L, Berto P, Ehreth J, Badia X, et al. Appraising the holistic value of Lenvatinib for radio-iodine refractory differentiated thyroid cancer: A multi-country study applying pragmatic MCDA. BMC Cancer. 2017;17(1):272.

220 Hongoh V, Gosselin P, Michel P, Ravel A, Waaub J-P, Campagna C, et al. Criteria for the prioritization of public health interventions for climate-sensitive vector-borne diseases in Quebec. PLOS ONE. 2017;12(12):e0190049.

221 UNESCO. International Standard Classification of Education (ISCED). [Internet]. [cited 2022 June 6]. Available from: https://uis.unesco.org/en/topic/international-standard-classification-education-isced

222 Agapova M, Bresnahan BW, Linnau KF, Garrison LP, Higashi M, Kessler L, et al. Using the Analytic Hierarchy Process for Prioritizing Imaging Tests in Diagnosis of Suspected Appendicitis. Academic Radiology. 2017;24(5):530-7.

223 Guarga L, Badia X, Obach M, Fontanet M, Prat A, Vallano A, et al. Implementing reflective multicriteria decision analysis (MCDA) to assess orphan drugs value in the Catalan Health Service (CatSalut). Orphanet Journal of Rare Diseases. 2019;14(1):157.

224 Howard S, Scott IA, Ju H, McQueen L, Scuffham PA. Multicriteria decision analysis (MCDA) for health technology assessment: the Queensland Health experience. Aust Health Rev. 2019;43(5):591-9.

225 Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. A Pilot Study of Multicriteria Decision Analysis for Valuing Orphan Medicines. Value in Health. 2013;16(8):1163-9.

226 Pieter WGB, Josee AMH. Designing multi-criteria decision analysis processes for priority setting in health policy. Journal of Multicriteria Decision Analysis. 2000;9(1-3):56.

227 Youngkong S, Teerawattananon Y, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. Health Research Policy and Systems. 2012;10(1):6.

228 Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. BMJ : British Medical Journal. 2013;346:f1049.

229 Gongora-Salazar P, Glogowska M, Fitzpatrick R, Perera R, Tsiachristas A. Commissioning [Integrated] Care in England: An Analysis of the Current Decision Context. International Journal of Integrated Care. 2022;22(4).

230 World Health Organization (WHO). Monitoring, evaluation and review of national health strategies: a country-led platform for information and accountability. 2011. [Internet]. [cited 2022 June 11]. Available from: https://apps.who.int/iris/handle/10665/85877.

231 Broqvist M, Garpenby P. To accept, or not to accept, that is the question: citizen reactions to rationing. Health expectations : an international journal of public participation in health care and health policy. 2014;17(1):82-92.

232 Bell DE, Keeney RL, Raiffa H. Conflicting objectives in decisions: John Wiley & Sons; 1977.

233 Belton V. Multi-Criteria Problem Structuring and Analysis in a Value Theory Framework. In: Gal T, Stewart TJ, Hanne T, editors. Multicriteria Decision Making: Advances in MCDM Models, Algorithms, Theory, and Applications. Boston, MA: Springer US; 1999. p. 335-66.

234 Regier DA, Peacock S. Theoretical Foundations of MCDA. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, editors. Multi-Criteria Decision Analysis to Support Healthcare Decisions. Cham: Springer International Publishing; 2017. p. 9-28.

235 Von Neumann J, Morgenstern O. Theory of Games and Economic Behavior (3rd editon). New York: John Wiley; 1953.

236 Dyer JS. Maut — Multiattribute Utility Theory. In: Figueira J, Greco S, Ehrogott M, editors. Multiple Criteria Decision Analysis: State of the Art Surveys. New York, NY: Springer New York; 2005. p. 265-92.

237 Willis T, Chandok R, Delgado R, Razak Y, Robertson L, Dornhorst A. The North West London Diabetes Transformation Programme (NWL-DTP): the journey so far in optimising diabetes care within North West London. Practical Diabetes. 2019;36(2):51-6.

238 NHS England & NICE. Implementing the Early Intervention in Psychosis Access and Waiting Time Standard: Guidance. NHS England, the National Collaborating Centre for Mental Health and the National Institute for Health and Care Excellence (NICE); 2016 April. [cited 2023 Feb 21]. Available from: https://www.rcpsych.ac.uk/docs/default-source/improving-care/nccmh/early-intervention-in-psychosis/nccmh-eip-guidance.pdf?sfvrsn=ae517db4_4.

239 Goeschel CA, Weiss WM, Pronovost PJ. Using a logic model to design and evaluate quality and patient safety improvement programs. International Journal for Quality in Health Care. 2012;24(4):330-7.

240 Gongora-Salazar P, Rocks S, Fahr P, Rivero-Arias O, Tsiachristas A. The Use of Multicriteria Decision Analysis to Support Decision Making in Healthcare: An Updated Systematic Literature Review. Value in Health. 2022.

241 Laugharne R, Eaves S, Mascas A, Psatha K, Dinnis G, Trower J, et al. Desperately seeking outcomes: quantifying the effectiveness of community mental healthcare using Health of the Nation Outcome Scales. BJPsych Open. 2018;4(3):91-4.

242 Tsiachristas A, Thomas T, Leal J, Lennox BR. Economic impact of early intervention in psychosis services: results from a longitudinal retrospective controlled study in England. BMJ Open. 2016;6(10):e012611.

243 Angrist J, Pischke J-S. Mostly Harmless Econometrics. New Jersey: Princeton University Press; 2009.

244 Cunningham, S. Causal inference: The mixtape. Yale University Press; 2021.

245 Leijten FRM, Struckmann V, van Ginneken E, Czypionka T, Kraus M, Reiss M, et al. The SELFIE framework for integrated care for multi-morbidity: Development and description. Health Policy. 2018;122(1):12-22.

246 Marsh K, Thokala P, Mühlbacher A, Lanitis T. Incorporating preferences and priorities into MCDA: selecting an appropriate scoring and weighting technique. Multi-criteria decision analysis to support healthcare decisions. 2017:47-66.

247 Hadorn DC. The role of public values in setting health care priorities. Soc Sci Med. 1991;32(7):773-81.

248 Mooney G. “Communitarian claims” as an ethical basis for allocating health care resources. Social Science & Medicine. 1998;47(9):1171-80.

249 Dworkin R. Liberal Community. California Law Review. 1989;77(3):479-504.

250 Skedgel C, Wailoo A, Akehurst R. Societal Preferences for Distributive Justice in the Allocation of Health Care Resources:A Latent Class Discrete Choice Experiment. Medical Decision Making. 2015;35(1):94-105.

251 Fleck LM. Just Caring: Health Reform and Health Care Rationing. The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine. 1994;19(5):435-43.

252 NHS England. Patient and public participation in commissioning health and care: Statutory guidance for clinical commissioning groups and NHS England. 2017. [cited 2022 Jan 23]. Available from: https://www.england.nhs.uk/publication/patient-and-public-participation-in-commissioning-health-and-care-statutory-guidance-for-ccgs-and-nhs-england/.

