OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Garfield, Kirsty M

Title:
Development of a modular resource-use measure for use in economic evaluations alongside randomised controlled trials

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

DEVELOPMENT OF A MODULAR RESOURCE-USE MEASURE FOR USE IN ECONOMIC EVALUATIONS ALONGSIDE RANDOMISED CONTROLLED TRIALS

Kirsty Marie Garfield

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy (PhD) in the Faculty of Health Sciences.

Bristol Medical School

January 2022

Word count: 67,001

Abstract

Patient-reported resource-use measures (RUMs) are designed for capturing resource-use data in economic evaluations of healthcare. RUMs are commonly developed on a trial-by-trial basis, with validation rarely performed. There does not currently exist a concise, generic RUM that is well-utilised in trial-based economic evaluations. The aim of this thesis was to develop the healthcare module of a new standardised, generic modular RUM (ModRUM) from items previously identified in a Delphi survey, and test the validity, acceptability, and feasibility of ModRUM.

In a review of existing RUMs, the reporting of the development and psychometric assessment of RUMs was found to be rare. The initial prototype of ModRUM was constructed from items identified in the Delphi study, with reference to existing RUMs and in consultation with experienced health economists and patient-reported outcome measure developers.

The measurement properties of ModRUM were tested in four evaluation studies, with modifications made within and between studies. The first study involved qualitative interviews with health economists to test content and face validity and assess whether questions were suitable for costing purposes in economic evaluations. The second study encompassed qualitative ‘think aloud’ interviews including retrospective probing with patients recruited from primary care, to assess content validity and acceptability. In the third study, a user guide was developed, and health economists piloted the adaptation process in recently funded trials. They also provided feedback in an online survey on the feasibility of using ModRUM for resource-use data collection. The fourth quantitative study involved piloting ModRUM with patients recruited from primary care to assess feasibility, acceptability, and construct and criterion validity.

This thesis has provided preliminary evidence for the validity, acceptability, and feasibility of ModRUM. Further testing is required within randomised controlled trials. Future development should include increasing the breadth of ModRUM, to cover sectors beyond healthcare.

Acknowledgements

First, I would like to thank my supervisors, Joanna Thorn, Sian Noble, Samantha Husbands and Will Hollingworth. Thank you for your support and guidance throughout this research, and for being so generous with your time to meet with me regularly and provide constructive feedback on my thesis.

This thesis would not have been possible without the contribution from health economists and members of the public who participated in this research. I am especially grateful to these individuals and to the GP practice staff who have helped me conduct this research.

This research was funded by the MRC Hubs for Trials Methodology Research (HTMR) (MR/L004933/2- PhD Award R9). Thank you to my funders and to Gill Cooper and Jamie Kirkham at the HTMR. Thank you to the cohort of HTMR PhD students and in particular to Nicola Farrar, my fellow HTMR student at the University of Bristol. Thank you also to staff and students at Health Economics Bristol and Bristol Trials Centre for their advice and encouragement.

Throughout this research I have gained invaluable feedback from a range of individuals, for which I am very grateful. Thank you to Kerry Avery, Mairead Murphy and Joanna Coast who acted as annual reviewers and provided insightful comments on this research. I am grateful to Colin Ridyard and Tracey Roberts who provided independent reviews on my research plans, which allowed me to gain support from the Clinical Research Network West of England. Thank you to Matt Ridd for advice and support with regards to recruiting patients via GP practices.

I have presented this research at several conferences. I am grateful to all conference attendees for their interest and comments on this work. In particular, thank you to Katharina Diernberger, Marjon van der Pol and Andrew Briggs for discussing and providing useful comments on this research at conferences.

Finally, I would like to thank my Mom, Dad, Nan Brenda, family and friends for their support and continued interest in my progress. A special thank you to Chris and Ruby for being there for me every day.

This thesis is dedicated to my Nan and Grandad MacLean, who always believed in me and supported me with my academic endeavours.

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE:.....

Table of Contents

Chapter 1 Background to resource-use measurement within health economics 1
1.1 Chapter overview 1
1.2 Introduction 1
1.3 Theoretical foundations of economic evaluation 3
1.3.1 Positive and normative economics 3
1.3.2 Theoretical approaches to economic evaluation 4
1.3.3 Economic evaluation methods 8
1.4 Resource-use identification, measurement, and valuation 13
1.4.1 Identification of resources to capture 13
1.4.2 Measurement of resource-use data 14
1.4.3 Valuation of resource-use data 19
1.5 Rationale and objectives of this thesis 20
1.5.1 Rationale for this thesis 20
1.5.2 Objectives of this thesis 22
1.6 Chapter summary 24
Chapter 2 Methods overview 25
2.1 Chapter overview 25
2.2 Instrument properties 25
2.3 Instrument development 30
2.3.1 Theoretical position of the research 30
2.3.2 RUM development procedures 31
2.3.3 Instrument development procedures 34
2.4 Research methods 39
2.4.1 Review of RUMs 41
2.4.2 Designing ModRUM 41
2.4.3 Qualitative interviews with health economists 43
2.4.4 Patient ‘think-aloud’ interviews with retrospective verbal probing 44
2.4.5 Piloting with health economists 45
2.4.6 Piloting with patients 45
2.5 Conclusion 46
Chapter 3 A review of the development and psychometric assessment of RUMs 48
3.1 Chapter overview 48
3.2 Background 48

3.2.1 Resource-use measurement 48
3.2.2 Resource-use measurement reviews 49
3.2.3 Objective of this review 52
3.3 Methods 52
3.3.1 Search strategy 52
3.3.2 Study selection 53
3.3.3 Data extraction 54
3.3.4 Data synthesis 55
3.4 Results 56
3.4.1 RUM selection 56
3.4.2 RUM characteristics 56
3.4.3 RUM development and assessment steps 60
3.5 Discussion 69
3.5.1 Main findings 69
3.5.2 Strengths and weaknesses of the review 70
3.5.3 Comparison to existing literature 73
3.5.4 Implications for research practice 75
3.5.5 Implications for this research 75
3.5.6 Unanswered questions and future research 76
3.6 Conclusion 76
Chapter 4 Designing ModRUM 78
4.1 Chapter overview 78
4.2 Background 78
4.3 Methods 79
4.3.1 Extraction from RUMs stored in DIRUM 79
4.3.2 Designing a prototype of ModRUM 80
4.3.3 Input from experienced PROM developers 81
4.3.4 Input from a professional design company 81
4.4 Results 82
4.4.1 Review of RUMs and DIRUM prototype design 82
4.4.2 Feedback from experienced PROM developers 89
4.4.3 Aesthetic enhancements and logo design from the design agency 92
4.5 Discussion 92
4.5.1 Chapter summary 92
4.5.2 Strengths and weaknesses 92

4.5.3 Comparison to existing literature 93
4.5.4 Implications for this research 94
4.6 Conclusion 95
Chapter 5 Qualitative interviews with health economic experts 96
5.1 Chapter overview 96
5.2 Background 96
5.3 Methods 97
5.3.1 Health economist identification and recruitment 97
5.3.2 Data collection 98
5.3.3 Qualitative interviews 99
5.3.4 Analysis of expert feedback 99
5.4 Results 100
5.4.1 Identification of experts 100
5.4.2 Recruitment of experts 100
5.4.3 Descriptive summary 104
5.4.4 Research team meeting 120
5.5 Discussion 127
5.5.1 Main findings 127
5.5.2 Strengths and weaknesses 128
5.5.3 Comparison to existing literature 130
5.5.4 Implications for this research 132
5.6 Conclusion 132
Chapter 6 Cognitive interviews with patients 133
6.1 Chapter overview 133
6.2 Background and objectives 133
6.3 Methods 135
6.3.1 Study design 135
6.3.2 Site identification and recruitment 135
6.3.3 Patient identification 135
6.3.4 Patient sampling and recruitment 136
6.3.5 Data collection 138
6.3.6 Data analysis 139
6.4 Results 145
6.4.1 Sites 145
6.4.2 Patients and participants 146

6.4.3 Interviews 149
6.4.4 Scoring results 149
6.4.5 Qualitative results 159
6.5 Discussion 163
6.5.1 Main findings 163
6.5.2 Strengths and weaknesses of this study 163
6.5.3 Comparison with the existing literature 167
6.5.4 Implications for research practice 168
6.5.5 Implications for this research 169
6.5.6 Unanswered questions and future research 169
6.6 Conclusion 170
Chapter 7 Piloting the adaptation process of ModRUM with health economists . 171
7.1 Chapter overview 171
7.2 Background 171
7.3 Methods 172
7.3.1 Study design 172
7.3.2 Health economist sampling and recruitment 173
7.3.3 Data collection 174
7.3.4 Analysis 175
7.4 Results 178
7.4.1 Respondent identification and recruitment 178
7.4.2 Respondent characteristics 179
7.4.3 Trial characteristics 179
7.4.4 Summary of adapted ModRUM 182
7.4.5 Summary of feedback 185
7.4.6 Summary of changes 190
7.5 Discussion 194
7.5.1 Main findings 194
7.5.2 Strengths and limitations of this study 194
7.5.3 Comparison with the existing literature 196
7.5.4 Implications for research practice 197
7.5.5 Implications for this research 198
7.5.6 Unanswered questions and future research 199
7.6 Conclusion 200
Chapter 8 Piloting with patients 201

8.1 Chapter overview 201
8.2 Background and objectives 201
8.3 Methods 202
8.3.1 Study design 202
8.3.2 Site identification and recruitment 202
8.3.3 Patient identification and recruitment 202
8.3.4 Data collection 204
8.3.5 Data analysis 207
8.4 Results 212
8.4.1 Participant characteristics, quality of life and resource use 212
8.4.2 Acceptability 217
8.4.3 Feasibility 219
8.4.4 Construct validity 227
8.4.5 Criterion validity 230
8.5 Discussion 236
8.5.1 Main findings 236
8.5.2 Strengths and weaknesses 236
8.5.3 Comparison to existing literature 241
8.5.4 Implications for research practice 242
8.5.5 Unanswered questions and future research 244
8.6 Conclusion 244
Chapter 9 Discussion and conclusion 245
9.1 Research summary 245
9.2 Strengths and limitations of this research 249
9.2.1 Strengths 249
9.2.2 Limitations 250
9.3 Contribution and implications for applied research 252
9.3.1 ModRUM in comparison to existing methods 252
9.3.2 Approach to developing and testing a new RUM 257
9.3.3 Launching ModRUM 257
9.3.4 Accessing ModRUM 257
9.4 Future development 258
9.4.1 Breadth bolt-on modules 258
9.4.2 Mode of administration 259
9.5 Future testing 260