253 Whitty JA, Lancsar E, Rixon K, Golenko X, Ratcliffe J. A Systematic Review of Stated Preference Studies Reporting Public Preferences for Healthcare Priority Setting. The Patient - Patient-Centered Outcomes Research. 2014;7(4):365-86.

254 Vass CM, Payne K. Using Discrete Choice Experiments to Inform the Benefit-Risk Assessment of Medicines: Are We Ready Yet? PharmacoEconomics. 2017;35(9):859-66.

255 Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. PharmacoEconomics. 2019;37(2):201-26.

256 Lancaster KJ. A new approach to consumer theory. The Journal of Political Economy. 1966;74(2):132–57.

257 McFadden D. Conditional logit analysis of qualitative choice behavior. In: P Z, editor. Frontiers in econometrics. New York: Academic Press; 1974. p. 105–42.

258 Johnson R, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value in Health. 2013;16(1):3-13.

259 Lancsar E, Fiebig DG, Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. PharmacoEconomics 2017 35:7. 2017;35(7):697-716.

260 Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. The European Journal of Health Economics. 2018;19(8):1053-66.

261 EuroQoL Group. [Internet]. [cited 2022 May 16]. Available from: http://www.euroqol.org/

262 Saint-Hilary G, Robert V, Gasparini M, Jaki T, Mozgunov P. A novel measure of drug benefit-risk assessment based on Scale Loss Score. Stat Methods Med Res. 2019;28(9):2738-53.

263 Menzies, T., Saint-Hilary, G., & Mozgunov, P. (2022). A comparison of various aggregation functions in multi-criteria decision analysis for drug benefit–risk assessment. Statistical methods in medical research, 31(5), 899-916.

264 Stokes J, Checkland K, Kristensen SR. Integrated care: theory to practice. Journal of Health Services Research & Policy. 2016;21(4):282-5.

265 Claxton K. Three questions to ask when examining MCDA. Value & Outcomes Spotlight. 2015;1(1):18-20.

266 Sculpher M, Claxton K, Pearson SD. Developing a Value Framework: The Need to Reflect the Opportunity Costs of Funding Decisions. Value in Health. 2017;20(2):234-9.

267 Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation: Oxford University Press; 2006.

268 Varkevisser CM, Pathmanathan I, Brownlee A. Designing and conducting health system research projects: volume 2, data analyses and report writing: IDRC, Ottawa, ON, CA; 2003.

269 Görgens-Albino M, Marelize GK, J.Zall; . Making Monitoring and Evaluation Systems Work: A Capacity Development Toolkit: World Bank Publications; 2009.

270 NHS England & NICE. Implementing the early intervention in psychosis access and waiting time standard. NHS England & The National Institute for Health and Care Excellence (NICE); 2023 16 February. [cited 2023 Feb 13]. Available from: https://www.england.nhs.uk/wp-content/uploads/2023/03/B1954-implementing-the-early-intervention-in-psychosis-access-and-waiting-time-standard.pdf.

271 NHS England and NHS Improvement. Transformation of urgent and emergency care: models of care and measurement. London: NHS England and NHS Improvement; 2020 December.

Contract No.: PAR122. [cited 2022 Feb 13]. Available from: https://www.england.nhs.uk/wp-content/uploads/2020/12/transformation-of-urgent-and-emergency-care-models-of-care-and-measurement.pdf.

272 Rea RD, Lumb A, Tan GD, Owen K, Thanabalasingham G, Latif A, et al. Using data to improve the care of people with diabetes across Oxfordshire. Practical Diabetes. 2020;37(1):27-31.

273 Office for Health Improvement & Disparities (OHID). Atlas of Variation. 2023. [Internet]. [cited 2023 June 13]. Available from: https://fingertips.phe.org.uk/profile/atlas-of-variation.

274 International Consortium for Health Outcomes Measurement (ICHOM). Patient-Centered Outcome Measures. 2023. [Internet]. [cited 2023 June 13]. Available from: https://www.ichom.org/patient-centered-outcome-measures/.

275 Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ. 2000;321(7262):694-6.

276 Airoldi M, Morton A, Smith J, Bevan G, Airoldi M, Bevan G, et al. Priority Setting for Population Health Healthcare prioritisation at the local level: a socio-technical approach Healthcare prioritisation at the local level: a socio-technical approach. Priority Setting for Population Health Working paper no 7 Department of Management, London School of Economics & Political Science. 2011.

277 Collins M, Mazzei M, Baker R, Morton A, Frith L, Syrett K, et al. Developing a combined framework for priority setting in integrated health and social care systems. BMC Health Services Research. 2023;23(1):879.

278 van de Wetering EJ, van Exel NJA, Rose JM, Hoefman RJ, Brouwer WBF. Are some QALYs more equal than others? The European Journal of Health Economics. 2016;17(2):117-27.

279 Spencer A, Rivero-Arias O, Wong R, Tsuchiya A, Bleichrodt H, Edwards RT, et al. The QALY at 50: One story many voices. Social Science & Medicine. 2022;296:114653.

280 NHS England and Improvement. 2021/22 priorities and operational planning guidance: Implementation guidance. 2021 March 25. [cited 2022 May 19]. Available from: https://www.england.nhs.uk/publication/implementation-guidance/.

281 Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. PharmacoEconomics. 2014;32(9):883-902.

282 Mott DJ. Incorporating Quantitative Patient Preference Data into Healthcare Decision Making Processes: Is HTA Falling Behind? The Patient - Patient-Centered Outcomes Research. 2018;11(3):249-52.

283 Louviere JJ, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications. Cambridge: Cambridge University Press; 2000.

284 Bliemer M, Rose J. Designing and Conducting Stated Choice Experiments. In: Hess S, Daly A, editors. Handbook of Choice Modelling, Second edition, forthcoming2022.

285 Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. PharmacoEconomics. 2008;26(8):661-77.

286 De Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. 2015.

287 Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. BMC health services research. 2014;14:360-.

288 Shah KK, Tsuchiya A, Wailoo AJ. Valuing health at the end of life: A stated preference discrete choice experiment. Social Science & Medicine. 2015;124:48-56.

289 Rowen D, Brazier J, Mukuria C, Keetharuth A, Risa Hole A, Tsuchiya A, et al. Eliciting Societal Preferences for Weighting QALYs for Burden of Illness and End of Life. Medical Decision Making. 2016;36(2):210-22.

290 Hampson G, Mott D, Devlin N, Shah K. Public Preferences for Health Gains and Cures: A Discrete Choice Experiment. OHE Consulting Report. . London: Office of Health Economics; 2019 January 2019.

291 Morrell L, Buchanan J, Rees S, Barker RW, Wordsworth S. What Aspects of Illness Influence Public Preferences for Healthcare Priority Setting? A Discrete Choice Experiment in the UK. PharmacoEconomics. 2021;39(12):1443-54.

292 Rowen D, Powell PA, Hole AR, Aragon M-J, Castelli A, Jacobs R. Valuing quality in mental healthcare: A discrete choice experiment eliciting preferences from mental healthcare service users, mental healthcare professionals and the general population. Social Science & Medicine. 2022;301:114885.