9.5.1 Testing ModRUM within randomised controlled trials 260
9.5.2 Testing ModRUM with under-represented groups 261
9.5.3 Testing of other measurement properties of ModRUM 261
9.6 Conclusion 262
References 263
Appendix 1. Review of RUMs 276
Appendix 2. Designing ModRUM 277
Appendix 3. Qualitative interviews with health economists 281
Appendix 4. Cognitive interviews with patients 296
Appendix 5. Piloting with health economists 307
Appendix 6. Piloting with patients 322
Appendix 7. Discussion and conclusion 336

List of Tables

Table 1.1 Design principles of ModRUM 24

Table 2.1 Properties of health measurement instruments 27

Table 2.2 RUM development and psychometric assessment steps 33

Table 3.1 A summary of the purpose of published reviews relating to resource-use measurement 51

Table 3.2 RUM characteristics 58

Table 3.3 RUM development steps and psychometric properties 62

Table 4.1 Items identified for inclusion in a new RUM 80

Table 4.2 PROM developer feedback and subsequent actions 90

Table 5.1 Participant characteristics 102

Table 5.2 Project characteristics 103

Table 5.3 Main themes and subsequent actions 121

Table 6.1 User-defined weights for inter-rater agreement 143

Table 6.2 Description and characteristics of participating sites 145

Table 6.3 Summary of participant characteristics 147

Table 6.4 Healthcare use in the last 3 months 148

Table 6.5 Issues and struggles, by participant 151

Table 6.6 Issues and struggles, by issue/struggle category 152

Table 6.7 Issues and struggles, by question 153

Table 6.8 Changes made between and during interview rounds 154

Table 7.1 Health economist responses 178

Table 7.2 Characteristics of responders 180

Table 7.3 Trial and resource-use data collection characteristics 181

Table 7.4 Summary of adapted versions of ModRUM 183
Table 7.5 Adaptations beyond those specified as permitted in the ModRUM User Guide 184
Table 7.6 Descriptive summary of survey responses 187
Table 7.7 Summary of respondent feedback and agreed changes 192
Table 8.1 Number of invites to be sent, by practice 204
Table 8.2 Participant characteristics 214
Table 8.3 EQ-5D-5L scores 215
Table 8.4 Healthcare utilisation and costs, by ModRUM version 216
Table 8.5 Response rates, by GP practice 218
Table 8.6 Participant-reported time to complete ModRUM 218
Table 8.7 Question completion rates, by ModRUM version 220
Table 8.8 Question completion issues, by ModRUM version 221
Table 8.9 Unit costs for the core module 224
Table 8.10 Unit costs for core module with depth questions responses 225
Table 8.11 Costing comparison for participants who completed ModRUM-CD 227
Table 8.12 Known-group analyses: rank test results 228
Table 8.13 Hypothesis testing: generalised linear regression results 229
Table 8.14 Sensitivity and specificity, ModRUM versus medical record data 231
Table 8.15 Agreement between ModRUM and medical record healthcare contacts and costs, by item 233
Table 9.1 Comparing ModRUM to alternatives for data collection 255

List of Figures

Figure 2.1 PhD aims and methods, by RUM development and assessment step 40

Figure 3.1 Flow diagram of RUM identification, screening, eligibility and inclusion .. 57

Figure 4.1 ModRUM core module prototype (first version developed with the research team) 83

Figure 6.1 ModRUM core module - pre-cognitive interviewing 140

Figure 7.1 ModRUM core module - pre-health economist pilot 176

Figure 8.1 ModRUM core module - patient pilot version 206

Figure 8.2 Bland-Altman plots (primary and community care contacts and costs) . 234

Figure 8.3 Bland-Altman plot (prescribed medication cost) 235

Figure 9.1 ModRUM core module (final version) 248

List of Abbreviations

ACQP:	Annotated cost questionnaire for completion by patients
A&E	Accident and Emergency
BNSSG:	Bristol, North Somerset and South Gloucestershire
CADTH:	Canadian Agency for Drugs and Technology in Health
CBA:	Cost-benefit analysis
CCA:	Cost-consequence analysis
CEA:	Cost-effectiveness analysis
CI:	Confidence interval
CINAHL:	Cumulative Index to Nursing and Allied Health Literature
CMA:	Cost-minimisation analysis
COS:	Core outcome set
COSMIN:	Consensus-based standards for the selection of health status measurement instruments
CPRD:	Clinical Practice Research Datalink
CQ:	Core question
CRN:	Clinical Research Network
CSRI:	Client Service Receipt Inventory
CSSRI-EU:	Client Socio-Demographic and Service Receipt Inventory-European Version
CUA:	Cost-utility analysis
DIRUM:	Database of Instruments for Resource Use Measurement
DQ:	Depth question
EQ-5D:	EuroQol 5 dimension
EQ-5D-3L:	EQ-5D 3-level

EQ-5D-5L:	EQ-5D 5-level
GLM:	Generalised linear model
GP:	General practitioner
HES:	Hospital Episode Statistics
HTA:	Health Technology Assessment
ICC:	Intraclass correlation coefficient
ICER:	Incremental cost-effectiveness ratio
ISPOR:	International Society for Pharmacoeconomics and Outcomes
ModRUM:	Modular resource-use measure
ModRUM-C:	ModRUM core module
ModRUM-CD:	ModRUM core module with depth questions
NIHR:	National Institute for Health Research
NHS:	United Kingdom National Health Service
NICE:	National Institute of Health and Care Excellence
PCO:	Primary care organisation
PHR:	Public Health Research
PRISMA:	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM:	Patient-reported outcome measure
PSS:	Personal social services
QALY:	Quality-adjusted life -year
RCT:	Randomised controlled trial
RUM:	Resource-use measure
SD:	Standard deviation
UK:	United Kingdom

Chapter 1 Background to resource-use measurement within health economics

1.1 Chapter overview

This chapter begins with an introduction which outlines the context and overall aim of this thesis. Section **1.3** sets out the theoretical foundations of economic evaluation, including positive and normative economics; approaches used, including welfarism and extra-welfarism; and methods for employing each of these approaches, including cost-benefit analysis, cost-effectiveness analysis and cost-utility analysis. In section **1.4**, literature on resource-use identification, measurement and valuation is summarised. In section **1.5**, the motivation for this thesis is described and the objectives of the research are set out.

1.2 Introduction

Healthcare systems are characterised by unlimited demands but finite resources (1). In the United Kingdom (UK), health expenditure per capita was £3237 in 2018, which was just less than half the per capita spending in the United States of America which was £7969, but considerably more than healthcare spending in the poorest countries in the world, such as Mozambique, where per capita spending in 2018 was £30 (2). Healthcare markets are typified by a number of market failures which mean that they are not consistent with the features of a perfectly competitive market, where an equilibrium between quantity demanded and quantity supplied determines prices (1). Market failures in the healthcare market include externalities, where the collective benefit is often greater than the individual benefit of consuming healthcare, imperfect markets, which include oligopolies in pharmaceuticals and monopsony purchasers, and imperfect information, which includes uncertainty about when healthcare will be required and insufficient knowledge on the value of healthcare (1).

Due to market failures, which mean that healthcare markets would be inefficient and inequitable, governments intervene (1). In high income countries, healthcare systems typically have a large component of public funding. Such government interventions include funding healthcare via taxation, as in the UK, and funding via a social health insurance system, as in Germany (1). In a government-funded

healthcare system, healthcare decision making bodies, such as the National Institute for Health and Care Excellence (NICE) in the UK and the Canadian Agency for Drugs and Technology in Health (CADTH) in Canada, are required to appraise technologies and make choices about how to allocate limited healthcare budgets (3). Governments also fund research, through bodies such as the National Institute for Health Research (NIHR) in the UK, to make sure that high quality evidence is available to decision makers.

Economic evaluation utilises economic theory to provide information that can aid the decision making bodies, so that scarce resources can be allocated efficiently (4). Three important forms of efficiency include technical, productive and allocative efficiency (5). Technical efficiency refers to using the minimum number of resources to achieve a specific output, or achieving maximum output from a set number of resources (6). If an outcome could be achieved with less of one type of input, then an intervention would be technically inefficient (5). When comparing interventions, where one intervention requires more (or less) of one resource, and more of another resource to produce the same (or a better) outcome, the costs of such resources need to be considered (5). Productive efficiency involves minimisation of costs to achieve a specific outcome, or maximisation of output for a specific cost (5). In addition to productive efficiency, allocative efficiency also incorporates efficiency with respect to the distribution of outcomes, with an allocation of resources that maximises outputs for society as a whole, considered allocatively efficient (5).

Economic evaluation involves the measurement and comparative analysis of both the costs and outcomes (or consequences) of alternative treatments (7). Within economic evaluations alongside randomised controlled trials (RCTs), data on the resources patients use (e.g. general practitioner (GP) visits) are combined with corresponding unit costs for each resource to estimate the costs of alternative treatments (8). Accurate data collection is required to ensure that interventions are valued correctly, so that valid conclusions can be drawn, which lead to an efficient use of scarce resources (1). Historically in RCTs, patients have primarily self-reported resource-use and outcome data (9). However, the way that resource use is measured using self-report is sub-optimal, particularly in comparison to the measurement of outcomes, where a wealth of research has been conducted and

research standards have been published (10-12). Self-report resource-use measures (RUMs) are often designed, or previous RUMs are adapted, for each new RCT (9). This lack of standardisation in the way resource-use data are measured inhibits comparability across RCTs (10). The validity and reliability of results obtained from most existing RUMs are also uncertain, as within the time constraints of an RCT, psychometric assessment is rarely conducted (9).

The aim of this thesis is to improve the way self-report resource-use data is measured in economic evaluations alongside RCTs, by developing and testing a new generic, standardised modular RUM (ModRUM), for use in a UK healthcare setting.

1.3 Theoretical foundations of economic evaluation

This section describes how economic theory has informed the development of methods for economic evaluation which are widely employed to help healthcare policy makers make informed resource allocation decisions. First, positive and normative economics approaches are described; then, normative theories, including welfarism and extra-welfarism are described; finally, economic evaluation methods that draw on these approaches (including cost-benefit analysis (CBA), cost-effectiveness analysis (CEA) and cost-utility analysis (CUA)) are presented.