293 Rutten-van Mölken M, Karimi M, Leijten F, Hoedemakers M, Looman W, Islam K, et al. Comparing patients' and other stakeholders' preferences for outcomes of integrated care for multimorbidity: a discrete choice experiment in eight European countries. BMJ Open. 2020;10(10):e037547.

294 NHS England and NHS Improvement. Core20PLUS5: An Approach to Reducing Health Inequalities. Supporting document to enable the completion of the Engage online survey on Core20PLUS5. 2021 December 6. [cited 2022 February 10]. Available from: https://www.england.nhs.uk/about/equality/equality-hub/national-healthcare-inequalities-improvement-programme/core20plus5/.

295 Train K. Discrete Choice Methods with Simulation. Discrete Choice Methods with Simulation. 2009;2nd Edition:1–388.

296 Vass C, Boeri M, Karim S, Marshall D, Craig B, Ho K-A, et al. Accounting for Preference Heterogeneity in Discrete-Choice Experiments: An ISPOR Special Interest Group Report. Value in Health. 2022;25(5):685-94.

297 Rivero-Arias O, Buckell J, Allin B, Craig BM, Ayman G, Knight M. Using stated-preferences methods to develop a summary metric to determine successful treatment of children with a surgical condition: a study protocol. BMJ Open. 2022;12(6):e062833.

298 Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in Health. 2016;19(4):300-15.

299 Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in Health. 2011;14(4):403-13.

300 Veldwijk J, Lambooij MS, van Til JA, Groothuis-Oudshoorn CGM, Smit HA, de Wit GA. Words or graphics to present a Discrete Choice Experiment: Does it matter? Patient Education and Counseling. 2015;98(11):1376-84.

301 Eppink FV, Hanley N, Tucker S. How Best to Present Complex Ecosystem Information in Stated Preference Studies? Ecological Economics. 2019;158:20-5.

302 Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Economics. 2019;28(3):350-63.

303 ChoiceMetrics. Ngene 1.3 User manual and reference guide. The Cutting Edge in Experimental Design. 2021.

304 Bliemer MCJ, Rose JM. Experimental design influences on stated choice outputs: An empirical study in air travel choice. Transportation Research Part A: Policy and Practice. 2011;45(1):63-79.

305 Rose J, Bliemer M. Stated choice experimental design theory: the who, the what and the why. Handbook of Choice Modelling: Edward Elgar Publishing; 2014. p. 152-77.

306 Kessels R, Jones B, Goos P, Vandebroek M. Rejoinder: The usefulness of Bayesian optimal designs for discrete choice experiments. Applied Stochastic Models in Business and Industry. 2011;27:173-88.

307 Crabbe M, Vandebroek M. Using appropriate prior information to eliminate choice sets with a dominant alternative from D-efficient designs. Journal of Choice Modelling. 2012;5(1):22-45.

308 Luyten J, Kessels R, Goos P, Beutels P. Public Preferences for Prioritizing Preventive and Curative Health Care Interventions: A Discrete Choice Experiment. Value in Health. 2015;18(2):224-33.

309 Gu Y, Lancsar E, Ghijben P, Butler JRG, Donaldson C. Attributes and weights in health care priority setting: A systematic review of what counts and to what extent. Social Science & Medicine. 2015;146:41-52.

310 Genie MG, Ryan M, Krucien N. To pay or not to pay? Cost information processing in the valuation of publicly funded healthcare. Social Science & Medicine. 2021;276:113822.

311 StataCorp. Stata Statistical Software: Release 17 [Software]. College Station, TX: StataCorp LLC; 2021.

312 Rose JM, Scarpa R, Bliemer MCJ. Incorporating model uncertainty into the generation of efficient stated choice experiments: A model averaging approach Institute of transport and logistics studies. 2009;Working Paper ITLS-WP-09-08

313 Hensher DA, Rose JM, Greene WH. Applied Choice Analysis. 2 ed. Cambridge: Cambridge University Press; 2015.

314 Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y-3L Health States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ? Medical Decision Making. 2021;41(5):584-96.

315 Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert rev. 2017;17(6):531-42.

316 Bateman I, Carson R, Day B, Hanemann M, Hanley N, Hett T, et al. Economic Valuation with Stated Preference Techniques: Edward Elgar Publishing; 2002.

317 Rakotonarivo S, Schaafsma M, Hockley N. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. Journal of Environmental Management. 2016;183.

318 Robinson J, Clouston B, Suh J, Chaloupka M. Are citizens' juries a useful tool for assessing environmental value? Environmental Conservation. 2008;35(4):351-60.

319 Bateman IJ, Day BH, Jones AP, Jude S. Reducing gain–loss asymmetry: A virtual reality choice experiment valuing land use change. Journal of Environmental Economics and Management. 2009;58(1):106-18.

320 Carlsson F, García JH, Löfgren Å. Conformity and the Demand for Environmental Goods. Environmental and Resource Economics. 2010;47(3):407-21.

321 Soliño M, Farizo BA, Vázquez MX, Prada A. Generating electricity with forest biomass: Consistency and payment timeframe effects in choice experiments. Energy Policy. 2012;41:798-806.

322 Orme B, Johnson R. Including holdout choice tasks in conjoint studies. Sawtooth Software: Research Paper Series. 2015.

323 Inza FSM, Ryan M, Amaya-Amaya M. “Irrational” Stated Preferences: A quantitative and qualitative investigation. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht: Springer Netherlands; 2008. p. 195-215.

324 Copsey B, Buchanan J, Fitzpatrick R, Lamb SE, Dutton SJ, Cook JA. Duration of Treatment Effect Should Be Considered in the Design and Interpretation of Clinical Trials: Results of a Discrete Choice Experiment. Medical Decision Making. 2019;39(4):461-73.

325 Jonker MF, Roudijk B, Maas M. The Sensitivity and Specificity of Repeated and Dominant Choice Tasks in Discrete Choice Experiments. Value in Health. 2022;25(8):1381-9.

326 Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2003;2(1):55-64.

327 Gonzalez JM, Johnson FR, Levitan B, Noel R, Peay H. Symposium Title: Preference Evidence for Regulatory Decisions. The Patient - Patient-Centered Outcomes Research. 2018;11(5):467-73.

328 Jacquemet N, Joule R-V, Luchini S, Shogren J. Preference elicitation under oath. Journal of Environmental Economics and Management. 2013;65(1):110-32.

329 Buckell J, Sindelar JL. The impact of flavors, health risks, secondhand smoke and prices on young adults' cigarette and e-cigarette choices: a discrete choice experiment. Addiction (Abingdon, England). 2019;114(8):1427-35.

330 Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Education and Counseling. 2008;73(3):448-55.

331 Özdemir S, Mohamed AF, Johnson FR, Hauber AB. Who pays attention in stated-choice surveys? Health Economics. 2010;19(1):111-8.

332 Johnson FR, Mohamed AF, Özdemir S, Marshall DA, Phillips KA. How does cost matter in health-care discrete-choice experiments? Health Economics. 2011;20(3):323-30.

333 Vass CM, Davison NJ, Vander Stichele G, Payne K. A Picture is Worth a Thousand Words: The Role of Survey Training Materials in Stated-Preference Studies. The Patient - Patient-Centered Outcomes Research. 2020;13(2):163-73.

334 Sawtooth S. The CBC System for Choice-Based Conjoint Analysis (Version 8). Sawtooth Software Technical Paper Series. 2013.

335 Chrzan K, Halversen C. Diagnostics for random respondents in choice experiments. Sawtooth Software; 2021.

336 Hess S. Conditional parameter estimates from Mixed Logit models: distributional assumptions and a free software tool. Journal of Choice Modelling. 2010;3(2):134-52.

337 Hess S, Train KE, Polak JW. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. Transportation Research Part B: Methodological. 2006;40(2):147-63.

338 Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B: Methodological. 2003;37(8):681-98.

339 Gonzalez J. A Guide to Measuring and Interpreting Attribute Importance. The Patient - Patient-Centered Outcomes Research. 2019;12:1-9.

340 Bryce, C., Loewenstein, G., Arnold, R., Schooler, J., Wax, R., & Angus, D. Quality of Death: Assessing the Importance Placed on End-of-Life Treatment in the Intensive-Care Unit. Medical Care, 2004: 42, 423-431. https://doi.org/10.1097/01.mlr.0000124245.62354.57.

341 Rocca, H., Rickenbacher, P., Muzzarelli, S., Schindler, R., Maeder, M., Jeker, U., Kiowski, W., Leventhal, M., Pfister, O., Osswald, S., Pfisterer, M., & Rickli, H. End-of-life preferences of elderly patients with chronic heart failure.. European heart journal, 2012: 33 6, 752-9. https://doi.org/10.1093/eurheartj/ehr404.

342 Brazier J, McCabe C. 'Is there a case for using visual analogue scale valuations in CUA' by Parkin and Devlin a response: 'yes there is a case, but what does it add to ordinal data?'. Health Economics. 2007;16(6):645-7.

343 Train K, Weeks M. Discrete choice models in preference space and willingness-to-pay space. In: Scarpa R, Alberini A, editors. Applications of Simulation Methods in Environmental and Resource Economic. 6. Netherlands: Springer; 2005.

344 Mott DJ, Chami N, Tervonen T. Reporting Quality of Marginal Rates of Substitution in Discrete Choice Experiments That Elicit Patient Preferences. Value Health. 2020;23(8):979-84.

345 Defechereux T, Paolucci F, Mirelman A, Youngkong S, Botten G, Hagen TP, et al. Health care priority setting in Norway a multicriteria decision analysis. BMC Health Services Research. 2012;12(1):39.

346 Hess S, Palma D. Apollo: A flexible, powerful and customisable freeware package for choice model estimation and application. Journal of Choice Modelling. 2019;32:100170.

347 RStudio Team. RStudio: Integrated Development for R [Software]. Boston, MA: PBC; 2020.

348 Office for National Statistics (ONS). Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland. In: Office for National Statistics, editor. 2021. [cited 2022 February 10]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2021.

349 Office for National Statistics (ONS). Population estimates by ethnic group, England and Wales. In: Office for National Statistics, editor. 2021. [cited 2022 February 10]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021.

350 Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: An EQ-5D-5L value set for England. Health Economics. 2018;27(1):7-22.

351 Office for National Statistics (ONS). Exploring local income deprivation: A detailed picture of disparities within English local authorities to a neighborhood level. 2021. [cited 2022 February 10]. Available from: https://www.ons.gov.uk/visualisations/dvc1371/#/E07000223.

352 Berndt ER, Hall BH, Hall RE, Hausman JA. Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement, Volume 3, number 4: NBER; 1974. p. 653-65.

353 Nieboer AP, Koolman X, Stolk EA. Preferences for long-term care services: willingness to pay estimates derived from a discrete choice experiment. Soc Sci Med. 2010;70(9):1317-25.

354 Nicolet A, Perraudin C, Wagner J, Gilles I, Krucien N, Peytremann-Bridevaux I, et al. Patient and Public Preferences for Coordinated Care in Switzerland: Development of a Discrete Choice Experiment. The Patient - Patient-Centered Outcomes Research. 2022.

355 Hoedemakers M, Karimi M, Jonker M, Tsiachristas A, Rutten-van Mölken M. Heterogeneity in preferences for outcomes of integrated care for persons with multiple chronic diseases: a latent class analysis of a discrete choice experiment. Quality of Life Research. 2022.

356 Department of Health and Social Care. Direct and Indirect Health Impacts of COVID-19 in England [Internet]. 2021 Sep 17 [cited 2022 May 19]. Available from: https://www.gov.uk/government/publications/dhsc-direct-and-indirect-health-impacts-of-covid-19-in-england-long-paper-9-september-2021.

357 Office for National Statistics (ONS). People with long-term health conditions, UK: January to December 2019. Office for National Statistics; 2020 March 26. [cited 2022 May 19]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/adhocs/11478peoplewithlongtermhealthconditionsukjanuarytodecember2019.

358 NIHR Evidence. Multiple long-term conditions (multimorbidity): making sense of the evidence 2021 [Internet]. [updated 2021 Mar 30; cited 2022 May 19]. Available from: https://evidence.nihr.ac.uk/collection/making-sense-of-the-evidence-multiple-long-term-conditions-multimorbidity/.

359 Ryan M, Kinghorn P, Entwistle VA, Francis JJ. Valuing patients' experiences of healthcare processes: Towards broader applications of existing methods. Social Science & Medicine. 2014;106:194-203.

360 Office for Health Improvement and Disparities. Patient-reported outcomes and experiences study [Internet]. London: UK Government; 2020 [cited 2023 Jun 13]. Available from: https://www.gov.uk/guidance/patient-reported-outcomes-and-experiences-study.

361 NHS England. Patient experience 2022 [Internet]. [updated 2022 Jul 22; cited 2023 Jun 13]. Available from: https://www.england.nhs.uk/gp/patient-experience/.

362 Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value in health. 2009;12:S10-S5.

363 Lancsar E, Gu Y, Gyrd-Hansen D, Butler J, Ratcliffe J, Bulfone L, et al. The relative value of different QALY types. Journal of Health Economics. 2020;70:102303.

364 Paolucci F, Mentzakis E, Defechereux T, Niessen LW. Equity and efficiency preferences of health policy makers in China—a stated preference analysis. Health Policy and Planning. 2015;30(8):1059-66.

365 Mentzakis E, García-Goñi M, Sequeira AR, Paolucci F. Equity and efficiency priorities within the Spanish health system: A discrete choice experiment eliciting stakeholders preferences. Health Policy and Technology. 2019;8(1):30-41.

366 Rawls J. A Theory of Justice. Oxford: Oxford University Press; 1999.

367 Marmot M, Allen J, Boyce T, Goldblatt P, Morrison J. Health equity in England: The Marmot Review 10 years on. London: Institute of Health Equity; 2020.