1.3.1 Positive and normative economics

Positive economics can be defined as a theoretical or empirical analysis that involves the *"description or prediction of behaviours and outcomes"* (1) (pg.14). It can provide insight into what is actually happening in the world, by describing or predicting trends in economic variables (1). As positive economics is descriptive and predictive in nature, it does not go beyond this to explore how the information generated should be interpreted to inform decision-making (1). In contrast, normative economics involves using an economic perspective to rank options from better to worse (13). Ranking requires value judgements to be made, where the value of one option is estimated relative to alternative options (1). Normative health economics is used by decision makers to support healthcare resource allocation decisions (14). The role of a health economist in normative health economics is to provide evidence to inform the decision-making process; this involves objectively estimating the

relative advantages of each alternative (13). Different theoretical frameworks are employed for the normative analysis of health and healthcare and include welfarism and extra-welfarism (15). In welfarism, healthcare interventions are ranked based solely on the indirect effect they have on individual expected utility, in other words, health is valued only in terms of the utility that is derived from it (16). While multiple definitions of 'utility' exist, in welfarism, it relates to the satisfaction of desires and is a measurement of preferences (17). Under extra-welfarism, healthcare interventions are ranked based on the direct impact they have on health status, regardless of utility derived from them (16). Welfarism, extra-welfarism and other approaches are described further in the sections below.

1.3.2 Theoretical approaches to economic evaluation

1.3.2.1 Welfarism

The theoretical foundations of welfarism are provided by neo-classical welfare economics (16). In a perfectly competitive market, under welfare economics, decisions are made solely with the aim of maximising societal welfare, which is the sum of individual utility (4). Hurley presented four key tenets that underlie neo-classical welfare economics: [1] *"utility maximisation"*, where individuals act as rational agents, whose aim is to maximise their utility, [2] *"individual sovereignty"*, which purports that individuals themselves make the best judgement on how to maximise their utility, while others' judgements (e.g. healthcare professionals) are considered irrelevant, [3] *"consequentialism"*, which states that individuals only consider the outcomes of their choices, with the route to achieving the outcome deemed irrelevant, and [4] *"welfarism"*, which means that judgements are made with consideration to individual utility only (13) (pg.327). Under a welfarist approach, when evaluating a health service, health is only considered in terms of the utility that it yields (17). To make decisions for society, welfare (utility) is summed across all members of society to derive an estimate of social welfare for each state of the world, with a social welfare ordering providing ranking of all states based on social desirability (1).

To make a value judgement, to determine the optimal allocation of resources (e.g. to make a decision whether to spend UK National Health Service (NHS) funds on a

new type of healthcare), the Pareto principle is widely employed in welfare economics (1). Brouwer et al. summarises the Pareto principle as:

> *"any increase of utility for one individual that involved no utility loss for another was an improvement and an optimum was where no reallocation of resources could be made without reducing at least one person's utility"* (15) (pg.328).

A re-allocation of resources where all individuals experience increased utility is labelled a weak Pareto improvement, while a re-allocation where at least one individual experiences an increase and no individual experiences a loss in utility is labelled a strong Pareto improvement (1). The strength of the Pareto improvement refers to the value judgement that is required, with weak indicating the least contentious judgement, where most people would agree with the judgement (1). An optimum allocation of resources, of which there may be many, is termed a Pareto efficient allocation (1). Equity amongst individuals is not considered under a Pareto framework, the only criterion is that no one is made worse off (1). This fundamental requirement highlights the limitations of this approach within healthcare decision making. Under the Pareto framework, states of the world which make at least one person worse off are not considered and cannot be ranked. For healthcare decision making, the Pareto criterion becomes redundant, as limited budgets mean that in deciding to fund one treatment, another treatment is likely foregone, resulting in utility losses for some individuals (13).

To address this issue and allow for judgement between states of the world that do not meet the Pareto principle, a compensation principle, labelled the Kaldor-Hicks criterion, was introduced (1, 18). The criterion asserts that a potential Pareto improvement can occur if those who experience utility improvements can hypothetically/in theory compensate those who experience utility losses, returning utility losers to their original utility level, while maintaining some improvement in their own utility (18). Even if the compensation does not occur in reality, a potential Pareto improvement means that overall losses are outweighed by gains, which would be considered an improvement to allocative efficiency (1, 18). To put a monetary value on utility changes, Hicks proposed two alternate methods, namely the compensating variation and the equivalent variation (18). Compensating variation considers ex post, following a change, what difference in monetary terms would take an individual back to their prior level of utility (1). Equivalent variation considers ex ante, what

amount would need to be given to an individual to take them to the level of utility that is anticipated following a change (1). Compensating variation can be measured as willingness to pay, which is the amount of money an individual is willing to give up for an improvement in health, to remain at the level of utility prior to the improvement (19). Conversely, willingness to accept is the amount of money an individual would need to be compensated for a reduction in health, to remain at the level of utility prior to the reduction (19).

Within economic evaluation, welfare economics and the compensation principle are most commonly operationalised using CBA, where costs and outcomes are monetised to estimate welfare change. If the estimated welfare change is positive, then a potential Pareto improvement means that an intervention should be implemented; if it is negative, there is not a potential Pareto improvement, so it should not be implemented. Further detail on economic evaluation including CBA is provided in section **1.3.3**.

There are some limitations to welfarism. First, the benefit of healthcare may not be limited to the utility it provides to the individual. Healthcare is considered a merit good, which means that the benefits of consuming healthcare may not only be experienced by the consumer themselves. Second, while efficiency is considered, equity is not. Finding that one state of the world offers a potential Pareto improvement when compared with another, does not necessarily mean it should be implemented, as although it may result in an efficient allocation of resources by increasing societal welfare, it may not be equitable as compensation is hypothetical and will not take place in reality (1). Third, the valuation of benefits in monetary terms may be problematic as it is difficult to undertake and not considered appropriate by governments (14). For example, operationalising the compensation principle through willingness to pay or accept may be difficult for people who are not familiar with paying for their healthcare at point of use, which is the case in a tax-funded healthcare system, such as the UK (14). Willingness to pay will also be impacted by ability to pay which raises equity concerns (14).

1.3.2.2 Extra-welfarism

Where welfarism has limitations for healthcare decision making, non-welfarist frameworks have been developed (1). Extra-welfarism was termed by Culyer in 1989 as an alternative approach to welfarism, which was not as restrictive, but built upon, Sen's work (see section **1.3.2.3** for further details) (14, 20). Extra-welfarism does not fit well under a welfarist framework as the concept of demand is changed to need, and utility to health (13). Morris et al. describe it as supplementing utility with further information about each state, including individual characteristics and preferences regarding consumption of goods, with health being included in the social welfare function to represent individual characteristics (1). While the theory asserts that health and utility are included separately under an extra-welfarism framework, in practice, health is used as a substitute for utility not a complement and the aim in economic evaluation is health maximisation (1). In many economic evaluations, health is quantified using the quality-adjusted life-year (QALY), which incorporates both quality and length of life in a single metric (21).

Extra-welfarism provides an alternative approach that is potentially more suitable for assisting decision makers with resource-allocation decisions, where the most efficient allocation of resources may not be the optimal allocation of resources for society (14). Economic evaluation methods operationalising an extra-welfarist approach, including CEA and CUA, have been adopted by several decision-making bodies, including NICE in the UK and CADTH in Canada (22). Under extra-welfarism, the focus on health means the decision-making bodies, such as NICE, focus on a decision makers perspective for health technology assessment, where the range of costs is limited to NHS and personal social services (PSS) (23). This approach has been considered a narrower approach than would be taken under welfarism, where capturing societal costs is more likely to measure welfare gains and losses of all members of society (15).

1.3.2.3 Other approaches

While economic evaluations are widely performed under an extra-welfarist approach, some health economists argue that the focus on health in economic evaluation under an extra-welfarist approach is too narrow, and that there may be benefits beyond health that should be included and considered in the decision-making process (14).

Most recently there has been significant interest in applying the capability approach in economic evaluation (16). The capability approach is considered a broader approach than extra-welfarism and an extension to the welfarist approach (14). The capability approach shares two commonalities with extra-welfarism; including the importance of function (health) and the preferences that are included are those from the community, as opposed to individual preferences under the welfarist approach (14). The focus of this approach is not simply on utility as in welfarism, or health as in the common application of extra-welfarism, but on *"a multidimensional approach to the measurement of well-being"*, with increased consideration of equity and distributional issues (16) (pg.168). Under this approach, interventions are evaluated with respect to their implications on capabilities, that is, whether an individual can function in a certain way, irrespective of whether they choose to or not (14).

1.3.3 Economic evaluation methods

As budgets for healthcare are limited, decision makers require information on what would be displaced if a new intervention was adopted. The term 'opportunity cost' is used to define the benefit foregone of the intervention that is no longer funded (1). Economic evaluation is a systematic approach used in healthcare decision making where the costs, which represent opportunity costs, and consequences of alternative uses for scarce resources are identified, measured, valued and compared (7). Drummond et al. succinctly define economic evaluation as:

> *"the comparative analysis of alternative courses of action in terms of both their costs and consequences"* (7) (pg.9).

The following sections define the four most commonly employed types of economic evaluation: CBA, CEA, CUA and cost-consequence analysis (CCA) (24).

1.3.3.1 Cost-benefit analysis

CBA is the primary method for operationalising welfare economic theory within economic evaluation (25). In CBA, the aim is to achieve allocative efficiency, that is, the optimal allocation of resources to maximise societal welfare (5). In a CBA, all costs and consequences (benefits) are estimated and converted into monetary values to determine whether net social welfare has increased or decreased (24). The

difference between costs and benefits may be presented as a net benefit (cost-benefit) ratio which represents welfare change (25).

CBA operationalises the Kaldor criterion; assuming there are no resource constraints, if societal benefits outweigh societal costs of a new healthcare intervention, then net social welfare has increased, which means an intervention should be implemented and the healthcare budget increased if necessary (7, 21, 24). However, when resources are scarce, which is the case for healthcare systems with fixed budgets, the cost-benefit ratio of different options should be compared and the options with the largest improvements in social welfare funded, until the budget is exhausted (24).

For valuation of outcomes, willingness-to-pay operationalises the compensating principle to assess potential Pareto improvements (7). Criticisms of the approach are described in section **1.3.2.1**, and include issues with putting a monetary value on health and human life, and with the willingness-to-pay approach favouring the wealthy (26). There are a number of distinguishing features of CBA, when compared with CEA and CUA, which can be considered advantageous properties. CBA requires that both costs and consequences are monetised in equivalent units. Consistent with CEA and CUA, this allows for a comparative analysis across interventions, but it also offers the advantage of providing an estimate of the absolute benefit or loss of each individual intervention in monetary terms (1). Another advantage is that the monetisation of all costs and outcomes, means that programme comparisons can not only be made within the healthcare sector, but across sectors, which is not an option for CEA and CUA which focus predominantly on health outcomes (1).