368 Care Quality Commission. Inequalities in care [Internet]. London: Care Quality Commission; 2022 Oct 21 [cited 2023 Jun 13]. Available from: https://www.cqc.org.uk/publication/state-care-202122/inequalities.

369 Nghiem SH, Connelly LB. Convergence and determinants of health expenditures in OECD countries. Health Economics Review. 2017;7(1):29.

370 Kruk ME, Kelley E, Syed SB, Tarp F, Addison T, Akachi Y. Measuring quality of health-care services: what is known and where are the gaps? Bull World Health Organ. 2017;95(6):389-a.

371 Harari D, Francis-Devine B, Bolton P, Keep M. Rising cost of living in the UK. UK Parliament; 2023 24 November 2023.

372 Bryan S, Buxton M, Sheldon R, Grant A. Magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. Health Economics. 1998;7(7):595-603.

373 Essers BAB, van Helvoort-Postulart D, Prins MH, Neumann M, Dirksen CD. Does the Inclusion of a Cost Attribute Result in Different Preferences for the Surgical Treatment of Primary Basal Cell Carcinoma? PharmacoEconomics. 2010;28(6):507-20.

374 Pedersen LB, Kjær T, Kragstrup J, Gyrd-Hansen D. Does the Inclusion of a Cost Attribute in Forced and Unforced Choices Matter?: Results from a Web Survey Applying the Discrete Choice Experiment. Journal of Choice Modelling. 2011;4(3):88-109.

375 Sever I, Verbič M, Klarić Sever E. Cost attribute in health care DCEs: Just adding another attribute or a trigger of change in the stated preferences? Journal of Choice Modelling. 2019;32:100135.

376 Correll CU, Galling B, Pawar A, Krivko A, Bonetto C, Ruggeri M, et al. Comparison of Early Intervention Services vs Treatment as Usual for Early-Phase Psychosis: A Systematic Review, Meta-analysis, and Meta-regression. JAMA Psychiatry. 2018;75(6):555-65.

377 Puntis S, Pappa S, Lennox B. What happens after early intervention? Mapping early intervention in psychosis care pathways in the 12 months after discharge. Early Interv Psychiatry. 2023.

378 NHS England. NHS Mental Health Implementation Plan 2019/20 – 2023/24. London: NHS England; 2019 Jul [cited 2023 Jun 13]. Available from: https://www.longtermplan.nhs.uk/wp-content/uploads/2019/07/nhs-mental-health-implementation-plan-2019-20-2023-24.pdf.

379 National Institute for Health and Care Excellence (NICE). Psychosis and schizophrenia in adults: prevention and management [Internet]. London: NICE; 2014 Mar. (NICE Clinical Guidelines, No. 178) [cited 2023 Jun 13]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK555203/.

380 NHS England. Guidance to support the introduction of access and waiting time standards for mental health services in 2015/16 [Internet]. London: NHS England; 2015 [cited 2023 Jun 10]. Available from: https://www.england.nhs.uk/wp-content/uploads/2015/02/mh-access-wait-time-guid.pdf.

381 Royal College of Psychiatrists. National Clinical Audit of Psychosis – England National Report for the Early Intervention in Psychosis Audit 2021/2022.: London: Healthcare Quality Improvement Partnership; 2022.

382 NIHR Maudsley Biomedical Research Centre. Clinical Record Interactive Search (CRIS) [Internet]. London: NIHR Maudsley BRC; 2023 [cited 2023 Jun 13]. Available from: https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/.

383 Camacho-Gomez M, Castellvi P. Effectiveness of Family Intervention for Preventing Relapse in First-Episode Psychosis Until 24 Months of Follow-up: A Systematic Review With Meta-analysis of Randomized Controlled Trials. Schizophr Bull. 2020;46(1):98-109.

384 Almond S, Knapp M, Francois C, Toumi M, Brugha T. Relapse in schizophrenia: costs, clinical outcomes and quality of life. The British Journal of Psychiatry. 2004;184(4):346-51.

385 Anderson KK, Fuhrer R, Schmitz N, Malla AK. Determinants of negative pathways to care and their impact on service disengagement in first-episode psychosis. Social Psychiatry and Psychiatric Epidemiology. 2013;48(1):125-36.

386 Public Health England. Psychosis Data Report: Describing variation in numbers of people with psychosis and their access to care in England [Internet]. London: Public Health England; 2016 [cited 2023 Jun 13]. Available from: https://assets.publishing.service.gov.uk/media/5c5176ff40f0b6254b1b101e/Psychosis_data_report.pdf.

387 Puntis S, Oke J, Lennox B. Discharge pathways and relapse following treatment from early intervention in psychosis services. BJPsych Open. 2018;4(5):368-74.

388 Penman-Aguilar A, Talih M, Huang D, Moonesinghe R, Bouye K, Beckles G. Measurement of Health Disparities, Health Inequities, and Social Determinants of Health to Support the Advancement of Health Equity. J Public Health Manag Pract. 2016;22 Suppl 1(Suppl 1):S33-42.

389 Cookson R, Propper C, Asaria M, Raine R. Socio-Economic Inequalities in Health Care in England. Fiscal Studies. 2016;37(3-4):371-403.

390 McLennan D, Noble S, Noble M, Plunkett E, Wright G, Gutacker N. The English Indices of Deprivation 2019. London: Ministry of Housing, Communities and Local Government 2019.

391 Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, et al. Unit Costs of Health and Social Care 2022 Manual. Technical report. Personal Social Services Research Unit (University of Kent) & Centre for Health Economics (University of York); 2023. Report No.: KAR id:100519.

392 Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983;70(1):41-55.

393 Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.

394 Heckman JJ, Ichimura H, Todd PE. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. The review of economic studies. 1997;64(4):605-54.

395 Baser O. Too much ado about propensity score models? Comparing methods of propensity score matching. Value Health. 2006;9(6):377-85.

396 Rosenbaum PR, Rubin DB. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician. 1985;39(1):33-8.

397 Dehejia RH, Wahba S. Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics. 2002;84(1):151-61.

398 Heckman J, Navarro-Lozano S. Using matching, instrumental variables, and control functions to estimate economic choice models. Rev Econ Stat. 2004;86(1):30-57.

399 Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. Am J Epidemiol. 2011;173(7):761-7.

400 Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015;34(28):3661-79.

401 Rubin DB. Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation. Health Services and Outcomes Research Methodology. 2001;2(3):169-88.

402 McCullagh P, & Nelder, J. A. Generalized Linear Models (Second edition): Chapman and Hall/CRC Press; 1989.

403 Claeskens G, Hjort NL. Model Selection and Model Averaging: Cambridge University Press; 2008.

404 Schwarz G. Estimating the Dimension of a Model. The Annals of Statistics. 1978;6(2):461-4.

405 Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike. New York, NY: Springer New York; 1998. p. 199-213.

406 Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. Value Health. 2012;15(8):1172-81.

407 Hansen P, Devlin N. Multi-Criteria Decision Analysis (MCDA) in Healthcare Decision-Making. Oxford University Press; 2019.

408 Hoedemakers M, Karimi M, Leijten F, Goossens L, Islam K, Tsiachristas A, et al. Value-based person-centred integrated care for frail elderly living at home: a quasi-experimental evaluation using multicriteria decision analysis. BMJ Open. 2022;12(4):e054672.

409 Sashidharan SP, White R, Mezzina R, Jansen S, Gishoma D. Global mental health in high-income countries. Br J Psychiatry. 2016;209(1):3-5.

410 Dattani S. How do researchers study the prevalence of mental illnesses?2023. Available from: https://ourworldindata.org/how-do-researchers-study-the-prevalence-of-mental-illnesses.

411 Phelps CE, Madhavan G. Using Multicriteria Approaches to Assess the Value of Health Care. Value in Health. 2017;20(2):251-5.

412 NHS England. Building an integrated care system intelligence function [Internet]. London: NHS England; 2023 Feb 27 [cited 2023 Jun 13]. Available from: https://www.england.nhs.uk/long-read/building-an-ics-intelligence-function/.

413 Michaels JA. Value assessment frameworks: who is valuing the care in healthcare? Journal of Medical Ethics. 2022;48(6):419-26.

414 Department of Health & Social Care. Data saves lives: reshaping health and social care with data. 2022 [updated 2022 June 15]. [cited 2022 May 19]. Available from: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data.

415 Office for National Statistics. Labour Force Survey. [Internet]. [cited 2022 February 10]. Available from: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurvey.

416 Calabrò GE, La Torre G, de Waure C, Villari P, Federici A, Ricciardi W, et al. Disinvestment in healthcare: an overview of HTA agencies and organizations activities at European level. BMC Health Services Research. 2018;18(1):148.

417 Harris C, Green S, Ramsey W, Allen K, King R. Sustainability in Health care by Allocating Resources Effectively (SHARE) 9: conceptualising disinvestment in the local healthcare setting. BMC Health Serv Res. 2017;17(1):633.

418 Harris C, Allen K, Brooke V, Dyer T, Waller C, King R, et al. Sustainability in Health care by Allocating Resources Effectively (SHARE) 6: investigating methods to identify, prioritise, implement and evaluate disinvestment projects in a local healthcare setting. BMC Health Services Research. 2017;17(1):370.

419 Garner S, Docherty M, Somner J, Sharma T, Choudhury M, Clarke M, et al. Reducing ineffective practice: challenges in identifying low-value health care using Cochrane systematic reviews. J Health Serv Res Policy. 2013;18(1):6-12.

420 Gustavsson E, Lindblom L. Justification of principles for healthcare priority setting: the relevance and roles of empirical studies exploring public values. Journal of Medical Ethics. 2023:jme-2022-108702.

421 Bake C. Mental health statistics (England) [Internet]. London: The House of Commons Library; 2021 Dec 13. Available from: https://dera.ioe.ac.uk/id/eprint/34934/1/SN06988%20(redacted).pdf.

422 NHS National Quality Board. NHS Patient Experience Framework [Internet]. London: NHS; 2012 Feb 22. Available from: https://assets.publishing.service.gov.uk/media/5a7c62b740f0b62aff6c14e9/dh_132788.pdf.

423 Sabik LM, Lie RK. Priority setting in health care: Lessons from the experiences of eight countries. International Journal for Equity in Health. 2008;7(1):4.

424 Hunter DJ, Marks L, Brown J, Scalabrini S, Salway S, Vale L, et al. The potential value of priority-setting methods in public health investment decisions: qualitative findings from three English local authorities. Critical Public Health. 2016;26(5):578-87.

425 NHS England. 2023/24 priorities and operational planning guidance. 2023 January 27. [cited 2023 June 23]. Available from: https://www.england.nhs.uk/publication/2023-24-priorities-and-operational-planning-guidance/.

426 Charlton V, DiStefano M, Mitchell P, Morrell L, Rand L, Badano G, et al. We need to talk about values: a proposed framework for the articulation of normative reasoning in health technology assessment. Health Economics, Policy and Law. 2023:1-21.

427 Mallender JA, Hatefi A, Allen LN, Amoore Z, Kak M, El-Saharty S, et al. Models of Integrated Care: The Global Experience (English). Washington, D.C.: World Bank Group; 2023.

428 NHS England. Delivering a 'Net Zero' National Health Service [Internet]. London: NHS England; 2022 Jul 4 [cited 2023 Nov 22]. Available from: https://www.england.nhs.uk/greenernhs/a-net-zero-nhs/#:~:text=On%201%20July%202022%2C%20the,now%20issued%20as%20statutory%20guidance.

429 Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. 2013;346:f1049-f.

430 Miot J, Wagner M, Khoury H, Rindress D, Goetghebeur MM. Field testing of a multicriteria decision analysis (MCDA) framework for coverage of a screening test for cervical cancer in South Africa. Cost effectiveness and resource allocation : C/E. 2012;10(1):2-.

431 Marsh K, Dolan P, Kempster J, Lugon M. Prioritizing investments in public health: a multi-criteria decision analysis. J Public Health (Oxf). 2013;35(3):460-6.

432 Nobre FF, Trotta LTF, Gomes LFAM. Multi-criteria decision making – an approach to setting priorities in health care. Statistics in Medicine. 1999;18(23):3345-54.

433 Le Gales C, Moatti JP. Searching for consensus through multi-criteria decision analysis. Assessment of screening strategies for hemoglobinopathies in southeastern France. Int J Technol Assess Health Care. 1990;6(3):430-49.

434 Dolan JG, Bordley DR. Isoniazid Prophylaxis:The Importance of Individual Values. Medical Decision Making. 1994;14(1):1-8.

435 Wenstøp F, Magnus P. Value focused rationality in AIDS policy. Health Policy. 2001;57(1):57-72.

436 Hummel JM, Snoek GJ, van Til JA, van Rossum W, Ijzerman MJ. A multicriteria decision analysis of augmentative treatment of upper limbs in persons with tetraplegia. J Rehabil Res Dev. 2005;42(5):635-44.

437 Rose JM, Bliemer MC. Constructing efficient stated choice experimental designs. Transp Rev. 2009;29(5):587-617.

438 Dolan JG, Frisina S. Randomized Controlled Trial of a Patient Decision Aid for Colorectal Cancer Screening. Medical Decision Making. 2002;22(2):125-39.

439 Van Wijk BLG, Klungel OH, Heerdink ER, de Boer A. A Comparison of Two Multiple-Characteristic Decision-Making Models for the Comparison of Antihypertensive Drug Classes. American Journal of Cardiovascular Drugs. 2006;6(4):251-8.

440 Singh S, Dolan JG, Centor RM. Optimal management of adults with pharyngitis – a multi-criteria decision analysis. BMC Medical Informatics and Decision Making. 2006;6(1):14.

441 Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Effectiveness and Resource Allocation. 2006;4(1):14.

442 Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. Health Econ. 2006;15(7):689-96.

443 Baltussen R, ten Asbroek AHA, Koolman X, Shrestha N, Bhattarai P, Niessen LW. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? Health Policy and Planning. 2007;22(3):178-85.