1.3.3.2 Cost-consequence analysis

CCA is a less well-utilised method (24). In a CCA, both costs and consequences are estimated and tabulated against one another; however, no attempt is made to combine the results to present them in an aggregated way (24). CCA is nested within extra-welfarist theory and is useful when there are a wide range of outcomes (e.g. for an RCT in a public health setting, where outcomes are potentially beyond utility and health) (22). To make resource allocation decisions, a trade-off is still required

between the costs and consequences. In a CCA, the decision maker takes responsibility for interpreting the tabulated information, from which they rank alternative options (21). While this approach provides decision makers with a comprehensive account of all relevant costs and consequences, for decision making the onus is on the decision maker to weight the importance of costs and consequences which may not lead to optimal, transparent or consistent decisions for patients or society (27).

1.3.3.3 Cost-effectiveness analysis

CEA provides a method for measuring costs and effects under an extra-welfarist framework (7). Extra-welfarist CEAs and welfarist CBAs are the most common applications of each framework (21). Under an extra-welfarist framework, in a CEA, costs are compared with effects for competing treatments/interventions (7, 21). Effects are measured in a common metric, which can include a broad range of health-related outcomes, such as depression score, life years, and deaths (7, 21). To perform a comparative analysis, the difference in costs (incremental cost) and difference in effects (incremental effect) are estimated (24). If one treatment is less costly and more effective than another, then it is said to dominate the more expensive treatment (24). It is often the case that a new intervention is more costly and more effective than existing intervention. In this case no intervention is dominant, so the incremental cost-effectiveness ratio (ICER) can be estimated, which is the ratio of incremental costs and incremental effects (24). The ICER equation for comparing two interventions a and b can be presented as:

$$\text{ICER} = \frac{Incremental\ cost}{Incremental\ effect} = \frac{Cost_a - Cost_b}{Effect_a - Effect_b}$$ (24).

When no intervention is dominant, due to budget constraints, the decision maker should decide whether the cost per unit of effect is acceptable and the more expensive intervention should be funded (24). A limitation of CEA is that it is restricted to addressing technical efficiency, that is decisions within patient groups, unless the outcome is QALYs, in which case allocative efficiency can be addressed, but only within the healthcare sector (21). In order to judge whether an intervention should be funded, an external threshold value is required to compare the ICER to

(21). In addition, it has limited capacity to influence resource allocation decisions between patient groups or interventions when outcome measures differ (21).

1.3.3.4 Cost-utility analysis

The conceptual framework that underpins CUA is also extra-welfarism (28). Utility in CUA reflects the preferences individuals or society have for different health states (7). CUA can address questions of technical efficiency and allocative efficiency within the healthcare system. CUA is a special case of CEA where quality-adjusted life expectancy is the measure of effect, with the QALY the most common measure (21, 24). QALYs encapsulate both length and quality-of-life in a single metric, to capture both changes in morbidity and mortality (21). One QALY is equivalent to one year in full health and zero is a health state equivalent to death (21).

CUA is recommended by NICE in the reference case for conducting economic evaluations (23). CUA addresses several of the limitations experienced in a CEA. The outcome measure in a CUA can capture multiple attributes of health; therefore interventions that impact multiple or different aspects of health can be compared (21). As the QALY can be used as a common outcome measure across a wide range of conditions, the cost per QALY of unrelated healthcare interventions can also be compared (21).

To capture quality of life, generic preference-based measures of health are widely employed (21). The most commonly used instrument is the EuroQol five dimension (EQ-5D) which includes five dimensions: ‘mobility’, ‘self-care’, ‘usual activities’, ‘pain/discomfort’, and ‘anxiety/depression’ (29). There are multiple versions of the EQ-5D, including the EQ-5D 3-level (EQ-5D-3L) which includes 3 levels per dimension and the more-recently developed EQ-5D 5-level (EQ-5D-5L), which includes five levels per dimension (29, 30). The EQ-5D-5L was developed to overcome concerns associated with the sensitivity of the EQ-5D-3L (30). The EQ-5D-3L defines 243 health states, while the increased number of levels in the EQ-5D-5L means that it defines 3125 different health states. Preference-weights for each health state of the EQ-5D-3L and EQ-5D-5L have been elicited from the general public (31, 32). However, at present, NICE recommends that, rather than using the

EQ-5D-5L preference-weights, preference-weights for the EQ-5D-5L should be derived using a validated mapping function from the EQ-5D-3L (33, 34).

Participants taking part in RCTs are usually asked to complete generic preference-based measures of health at several timepoints during the study period, including baseline. To estimate QALYs, preference-weights are applied to scores and scores are combined with length of time in each state using the area under the curve approach (35). As with a CEA, when no arm is dominant, an ICER can be estimated which presents incremental cost over incremental QALYs of two interventions. In order to make funding decisions, policy makers compare this figure to a threshold which defines the maximum willingness to pay for a unit of effect. In England, the threshold for one additional QALY is £20,000 to £30,000 (36). Interventions with an ICER under £20,000 will generally be funded, interventions with an ICER between £20,000 and £30,000 would be considered by the decision maker alongside other information, and interventions with an ICER above £30,000 would require a strong case to be funded (36). In some instances, there are modifiers which mean NICE apply a much higher threshold. For example, under the Highly Specialised Technologies programme, which covers interventions that treat very rare conditions, a threshold of £100,000 is applied (37). The net-benefit framework allows costs and QALYs to be combined with the threshold willingness-to-pay value and expressed as an incremental net monetary benefit (iNMB):

$$\text{iNMB} = (\text{incremental QALYs} * \text{threshold value}) - \text{incremental cost}$$ (24).

If the iNMB is positive at the specified threshold, the intervention should be funded and if it is negative it should not be funded (24).

Although in practice unlikely, due to uncertainty around estimates of outcomes, when the outcomes in a CEA/CUA are equivalent, a CEA/CUA can reduce to a cost-minimisation analysis (CMA) (38). CMA may be implemented when no difference is observed in outcomes between two interventions and the objective of the economic evaluation becomes the minimisation of costs (7). A CMA is not usually chosen a priori as it would require prior evidence showing equivalence of interventions with

respect to their outcomes (7). Furthermore, it has been shown that CMA leads to biased estimates of uncertainty which impact the probability that an intervention is considered cost-effective (39). CMA should only be considered when the difference in cost is so large that no plausible difference in efficacy could make the more expensive intervention cost-effective (39).

1.4 Resource-use identification, measurement, and valuation

Extensive research has been conducted in the area of outcome measurement, while research in the area of cost measurement for economic evaluation has been relatively limited (10). The focus of this thesis is to address issues on the cost side, for application in trial-based economic evaluations, but which also aims to improve the quality of data from RCTs that may ultimately be used in model-based economic evaluations, which extend beyond the trial period and can include a broader range of comparators (40, 41). Several decisions need to be made on the cost side of economic evaluation including what resources to measure, how to measure resource usage and how to generate cost data from resource-use data (42). The following three sections summarise research in each of these areas.

1.4.1 Identification of resources to capture

Estimation of the incremental difference in costs and outcomes is key to conducting a comparative analysis in economic evaluation (7). To estimate the difference in costs it is important to identify and measure key cost drivers; these are resources where there is likely to be a difference in resource consumption between intervention arms (43). It is also important to measure resources that contribute a large proportion to the total cost of care (43). For economic evaluation, including as many resources as possible is unlikely to be a practical or appropriate approach, as the purpose of an economic evaluation is to conduct a comparative analysis and not to estimate the full cost of a disease, which would be undertaken in a cost-of illness study (43).

The first stage of identification is to create a list of all resources that an intervention is likely to have an effect on (42). The scope of resources to include will depend upon the perspective of the economic evaluation (42). The most appropriate perspective to take varies between countries and is impacted by how the healthcare

system is organised (44). In the UK, the NICE technology assessment process takes what could be considered a narrow NHS and PSS perspective in the reference case when interventions are funded by the NHS and PSS and there are health outcomes (23). When interventions are funded by the public sector, NICE advocates a broader perspective, which may encompass a public sector or societal perspective (23). From a welfare economics stance, it has been argued that a societal perspective should be routinely taken as it can capture the impact on societal welfare, as a healthcare intervention may impact societal welfare in sectors beyond the NHS and PSS, such as criminal justice or education (45).

When identifying which resources to collect, researchers also need to decide whether to collect condition-specific or all-cause resource use (43). Collecting condition-specific data requires ease in distinguishing related and unrelated resource-use which will vary dependent on the condition and who (e.g. a patient, clinician or analyst) is categorising the data as relevant or not relevant. Collecting condition-specific resource-use data only may reduce participant burden and can increase precision (46).

1.4.2 Measurement of resource-use data

Once all relevant items have been identified, the next stage is to measure the usage of these resources, which will allow identification of changes in resource consumption, and subsequently estimated costs, throughout the trial period between the trial arms (42). Prior to deciding where the information will be obtained from a decision needs to be made as to whether a bottom-up or top-down costing approach will be employed. Bottom-up costing (or micro-costing) is a more time-consuming approach that may be necessary when new services are established (e.g. a new surgical intervention) or existing services are changed (e.g. single- versus multi-port laparoscopic surgery) (42, 47). When comparing similar procedures in a trial, micro-costing may be required when available aggregate-level national unit costs for similar procedures are the same (e.g. are grouped under the same Healthcare Resource Group (HRG) in the National Schedule of NHS costs) (42, 47, 48). Top-down costing is a simpler approach and considerably less time consuming as less detailed information is required, it is more appropriate for resources where less sensitivity in is required (42). A combination of costing methods is often used in

trials, where the trial intervention and comparator are often costed with greater sensitivity using the bottom-up approach, and all subsequent resource-use over the trial period is costed using a top-down approach (42).

There are multiple methods that are used to obtain resource-use data in economic evaluations. Data can be obtained from administrative records, including national databases and primary care electronic medical records, or directly from patients themselves either prospectively using a diary or retrospectively using a questionnaire (49). Expert opinion may also be used for resource-use estimation but is not commonly utilised as it is not generally deemed an unbiased and reliable method (9). To capture data at the required depth for micro-costing, case report forms or more intensive observation, such as time-motion studies, may be completed prospectively by the research team or healthcare providers, or data from hospital administrative or information systems, may be used (47). Researchers may choose to employ one or a combination of these methods to obtain resource-use data in an RCT (42).

Franklin and Thorn have outlined 17 aspects to consider when choosing between administrative sources and self-report for the collection of resource-use data, including the level of detail required for costing, perspective, access to the required data and patient characteristics (49). Many studies have compared resource-use measurement captured via self-report and administrative data and some authors conclude that support can be given for equivalence of the methods (50), while others have concluded that the preferred method is dependent upon the resources being captured (51, 52). While there is still uncertainty surrounding the optimal method, in a Delphi study including health economic experts from around Europe, it was concluded that patient self-report is the optimal method for resource-use data collection in economic evaluations, due to resource coverage and availability (53). The following two sections further outline administrative and self-report methods for measuring resource-use data and discuss the strengths and weaknesses of each approach.

1.4.2.1 Administrative data

Administrative data encompasses a wide array of sources, including data directly from healthcare providers, such as records from individual GP practices, to national

databases, where data from multiple healthcare providers are collated in one dataset (e.g. the Hospital Episode Statistics (HES) (54)), and healthcare insurance companies, which may be more relevant to non-UK based evaluations (49). Administrative data may be preferred to self-report data when the participant is unlikely to have sufficient knowledge to provide the level of detail required on the resources of interest (10). Where participants in RCTs are required to complete many questionnaires, collecting information on resource utilisation from administrative data can also reduce participant burden. As data are retrieved directly from healthcare records, it may be more accurate than relying upon participant recall; however, the accuracy of healthcare records is also rarely tested (49).

While at face value it may seem that collecting resource-use data directly from administrative data could save time as opposed to collecting self-report data, in reality collecting data from multiple sources can be a time-consuming process (49). In England, there is no central database where all data of NHS funded healthcare is recorded. The variety of potential administrative data sources is highlighted by Franklin and Thorn and includes multiple sources for primary care, secondary care and mental health care data (49). Administrative datasets are not primarily developed with research in mind, as their purpose is for keeping a record of patient healthcare and for reimbursement. As such, the dataset may not include all relevant information for an economic evaluation and may require substantial data processing to transform it into a suitable format for analysis. Oftentimes the collection of all relevant data will require collection from multiple sources (49). To access administrative data, researchers require the relevant approvals; however, as this is a time-consuming and expensive process, particularly if data are being obtained from multiple administrative sources, it may not be feasible to use some data sources and prepare them for use in an economic evaluation within the scope of an RCT, where timelines are not usually informed by the economics (49).

Administrative data available for use in UK-based RCTs primarily relate to healthcare resources. Beyond healthcare, the availability and existence of administrative data become sparse. Even from the NICE reference case, where an NHS and PSS perspective is recommended, while collecting NHS data may be feasible, collecting administrative PSS data may be more troublesome (49). Administrative PSS data

are less suitable for use in economic evaluations due to reasons including more fragmentation in social care provision, less standardisation in coding and absence of national data dictionaries (49). When taking a societal perspective, there are also likely to be resources for which no administrative data are available, such as informal care.

1.4.2.2 Patient-reported data

In RCTs, patient-reported resource-use data can be collected via self-report, interviewer-administration, and proxy completion. Data can be collected retrospectively or prospectively. Trial participants can retrospectively report their resource use in resource-use questionnaires (labelled resource-use measures (RUMs) hereinafter) or in interviews. Prospective data collection includes resource-use logs and resource-use diaries. While a resource-use diary may be used as the primary source of data, a resource-use log is generally designed to act as a memory aid for participants to help participants complete RUMs retrospectively, but is not collected from participants as a source of data (55). RUMs can include questions framed in variety of ways, including open-ended and closed questions, standalone questions, tables and skip logic. Skip logic guides respondents through a questionnaire and allows them to skip questions that are irrelevant based on previous responses (e.g. if a respondent answers ‘no’ to a ‘yes’ or ‘no’ question, they would skip a follow-up question that is only relevant to respondents who answered ‘yes’).

Patient-report is generally considered a less costly and more practical method for obtaining resource-use data (49, 53). While ethical approval is required to obtain resource-use data from trial participants, it is generally more readily accessible than administrative data which is often subject to stringent information governance procedures, and subject to delays between resources being used and corresponding data being available for research. Trial participants can provide data, without delay, on their utilisation of a wide range of resources which may be included in an economic evaluation from the societal perspective (49).

Limitations associated with patient-report include comprehension and recall issues, patient burden, accuracy, missing data, and insufficient detail for cost estimation. In

trials, participants are required to recall their healthcare utilisation over potentially long time periods, often ranging from two to six months (9). Extended recall periods are associated with reduced accuracy and less detailed information (56). Recall issues may be exacerbated for less salient events (e.g. GP visits), when compared with more salient events (e.g. hospital admissions), as information may be more difficult to recall and as a consequence, responses may be less accurate (56). As recall periods extend, telescoping and reverse telescoping may become more common (56). The former leads to the inclusion of records from outside the recall period, while the latter refers to exclusion of records from inside the recall period (56). The optimal recall length may be dependent on the resource and the level of detail being measured (57). Deciding on the optimal recall period is a balance between recall bias, due to patients forgetting an encounter or incorrectly recalling it, and data completeness, where a short recall period is not sufficient to capture information from the whole period required (41). For economic evaluations alongside RCTs, these inaccuracies are more problematic if there is a systematic difference between trial arms as they lead to biased estimates of cost-effectiveness (49).

Missing data is a key limitation that is common when collecting resource-use data via self-report (58). Research has found that as the detail that is requested for each item increases, so does the amount of missing data (59). However, it is unclear whether the reason for this is increased burden leading to none or partial completion, or whether patients cannot recall more detailed information about their healthcare (59). Increasing the demand on memory by asking for further detail can reduce acceptability of a RUM, increase the proportion of missing data and lessen the accuracy of reported data (60, 61).

A systematic review of RUMs, with respect to their development and psychometric assessment is provided in **Chapter 3**. The review includes RUMs stored within the Database of Instruments for Resource Use Measurement (DIRUM) (62). Prior to the introduction of DIRUM, there has long been recognised a paucity of evidence focused on resource-use data collection methods and inconsistency in how RUMs are described (10, 63, 64). RUMs were rarely published and if they were, they were often nested within the appendices of funder reports. DIRUM has increased

transparency in resource-use measurement, by providing a repository where developers can upload their measure in addition to details on the psychometric properties of the measure (62). It has also highlighted a large overlap in the questions included in existing RUMs (62).

1.4.3 Valuation of resource-use data

As reported in section **1.3.3**, the value of a resource is the opportunity cost, which is defined as the benefits foregone of not using the resources in their best alternative use (7). Valuation of resources involves applying relevant prices as a proxy for opportunity cost. Under perfect competition, market prices provide reasonable estimates of opportunity cost, as a perfect market is likely to be in equilibrium (65). However, as healthcare markets are subject to market failure (see section **1.1** for further detail), a pareto efficient output is not achieved and market prices are distorted so they do not represent opportunity costs (1, 65).

For some resources, such as healthcare contacts and services, there are average unit costs available which can be used as an estimate for opportunity cost. For micro-costing, some costs may need to be sourced directly from suppliers or from hospital finance departments, and staff costs may be sourced from national data sources. Unit costs are monetary values attached to each item of resource. For UK-based economic evaluations there are multiple sources for obtaining unit costs; the National Schedule of NHS costs (formerly known as the National Schedule of Reference Costs) is commonly used for obtaining secondary care unit costs, the Unit Costs of Health and Social Care is commonly used for obtaining primary and community care unit costs and the British National Formulary is commonly used for obtaining unit costs of prescribed medications (48, 66, 67). Within each source there are often multiple unit costs available for valuing resources at various levels of granularity.

When taking a broader perspective, identifying unit costs becomes more challenging, particularly for ‘non-market’ goods where a cost cannot be directly observed (7). Two of the most predominant non-market resources included in economic evaluations from a patient or societal perspective are lost time from usual activities and informal care (7). Where a price is not available, shadow pricing methodology may be used to

estimate a monetary value for ‘non-market’ resources which is then used as an estimate of opportunity cost (65). For the valuation of informal care, methods include contingent valuation to measure willingness-to-pay or willingness-to-accept, the opportunity cost method where informal care is valued as wages foregone; and the replacement cost method where informal care is valued using professional care prices, which is the nearest substitute for informal care (25, 68).

1.5 Rationale and objectives of this thesis

1.5.1 Rationale for this thesis

Research in the area of resource-use measurement has trailed behind outcome measurement research (10). For outcome measurement, there is a precedent for the psychometric assessment of outcome measures, with guidelines provided in the consensus-based standards for the selection of health status measurement instruments (COSMIN) taxonomy of measurement properties (69), and the development of condition-specific standardised core outcome sets, which include the minimum set of outcomes that should be measured and reported in all trials of a condition (70). Uptake of these standards has not yet translated to resource-use measurement. At present there is a lack of consistency in the way resource utilisation is measured, with researchers developing bespoke RUMs for each new trial, or adapting existing RUMs (9). This lack of standardisation inhibits the comparability of results across trials, which impedes the ability of decision makers to efficiently allocate resources across the healthcare system (71). Several reviews of RUMs have been conducted and have highlighted that psychometric assessment is rarely performed (9, 50, 51, 72). For the minority of RUMs where psychometric assessment has been performed, it is most commonly conducted by validating the RUM with administrative data and often only includes a subset of items from the RUM where administrative data are available (51). Validating against another data source may be considered insufficient, as this is only a test of criterion validity (assuming administrative data is considered a ‘gold standard’ compared with self-report data); for outcome measures it is first important to assess the content validity of an instrument to ensure it is relevant, comprehensive and comprehensible (73).

To date, the most commonly employed and extensively validated RUM is the Client Service Receipt Inventory (CSRI) (72, 74, 75). The CSRI was originally designed to

be interviewer-administered for collecting data on the health and social care utilisation of people with mental health problems (74). Several other categories were collected including accommodation, employment and informal care (74). Since the inception of the CSRI, it has been translated for use in a wide range of countries, patient groups, and modes of administration (10, 75). When citing the CSRI, many researchers report using a modified or adapted version; however, it is rarely clear what modifications have been made and whether modifications inhibit the validity of the results obtained. The lack of clarity also hampers the ability to make comparisons across trials.

The annotated cost questionnaire for completion by patients (ACQP) was developed in response to health economists indicating it would be a valuable resource (76). Despite this, uptake has been low. The reason for this is unclear; however, it has been speculated that there may be issues with usability, as preparation of the measure for use in a trial involves the formulation of the RUM from a large battery of questions, provided in a lengthy document (71). Other standardised RUMs have been developed for specific conditions; however, uptake of such measures has also been limited (71).

In a review of methods for conducting economic evaluation alongside RCTs, Hughes et al. highlighted a need for more robust methods for collecting resource-use data, which would increase the accuracy of results from which valid conclusions can be drawn (44). Assuming self-report is likely to continue to be the most commonly used data collection method for resource-use data in RCTs due to time constraints, expense and coverage of alternative data sources, a new generic standardised RUM, that has undergone a thorough development and assessment process, could be a valuable resource (49, 53). Creating a generic standardised RUM that could be used in all trials, irrespective of the clinical setting, patient group and intervention under study, would increase transparency and enhance the comparability of results. A new RUM would require a balance between standardisation, to facilitate comparison of results across trials, and flexibility, to encourage uptake among researchers, as there may be resources that are pertinent to capture in some trials, that are not relevant to others, and the level of granularity required for each question may vary by trial, dependent on the precision required for costing (77). While

different resources and sectors may be relevant across conditions, increased standardisation could be implemented for disease groups, as relevant resources are likely to be consistent within disease groups.