444 Wilson ES, J.; Macleod, C.; Fordham, R. Prioritizing health technologies in a Primary Care Trust. Journal of Health Services Research & Policy. 2007;12(2):80-5.

445 van Til JA, Renzenbrink GJ, Dolan JG, Ijzerman MJ. The Use of the Analytic Hierarchy Process to Aid Decision Making in Acquired Equinovarus Deformity. Archives of Physical Medicine and Rehabilitation. 2008;89(3):457-62.

446 Dalalah DM, S. A Remote Fuzzy Multicriteria Diagnosis of Sore Throat. Telemedicine and e-Health. 2008;14(7):656-65.

447 González-Zapata LI, Alvarez-Dardet C, Ortiz-Moncada R, Clemente V, Millstone E, Holdsworth M, et al. Policy options for obesity in Europe: a comparison of public health specialists with other stakeholders. Public Health Nutrition. 2009;12(7):896-908.

448 Felli JC, Noel RA, Cavazzoni PA. A Multiattribute Model for Evaluating the Benefit-Risk Profiles of Treatment Alternatives. Medical Decision Making. 2009;29(1):104-15.

449 Kroese M, Burton H, Whittaker J, Lakshman R, Alberg C. A Framework for the Prioritization of Investment in the Provision of Genetic Tests. Public Health Genomics. 2010;13(7-8):538-43.

450 Goetghebeur MM, Wagner M, Khoury H, Rindress D, Grégoire J-P, Deal C. Combining multicriteria decision analysis, ethics and health technology assessment: applying the EVIDEM

decision-making framework to growth hormone for Turner syndrome patients. Cost effectiveness and resource allocation : C/E. 2010;8:4-.

451 Baeten SA, Baltussen RMPM, Uyl-de Groot CA, Bridges J, Niessen LW. Incorporating Equity–Efficiency Interactions in Cost-Effectiveness Analysis—Three Approaches Applied to Breast Cancer Control. Value in Health. 2010;13(5):573-9.

452 Tony M, Wagner M, Khoury H, Rindress D, Papastavros T, Oh P, et al. Bridging health technology assessment (HTA) with multicriteria decision analyses (MCDA): field testing of the EVIDEM framework for coverage decisions by a public payer in Canada. BMC Health Services Research. 2011;11(1):329.

453 Diaby V, Lachaine J. An application of a proposed framework for formulary listing in low-income countries. Applied Health Economics and Health Policy. 2011;9(6):389-402.

454 Cunich M, Salkeld G, Dowie J, Henderson J, Bayram C, Britt H, et al. Integrating Evidence and Individual Preferences Using a Web-Based Multi-Criteria Decision Analytic Tool. The Patient: Patient-Centered Outcomes Research. 2011;4(3):153-62.

455 Airoldi M, Morton AD, Smith J, Bevan G. Healthcare prioritisation at the local level: a socio-technical approach. 2011

456 Airoldi M, Morton A, Smith JA, Bevan G. STAR—people-powered prioritization: a 21st-century solution to allocation headaches. Med Decis Making. 2014;34(8):965-975.

457 Youngkong S, Baltussen R, Tantivess S, Koolman X, Teerawattananon Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. BMC Health Services Research. 2010;10(1):197.

458 van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multicriteria benefit–risk assessment using network meta-analysis. Journal of Clinical Epidemiology. 2012;65(4):394-403.

459 Guest J, Harrop JS, Aarabi B, Grossman RG, Fawcett JW, Fehlings MG, et al. Optimization of the decision-making process for the selection of therapeutics to undergo clinical testing for spinal cord injury in the North American Clinical Trials Network. J Neurosurg Spine. 2012;17(1 Suppl):94-101.

460 Hummel MJM, Volz F, van Manen JG, Danner M, Dintsios C-M, Ijzerman MJ, et al. Using the Analytic Hierarchy Process to Elicit Patient Preferences. The Patient: Patient-Centered Outcomes Research. 2012;5(4):225-37.

461 Hummel JM, Boomkamp ISM, Steuten LMG, Verkerke BGJ, IJzerman MJ. Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis. Journal of Orthopaedic Research. 2012;30(9):1453-8.

462 Holdsworth M, El Ati J, Bour A, Kameli Y, Derouiche A, Millstone E, et al. Developing national obesity policy in middle-income countries: a case study from North Africa. Health policy and planning. 2012;28(8):858-70.

463 Tervonen T, Naci H, van Valkenhoef G, Ades AE, Angelis A, Hillege HL, et al. Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing among Statins in Primary Prevention. Medical Decision Making. 2015;35(7):859-71.

464 Strømme EM, Bærøe K, Norheim OF. Disease Control Priorities for Neglected Tropical Diseases: Lessons from Priority Ranking Based on the Quality of Evidence, Cost Effectiveness,

Severity of Disease, Catastrophic Health Expenditures, and Loss of Productivity. Developing World Bioethics. 2014;14(3):132-41.

465 Garety PA, Fowler DG, Freeman D, Bebbington P, Dunn G, Kuipers E. Cognitive-behavioural therapy and family intervention for relapse prevention and symptom reduction in psychosis: Randomised controlled trial. The British Journal of Psychiatry. 2008;192(6):412-23.

466 Arias de la Torre J, Vilagut G, Ronaldson A, Dregan A, Ricci-Cabello I, Hatch SL, et al. Prevalence and age patterns of depression in the United Kingdom. A population-based study. Journal of Affective Disorders. 2021;279:164-72.

467 British Heart Foundation. UK Factsheet November 2018 [Internet]. London: British Heart Foundation; 2018 [cited 2023 Jun 13]. Available from: https://www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications/cardiovascular-disease-statistics-2018.

468 Cruden G, Frerichs L, Powell BJ, Lanier P, Brown CH, Lich KH. Developing a Multi-Criteria Decision Analysis Tool to Support the Adoption of Evidence-Based Child Maltreatment Prevention Programs. Prevention Science. 2020;21(8):1059-64.

469 World Health Organization. Quality of Care. [Internet]. [Date of access, e.g., 2023 Jan 25]. Available from: https://www.who.int/health-topics/quality-of-care#tab=tab_1

470 Davidson L, Scott J, Forster N. Patient experiences of integrated care within the United Kingdom: A systematic review. International Journal of Care Coordination. 2021;24(2):39-56.

471 Care Quality Commission. Adult Inpatient Survey 2022: NHS; 2022 [Available from: https://nhssurveys.org/surveys/survey/02-adults-inpatients/.

472 Office for Health Improvement & Disparities. Atlas of Variation 2022 [Available from: https://fingertips.phe.org.uk/profile/atlas-of-variation.

473 Pauwels K, Huys I, Casteels M, Denier Y, Vandebroek M, Simoens S. What Does Society Value About Cancer Medicines? A Discrete Choice Experiment in the Belgian Population. Applied Health Economics and Health Policy. 2019;17(6):895-902.

474 National Institute for Health and Care Excellence. Achieving and demonstrating compliance with NICE TA and HST guidance 2022 [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/achieving-and-demonstrating-compliance.