Key considerations in the development of a new RUM would be to ensure it is well-utilised by researchers and well-received by participants in trials. To encourage uptake of a new RUM, the instructions should be clear and concise, with usability testing conducted to allow the identification of any impediments to use. In addition to establishing a RUM that is fit for purpose, increasing uptake of a new RUM would involve creating awareness among potential users of the new RUM and the advantages it could offer over existing methods. Once the RUM has been adopted in a trial, the extensive development and assessment process would help to ensure questions are concise, comprehensible and acceptable to patients, which could improve the quality of data obtained (78). Consideration of how data will be costed at the development stage could also mean that data obtained from the RUM should be easy to value using available unit costs, particularly for concise questions which do not require free-text responses.

1.5.2 Objectives of this thesis

A well-utilised, standardised generic RUM, designed for self-completion, does not currently exist. Research has suggested that a concise list of key healthcare resources could encapsulate the majority of total costs (78). In a Delphi consensus survey conducted prior to this PhD, described in detail in **Box 2.1**, health economists identified ten core items, listed in **Table 4.1**, that should be collected in all trial-based economic evaluations (71). The aim of this PhD research was to develop these items into, and perform initial validation of, a new standardised, generic modular RUM (ModRUM), that can be used for measuring healthcare utilisation data in economic evaluations conducted alongside UK-based RCTs. In **Table 1.1**, the design principles of ModRUM and the purpose of each principle are outlined. With reference to the results of the Delphi study, I outlined the design principles, which were subsequently refined and agreed in collaboration with the research team (PhD supervisors) (71). To meet my aim of developing and performing initial validation of ModRUM, the objectives of my research were to:

1. Systematically review the development and assessment of existing RUMs (**Chapter 3**).
2. Design a prototype of ModRUM from items identified in the Delphi survey (71) (**Chapter 4**)
3. Design a ModRUM User Guide (**Chapter 7**)
4. Assess whether ModRUM:
    a. appears to measure what it is intended to measure (face validity) (**Chapter 5**)
    b. measures all important and intended resources (content validity) (**Chapter 5** and **Chapter 6**)
    c. is acceptable to patients (**Chapter 6** and **Chapter 8**)
    d. is suitable for costing purposes in trial-based economic evaluations (**Chapter 5**, **Chapter 7** and **Chapter 8**)
    e. is feasible to adapt for use in trial-based economic evaluations (**Chapter 7**)
    f. correlates to other constructs as hypothesised (construct validity) (**Chapter 8**)
    g. produces results that adequately reflect a ‘gold standard’ (criterion validity) (**Chapter 8**).

Table 1.1 Design principles of ModRUM

Design principle	Purpose
Standardised	To increase consistency in the way resource-use data is captured, which will improve the comparability of RCT results.
Generic	To be relevant for collecting resource-use in a wide range of RCTs of healthcare.
Flexible	To increase uptake, by allowing users to make pre-defined adaptations, including adding bolt-on modules, to ensure relevance for a wide range of healthcare RCTs.
Precise	To allow for more detail to be captured for key cost drivers or highly utilised resources, for increased accuracy in cost estimates.
Concise	To minimise patient burden and reduce missing data.
Comprehensible	To minimise patient burden, improve accuracy and reduce missing data.
Transparent	To allow greater clarity of what and how resource-use data have been captured.

1.6 Chapter summary

In this chapter, I have outlined the context, and overall aim of this thesis, which is to develop and perform the initial validation of a new modular RUM (ModRUM) which can be used to collect participant-reported healthcare resource-use data in a wide range of economic evaluations alongside RCTs. In the following chapter, I provide an overview of the methods used in this thesis. More detailed methods are provided in chapters that follow the methods chapter.

Chapter 2 Methods overview

2.1 Chapter overview

This chapter is separated into four sections. In section **2.2**, the desired properties of a measurement instrument are defined. The relevance of each property to a RUM is also considered. In section **2.3**, the theoretical position from which this PhD is conducted is described, followed by a summary of literature on RUM development and a summary of general guidance on instrument development. In section **2.4**, the aims and methods of this PhD research are outlined. The chapter is concluded in section **2.5.**

2.2 Instrument properties

Prior to administering a new instrument, it is important to establish that the instrument is suitable and valid for measuring the construct it is intended to capture, in the target population. In **Table 2.1**, I have summarised definitions from several sources which define the desirable properties of an instrument (11, 69, 73, 79-81). To my knowledge, there does not exist a well-defined list of desirable properties of a RUM; the list below was informed by literature advising on health measurement scale and patient-reported outcome measure (PROM) development. There are distinct differences between outcome measures and RUMs, which are outlined further in the next paragraph. As a result of these differences, some properties that are considered important for an outcome measure may not be considered relevant for a RUM. Properties that I did not consider relevant to RUMs are indicated in **Table 2.1**. Further information on the properties I considered to be relevant to the development of ModRUM is reported in sections **2.3.3.1** to **2.3.3.4**.

When completing a PROM, individuals typically report their health on a range of health dimensions (21). PROMs *"do not include biomedical measures"* but *"they can assess symptoms, function or well-being"* (21) (pg.14-15). PROMs typically measure unobservable constructs, including unidimensional (e.g. pain) and multidimensional constructs (e.g. health-related quality-of-life) (11). Summary scores of health are then estimated by applying numerical scoring systems to individual responses (21). PROMs, and specifically generic preference-based measures from which QALYs

can be generated for economic evaluations, are generally concise and are collected at intervals (e.g. every 3 months), with interpolation the accepted method to estimate QALYs over the whole study-period. In contrast, RUMs are often lengthy and are administered at several time periods throughout a trial, with the aim of capturing all resource utilisation over the entire trial period. Interpolation of resource-use data is less preferable than data collection that covers the entire trial period, as intermittent resource-use data can lead to biased estimates, as rare, high-cost events (e.g. inpatient stays) may or may not be observed within the data collection period (82). Data collected in RUMs are used to estimate total costs, by applying unit costs to each resource, and summing costs across resources. Once developed, adaptation of a PROM is not typically considered acceptable. In contrast, items included in RUMs may need to evolve over time to reflect changes in health services (e.g. NHS 111 replaced NHS Direct in England in 2014).

Table 2.1 Properties of health measurement instruments

Property	Definition[1]
Content validity	While some attributes can be directly measured (e.g. blood pressure), others cannot (e.g. anxiety) (79). For the latter, we can only observe behaviours that are hypothesised to be associated with the attribute. Behaviours (e.g. going to work despite illness) can inform factors (e.g. motivation) associated with an attribute, and the factors can be referred to as a 'construct' or 'hypothetical construct' (79). Content validity is the extent to which the content of an instrument adequately covers the important and relevant aspects of the construct it is intended to measure (69, 79-81). Content validity is considered to be the most important psychometric property, as the assessment of other properties cannot be tested until it has been established that the measure is relevant, comprehensive and comprehensible (73).
Face validity	The extent to which, on the surface, an instrument appears as though it measures what it is intended to measure (69, 79). While analogous to content validity, face validity is a complementary property that should also be tested (81).
Construct validity (hypothesis testing)	Assuming that hypotheses are valid, the extent to which scores obtained from an instrument are consistent with predicted hypotheses about the relationship between the instrument and other variables or instruments (69). It can encompass convergent validity (the instrument should be related to other variables and instruments which measure the same construct) and discriminant validity which includes divergent validation (there should be no relationship with unrelated constructs) and known group validation (the instrument is able to discriminate between groups where a difference in outcome is anticipated) (69, 79).
Construct validity (structural validity)	The extent to which scores of an instrument adequately reflect the dimensions of the construct it is designed to measure (69). Only relevant for multi-item PROMs (69).

Table 2.1 continued

Property	Definition[1]
Construct validity (cross-cultural validity)	When an instrument is translated or adapted for use in a different language and/or culture, this property is concerned with whether the adapted instrument is an adequate reflection of the original instrument (69). Scores from equivalent questions in adapted and original versions of an instrument should be the same for patients with the same severity of illness (11). Different scores would suggest that the items measure different things for different populations (11).
Criterion validity	The extent to which the scores of an instrument correlate with an accepted 'gold standard', where the 'gold standard' is considered to measure the construct of interest accurately (69, 79, 81). Criterion validity can be divided into concurrent validity and predictive validity (79). The former considers the correlation between instruments administered at the same time, the latter considers the new measure against a 'gold standard' which will be available at a later date (79). Some have contested whether 'gold standards' actually exist for health measurement, with the only exception being a longer version of an instrument (81). In the absence of a 'gold standard', assessment of the relationship of two instruments measuring the same construct, can be labelled as the assessment of construct validity (convergent validity).
Responsiveness	Over time, a change in the construct of interest should be detected by an instrument designed to measure the construct (69). Responsiveness differs from validity, as in validity testing the focus is on assessing the validity of a single score at one point in time, whereas for responsiveness the focus is on assessing the validity of a change score (11). Responsiveness has been used interchangeably with 'sensitivity to change', with some describing a distinction between the two related to whether the change is important or meaningful (79).
Reliability (including test-retest, inter-rater, intra-rater)	Assuming circumstances remain the same, the extent to which scores are reproducible when an instrument is administered at different times (test-retest), to different raters at the same time (inter-rater) or to the same rater at different times (intra-rater) (69, 79). For a self-report instrument, it is relevant to assess reliability via test-retest reliability.

Table 2.1 continued

Property	Definition[1]
Reliability (internal consistency)	The extent to which items are interrelated (69). Relevant to instruments where multiple items are designed to measure the same dimension (e.g. physical function) (79).
Interpretability	The extent to which qualitative meaning (e.g. minimally clinically important difference) can be applied to quantitative scores or change in scores from an instrument (69, 81). Historically, interpretability is an issue with PROMs rather than more objective measures (e.g. blood pressure) (81).
Acceptability	The acceptability of an instrument concerns respondent burden and considers aspects including completion time and difficulties in responding (81). Acceptability can be indicated via instrument response rates and item completion rates (81).
Feasibility	The extent to which it is feasible, with reference to time and cost, for a researcher to administer and analyse instrument responses, and for a respondent to complete the instrument (79, 81).
Precision	The extent to which an instrument has a sufficient number of response categories to make distinctions between health states (81). Relevant to PROMs (81).

1: Definitions included in this table are summarised from several sources, exact terminology and definitions vary in the literature.

2.3 Instrument development

In their comprehensive guide to the development of health measurement scales, Streiner and Norman assert that the process of developing an instrument is a laborious, time-consuming and iterative process that involves the use of both qualitative and quantitative methods (79). In section **2.3.1,** the theoretical position from which this PhD is conducted is described. In section **2.3.2**, literature which provides guidance on RUM development is summarised. Compared with other fields there is a paucity of literature on RUM development methods. As many of the properties of health measurement scales are also desirable for RUMs, in section **2.3.3**, relevant guidance and methods from other fields, including PROM and core outcome set (COS) development, are summarised.

2.3.1 Theoretical position of the research

Prior to this PhD research, my research predominately consisted of employing quantitative methods to perform economic evaluations alongside RCTs. Research in economics is most often performed under a positivist epistemological paradigm, which is the belief that there is a single reality which can be researched (83, 84). From literature on RUM and health measurement scale development, it was evident that a comprehensive approach to the development and assessment of a RUM should involve both qualitative and quantitative methods (11, 79, 85). To employ these methods and meet the aims of this PhD thesis to develop and assess a new RUM, this PhD research aligns with a theoretical perspective of subtle realism (83), as advocated in previous health economics research which combines quantitative and qualitative methodology (86). This approach acknowledges that there is an independent, knowable reality; however, as researchers it is not possible to access this reality or represent it with certainty (83). Instead, it is only possible to provide a representation of the reality, which relies on the perspectives or cultural assumptions of research participants and is likely to be influenced by the observations and assumptions of the researcher (83). Using this approach, I was able to employ qualitative and quantitative methods to develop a version of ModRUM, that may not be the ‘right’ (only possible) version of ModRUM, but is a credible version that includes items that are relevant and acceptable to the study population.

2.3.2 RUM development procedures

Relative to health measurement scale development, there is a lack of literature which describes the process of developing a RUM. There are several journal articles, that I am aware of, where some information on the development and assessment steps for a RUM are provided. I have summarised these articles in **Table 2.2**, where the information provided in each article has been divided under two development steps: 'item identification and formulation' and 'piloting and psychometric assessment'. Each article is also described below, in further detail.

The first article, by Chisholm et al. (2000), is not defined by the authors as a guidance document but rather reports on the development of an adapted version of the CSRI (the Client Socio-demographic and Service Receipt Inventory – European Version (CSSRI-EU)) (75). Within the article the authors included a table detailing the stages of development for an instrument designed to collect resource-use data. Although the stages of development were defined, limited detail was provided on the assessment of measurement properties during these steps. The only measurement properties included in the table were face validity and semantic equivalence (cross-cultural validity). However, in the discussion, the authors indicated that the paper reported on the earlier stages of development, and assessment of reliability and validation with an alternative source (criterion validity) should follow in further research. In the DIRUM listing for the CSSRI-EU, evidence of criterion validity is reported and a link to a publication by Patel et al. provided, which describes a study comparing data collected via the CSRI, which the CSSRI-EU was adapted from, and primary care records (87).

In 2010, Ridyard and Hughes systematically reviewed the health economics methods of 85 publications citing Health Technology Assessment (HTA) studies reporting an economic analysis using patient-reported resource-use data (9). From their review of methods employed in existing studies, the authors devised a good practice checklist for resource-use data collection in HTA trials. The checklist starts from selecting the appropriate perspective (e.g. healthcare provider perspective) to standards for reporting the results of an economic evaluation. The authors state that items for inclusion *"should be identified a priori from consultation with health-care professionals, pilot studies, or literature searches"* to identify main cost-drivers,

provide justification for included resources and achieve external validity (9) (pg.871). They state that if patient-report is selected as the appropriate source of resource-use data, piloting should be performed to assess clarity, ease of use, completion rates and inform main cost-drivers. They also indicated that where feasible, validation with administrative data should be performed.

In 2013, Thorn et al. summarised the themes raised at a workshop on resource-use measurement of patient level data in UK-based economic evaluations (10). The authors summarised the decisions, challenges and processes associated with the development and application of RUMs. Decisions to be made included defining the perspective, who to obtain data from (e.g. self-complete or proxy complete) and recall period. They defined multiple procedural steps which included item identification and formulation, and an iterative process of assessment and testing. Methods reported in the article included clinician and patient focus groups during item identification, cognitive interviewing with patients or volunteers and patient piloting. The authors outlined the following properties to assess during the process: validity (face, content and criterion), acceptability, completion time (respondent burden) and reliability (such as test-retest). Detailed descriptions of the methods suggested were beyond the scope of the article; however, the authors cited several textbooks where the methods are comprehensively described (79, 88, 89).

Most recently in 2021, Janssen et al. summarised literature on aspects and challenges to consider for resource-use measurement, which included things to consider when developing a RUM (46). While information on item selection was not covered, reference to item formulation was considered with the authors stating that validation is required to assess whether participants have the same understanding of terminology included in a RUM, which would be required for a RUM to have content validity. The article also details the relative advantages and disadvantages of alternative modes of administration, and discusses whether resource use related to a specific condition or all resource use should be captured. For RUMs, the article also asserts that the measurement properties of the RUM, that is validity and reliability, should be considered.

Table 2.2 RUM development and psychometric assessment steps

Summary of steps	Chisholm et al. (2000) (75)	Ridyard and Hughes (2010) (9)	Thorn et al. (2013) (10)	Janssen et al. (2021) (46)
Item identification and formulation	- Identification of the *main desirable characteristics* of a measure in this area. - Examine existing measures and select a measure to adapt or develop a new measure, aiming for *face validity*. - Translation to other languages. - Focus groups to refine the content and translated version.	- Identification of resource use items, including *main cost-drivers*, through discussion with healthcare professionals, literature searches or pilot studies.	- Identify items for inclusion, including *main cost-drivers* via literature review, identification of existing measures and focus groups with healthcare professionals and patients. - Compare, contrast discuss existing questionnaires, agree on cost-drivers, agree on scales and draft measure.	- Terminology of items should be validated to assess whether interpretation is as expected (*content validity*). - Consideration given to mode of administration, recall period and type of resources to capture
Piloting and psychometric assessment	- Assess performance of the measure with respect to *response rates*, *ease of completion*, *preferred aggregation*. - Make further revisions, ready for widespread use.	- Patient/carer piloting to test *clarity*, *ease of use*, *completion rates* and determine the main cost drivers. - Assess *validity* via comparison to alternative of resource use data collection methods.	- Cognitive interviewing of patients/volunteers to assess *acceptability*, *content validity, completion time* and *administration mode*. - Revise draft instrument (with *content and face validation*). - *Reliability* testing with patients/volunteers. - *Patient piloting* to test *reliability* and *criterion validity.*	- Assessment of the psychometric quality, including *validity* and *reliability*, required, potentially validating using multiple methods for collecting resource-use data.

2.3.3 Instrument development procedures

The articles summarised in the previous section highlight that some guidance on RUM development and assessment is available; however, the literature is limited to outlining an approach to take, without detail on how to operationalise the information to develop and test a RUM. As such, this section reports on guidelines to develop health measurement scales, PROMs and COSs, which I utilised to formulate the approach to RUM development implemented in this PhD research.

In their highly cited textbook on health measurement scale development, Streiner and Norman suggest that the development process should include item generation, testing and revision of items, and testing of reliability, generalisability and validity (79). In the development of PROMs, de Vet et al. defined six intertwined steps including: [1] construct definition, [2] item development, [3] response option development, [4] pilot testing, [5] field testing and [6] evaluation of measurement properties (11). A summary of the best practices in PROM development and content validation are also provided in two International Society for Pharmacoeconomics and Outcomes (ISPOR) task force reports (80, 85).

As the steps listed above are in the context of health measurement scale and PROM development and assessment, some may be deemed irrelevant for the development of a RUM. For example, field testing, where the aim is to reduce items and assess the structure of data, potentially using techniques such as factor analysis, is only relevant to multi-item measures of unobservable constructs (11). I judged the following steps to be relevant and important to undertake during the development of a RUM: [1] item identification, [2] RUM formulation (including instructions, questions and response options), [3] psychometric assessment (including face validity, content validity, suitability and acceptability) and [4] piloting (user testing, with further assessment of psychometric properties). In the sections below, methodological options and considerations are presented for each step.

2.3.3.1 Item identification

Once the phenomenon to be measured has been defined, it is important to ensure that all important and relevant items are included so that the instrument formulated has content validity (79). Item development has already been highlighted as an

important stage in the RUM development process as it allows the identification of key cost drivers (9). For multi-item measurement instruments, recommendations for the identification of items include literature reviews, reviews of existing instruments and expert clinical opinion (11). For instrument development, patients with the condition of interest may also be involved in the item identification process via qualitative methods including focus groups and interviews (79). Clinicians may also participate in qualitative studies including interviews and focus groups, or they may be observed in clinical practice (for example, to identify items to micro-cost surgery) (79). Inclusion of both patients and clinicians may be appropriate, as patients can provide insight into subjective elements of a condition, while clinicians can provide insight into the outward signs and symptoms of a condition (11). Use of the appropriate methods during item development can contribute to the content validity of an instrument, if all relevant items are identified (79). Qualitative methods are advantageous in the early stages of developing an instrument as they are more flexible and allow for greater exploration than quantitative methods, with considerably smaller sample sizes (90). They allow research questions to be answered that could not be answered via quantitative research (90).

Items for inclusion in ModRUM were initially identified in a Delphi study conducted prior to this PhD research by several members of the supervisory team (Thorn, Noble and Hollingworth) and colleagues (71). The aim of study was to:

> *"identify a minimum set of core resource items that should be included in a standardised adult instrument for UK health economic evaluation from a provider perspective"* (71) (pg.640).

A summary of this study is provided in **Box 2.1**. The authors recommended that the new RUM would have a modular structure with a core healthcare module, including the 10 items identified in the Delphi study (71). The authors suggested optional extended modules could be developed, including depth questions which would allow the end-user to capture more detail on items included in the core module, and healthcare resources that did not make it through the final round of the Delphi study (e.g. paramedic contacts) (71). They also suggested bolt-on modules which would increase the breadth of the RUM, covering sectors such as social and residential care, which may not be relevant to capture in all trials (71).