475 Flanagan S, Damery S, Combes G. The effectiveness of integrated care interventions in improving patient quality of life (QoL) for patients with chronic conditions. An overview of the systematic review evidence. Health and Quality of Life Outcomes. 2017;15(1):188.

476 Carreira H, Williams R, Dempsey H, Bhaskaran K. Recording of patients' mental health and quality of life-related outcomes in primary care: a cross-sectional study in the UK. BMJ Open. 2022;12(12):e066949.

477 NHS England and NHS Improvement. NHS commissioning: Specialised services 2022 [Available from: https://www.england.nhs.uk/commissioning/spec-services/.

478 Norheim OF. Ethical Perspective: Five Unacceptable Trade-offs on the Path to Universal Health Coverage. Int J Health Policy Manag. 2015;4(11):711-4.

479 Norheim OF. Ethical priority setting for universal health coverage: challenges in deciding upon fair distribution of health services. BMC Medicine. 2016;14(1):75.

480 Kieslich K, Bump JB, Norheim OF, Tantivess S, Littlejohns P. Accounting for Technical, Ethical, and Political Factors in Priority Setting. Health Systems & Reform. 2016;2(1):51-60.

481 Trapero-Bertran M, Rodríguez-Martín B, López-Bastida J. What attributes should be included in a discrete choice experiment related to health technologies? A systematic literature review. PLOS ONE. 2019;14(7):e0219905.

482 Goto R, Mori T. Comparison of Equity Preferences for Life Expectancy Gains: A Discrete Choice Experiment Among the Japanese and Korean General Public. Value in Health Regional Issues. 2019;18:8-13.

483 Bourke SM, Plumpton CO, Hughes DA. Societal Preferences for Funding Orphan Drugs in the United Kingdom: An Application of Person Trade-Off and Discrete Choice Experiment Methods. Value in Health. 2018;21(5):538-46.

484 López-Bastida J, Ramos-Goñi JM, Aranda-Reneo I, Trapero-Bertran M, Kanavos P, Rodriguez Martin B. Using a stated preference discrete choice experiment to assess societal value from the perspective of decision-makers in Europe. Does it work for rare diseases? Health Policy. 2019;123(2):152-8.

485 Nicolet A, van Asselt ADI, Vermeulen KM, Krabbe PFM. Value judgment of new medical treatments: Societal and patient perspectives to inform priority setting in The Netherlands. PLOS ONE. 2020;15(7):e0235666.

486 Babashahi S, Hansen P, Sullivan T. Creating a priority list of non-communicable diseases to support health research funding decision-making. Health Policy. 2021;125(2):221-8.

487 Huppelschoten AG, Verkerk EW, Appleby J, Groenewoud H, Adang EMM, Nelen WLDM, et al. The monetary value of patient-centred care: results from a discrete choice experiment in Dutch fertility care. Human Reproduction. 2014;29(8):1712-20.

488 Care Quality Commission. Adult Inpatient Survey 2021. NHS; 2021.

489 Ipsos MORI. GP Patient Survey 2021. In: NHS, editor. Technical Annex2022.

490 Green C. Investigating public preferences on ‘severity of health’ as a relevant condition for setting healthcare priorities. Social Science & Medicine. 2009;68(12):2247-55.

491 Lim MK, Bae EY, Choi S-E, Lee EK, Lee T-J. Eliciting Public Preference for Health-Care Resource Allocation in South Korea. Value in Health. 2012;15(1, Supplement):S91-S4.

492 Lal A, Siahpush M, Moodie M, Peeters A, Carter R. Weighting Health Outcomes by Socioeconomic Position Using Stated Preferences. PharmacoEconomics - Open. 2018;2(1):43-51.

493 Ezeife DA-OX, Dionne F, Fares AF, Cusano ELR, Fazelzad R, Ng W, et al. Value assessment of oncology drugs using a weighted criterion-based approach. 2020(1097-0142 (Electronic)).

494 McNamara S, Tsuchiya A, Holmes J. Does the UK-public's aversion to inequalities in health differ by group-labelling and health-gain type? A choice-experiment. Social Science & Medicine. 2021;269:113573.

495 Linley WG, Hughes DA. SOCIETAL VIEWS ON NICE, CANCER DRUGS FUND AND VALUE-BASED PRICING CRITERIA FOR PRIORITISING MEDICINES: A CROSS-SECTIONAL SURVEY OF 4118 ADULTS IN GREAT BRITAIN. Health Economics. 2013;22(8):948-64.

496 Campbell HE, Gray AM, Watson J, Jackson C, Moseley C, Cruickshank ME, et al. Preferences for interventions designed to increase cervical screening uptake in non-attending young women: How findings from a discrete choice experiment compare with observed behaviours in a trial. Health Expectations. 2020;23(1):202-11.

497 Oremus M, Tarride J-E, Raina P, Thabane L, Foster G, Goldsmith CH, et al. The General Public's Willingness to Pay for Tax Increases to Support Unrestricted Access to an Alzheimer's Disease Medication. PharmacoEconomics. 2012;30(11):1085-95.

498 Mentzakis E, Stefanowska P, Hurley J. A discrete choice experiment investigating preferences for funding drugs used to treat orphan diseases: an exploratory study. Health Economics, Policy and Law. 2011;6(3):405-33.

499 Jouyani Y, Hadiyan M, Salehi M, Souri A. Using discrete choice model to elicit preference for health-care priority setting. J Educ Health Promot. 2019;8:117-.

500 Busse R, Stahl J. Integrated Care Experiences And Outcomes In Germany, The Netherlands, And England. Health Affairs. 2014;33(9):1549-58.

501 Keeble E, Bardsley M, Durand MA, Hoomans T, Mays N. Area level impacts on emergency hospital admissions of the integrated care and support pioneer programme in England: difference-in-differences analysis. BMJ Open 2018;9.

502 Nolte E, Pitchforth E. What is the evidence on the economic impacts of integrated care? ; 2014.

503 Liljas AEM, Brattström F, Burström B, Schön P, Agerholm J. Impact of Integrated Care on Patient-Related Outcomes Among Older People – A Systematic Review. International Journal of Integrated Care. 2019;19(3).

504 Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. Journal of health economics. 2011;30:466-78.

505 Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53-72.

506 King J, Gibbons E, Graham C, Walsh J. Developing measures of people's self-reported experiences of integrated care Picker Institute Europe & University of Oxford; 2013 October, 2013.

507 Department of Health & Social Care. The UK Rare Diseases Framework 2021 [updated 9 January 2021. Available from: https://www.gov.uk/government/publications/uk-rare-diseases-framework/the-uk-rare-diseases-framework#:~:text=A%20rare%20disease%20is%20defined,being%20identified%20as%20research%20advances.

508 NICE. Diabetes - type 2: How common is it? 2022 [updated June 2022. Available from: https://cks.nice.org.uk/topics/diabetes-type-2/background-information/prevalence/.

509 Torrance GW. Measurement of health state utilities for economic appraisal: A review. Journal of Health Economics. 1986;5(1):1-30.

510 Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health Serv Res. 1973;8(3):228-45.

511 Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-22.