As the aim of the Delphi study was to identify a standardised core set of items for a generic RUM, the authors drew parallels with the process of developing a COS (71). COSs are standardised sets of outcomes that should be collected at a minimum for trials in specific clinical areas (91). Following the Delphi study that preceded this PhD, recommendations for the development of COSs were published, which included three domains: [1] scope specification requires the specification of the setting, health condition, population for which the COS is relevant (92); [2] stakeholders involvement requires the involvement of potential users (i.e. health economists for a RUM), clinicians and patients in the COS development process; [3] consensus process requires a transparent, rigorous and unbiased consensus process, which includes the initial identification of a list of potential outcomes (or items in the case of a RUM) and the consensus procedures (e.g. pre-defined criteria for inclusion, omission and addition of outcomes) (92). The most appropriate method for developing a COS has not been determined; however, the Delphi technique, often in combination with other methods has become the most popular approach for COS development (70, 93). During a Delphi study to develop a COS, participants who are considered to have the relevant expertise are sent a survey and asked to provide their opinion on inclusion of outcomes (94). Following the first round, participants are presented with their opinions alongside group feedback from the previous round and are asked whether their opinion has changed or not (94). Rounds are repeated until consensus is reached (94).

The Delphi study that proceeded this work had strengths in that the authors recruited 45 participants, with low attrition between rounds, pre-defined consensus criteria and clear consensus for the final items (71). While the sample was representative in some respects (e.g. varied NHS research experience), there was low representation from non-academic health economics (<7 percent) (71). A further potential issue with the survey, was that items were considered individually, as opposed to in the context of other items being captured (71). In the final consensus meeting, items were considered together but only one member of the Delphi panel was able to attend this meeting alongside the research team (71).

Box 2.1. Summary of the Delphi study conducted by Thorn et al. (71)

The study was conducted in three phases. In phase one, items were extracted from 59 existing instruments that were used in RCTs of health interventions. Similar items were combined, and non-NHS or PSS items were dropped. The resulting list of 60 items formed questions for the Delphi survey.

In phase two, 45 practising health economists took part in round one of the Delphi survey, where they rated the importance of items for inclusion in a core standardised set of resource use items from an NHS and PSS perspective. Items were retained if they met prespecified criteria. In round two, participants were sent the revised list of 34 items alongside their original score and the median score of all participants and a summary of relevant comments from round one. Forty-two participants rerated items in round two. Items were retained if they met prespecified criteria.

In phase three, members of the research team and one Delphi participant attended a meeting to agree items for a brief standardised RUM. The group agreed on ten items that made up a core set of healthcare resource items of a new standardised RUM, that may be relevant for most trials, conditions, and patient groups. They also identified items that could form extended and bolt-on modules.

2.3.3.2 Instrument formulation

When formulating an instrument from a list of items, it is important to consider comprehensibility, clarity and conciseness (11). With respect to comprehensibility, things to consider include: reading level, ambiguity, double-barrelled questions, jargon and value-laden words (79). A balance should be sought when deciding the length of items, so that they are as concise as possible, without losing meaning (79). A further consideration, which applies to RUMs, is heterogeneity of items, which for a RUM is required to omit the issue of double-counting (79). Formulation of an instrument also requires response options to be developed, which may be a more onerous task for PROMs than RUMs, as developers also need to consider how to scale responses and how overall scores will be estimated from responses to items (11, 79).

2.3.3.3 Psychometric assessment

To explore the process of answering questions and identify problems with items, in health measurement scale development several cognitive methods have been derived from the field of psychology, to use with representatives from the target population, including *"rephrasing, double interviewing, thinking aloud interviews, and probing"* (79) (pg.123). In a rephrasing interview, participants are asked to rephrase questions using their own words, while in a double-interview, participants complete the instrument and the researcher asks follow-up questions on several items to gain greater understanding on how questions were answered (79). 'Think-aloud' interviews involve respondents completing questionnaires while verbalising their thought processes (88). Probing can be performed concurrently or retrospectively to instrument completion, to gain greater understanding or clarity about answers provided and to ask about areas of interest to the researcher (88). Concurrent probes avoid retrospection problems but have the potential to influence answers to subsequent questions, retrospective probes can overcome this interviewer-imposed bias as minimal input is required from the interviewer during questionnaire completion (88).

For PROM development, agreement was reached in a consensus study that cognitive interviewing, using think aloud and/or verbal probing, should be undertaken to complete the content validation process of an instrument prior to quantitative testing of other measurement properties (85). Cognitive interviewing provides the opportunity to assess comprehension of the wording and formatting of an instrument, and evaluate the comprehensiveness to confirm that no important items are missing (85). Sampling strategies for cognitive interviewing studies should target the patients who are similar to those likely to complete the instrument in an RCT, with purposive sampling to recruit patients who are likely to offer unique perspectives (e.g. for ModRUM, patients who have accessed different services) and patients who are more likely to experience difficulties comprehending or completing the instrument (85). There is no set sample size for such studies; however, the number of interviews may be higher for more complex instruments and for more diverse target populations (85).

For the assessment of content validity within cognitive interviews, analysis using a standardised classification scheme based on Tourangeau's survey response model has been recommended (85, 89). The classification scheme breaks down the cognitive process of answering questions into four actions including comprehension of the question, retrieval of information over the recall period, judgement of what is relevant or irrelevant and response formatting (89). There are also a range of approaches that can be used to analyse interview data qualitatively, including some that are inductive and theory generating (e.g. grounded theory) and others that are deductive, where theory or hypotheses are tested, with pre-developed classification systems (e.g. content analysis) (95). The constant comparison technique, which draws on grounded theory, has been used to assess opinions about economic outcome measures following completion of such measures (96). Techniques of constant comparison involve continual comparison of participant responses and emerging themes, to develop key patterns and themes, to enhance understanding (97, 98).

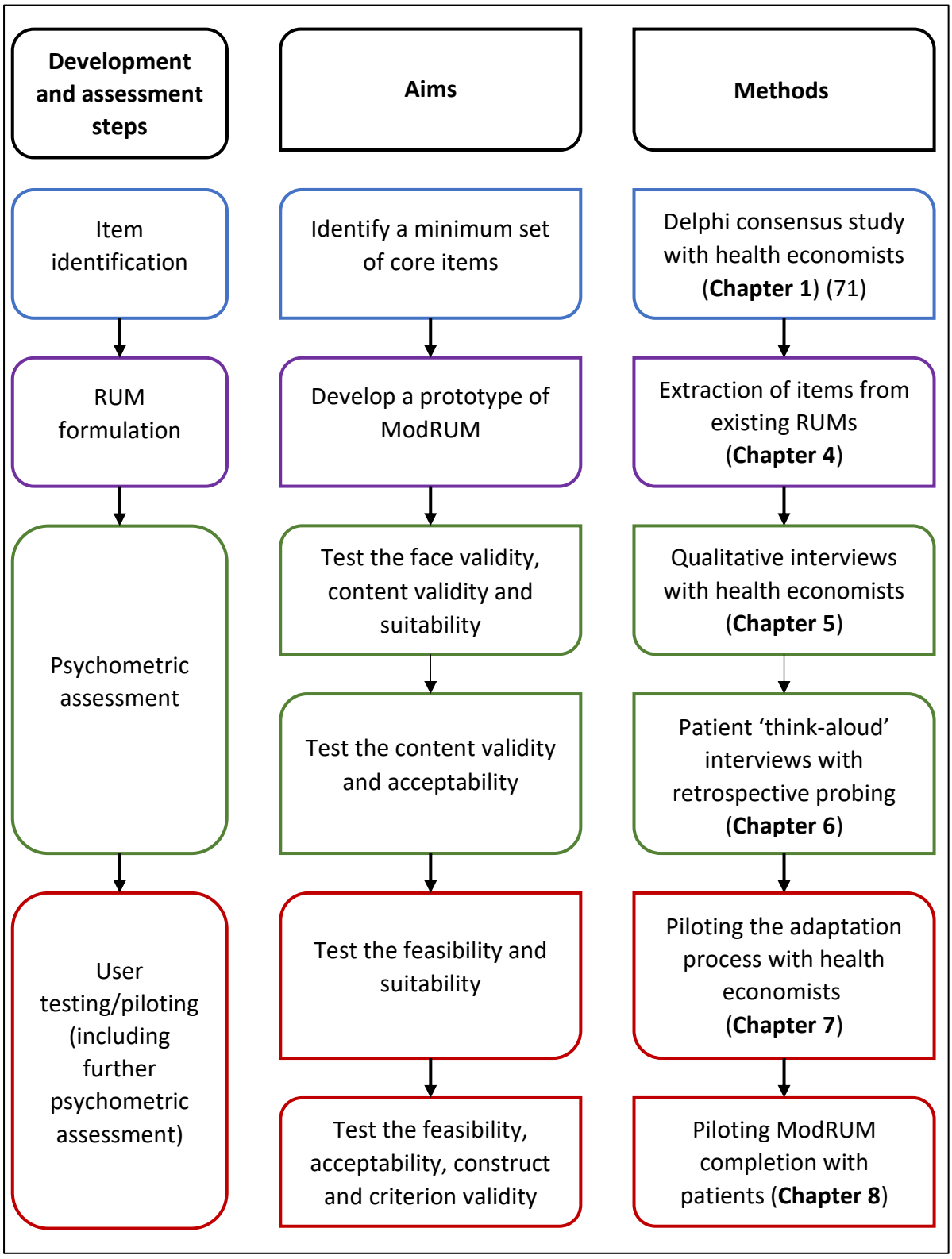
2.3.3.4 Pilot testing

Pilot testing, with representatives from the target population, provides the opportunity to identify and rectify problems with an instrument prior to wider usage (11, 79). It is often conducted when an instrument is almost in its final format (11). In comparison to qualitative methods that are used in earlier development stages for item identification and content validation, a larger quantitative study permits the assessment of other measurement properties including feasibility, acceptability (completion and response rates), construct validity, criterion validity and reliability, which are defined in **Table 2.1** (11). Assessment of missing data in a larger study can also reveal whether there are issues with particular items, which may provide an indication of respondent burden (11).

2.4 Research methods

In this section, the aims and methods of each chapter are summarised. More detail is provided in each chapter. **Figure 2.1** provides an overview of the steps taken to develop and assess ModRUM.

Figure 2.1 PhD aims and methods, by RUM development and assessment step



2.4.1 Review of RUMs

In the absence of comprehensive guidance on RUM development, the aim of the review, reported in **Chapter 3**, was to identify:

1. what methods have been employed in the development of existing RUMs which were designed to capture healthcare utilisation from patients,
2. what psychometric properties have been assessed for included RUMs and
3. what methods were used to perform psychometric assessment.

The review followed the four phases outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for systematic reviews: [1] identification, [2] screening, [3] eligibility and [4] inclusion (99). To identify records reporting the development of RUMs for inclusion in the review, first I conducted an extensive search of multiple sources including DIRUM (62), electronic bibliographic databases and reference searching. Next, I conducted title and abstract screening of identified records, followed by full text eligibility assessment. Articles were included if they were original research that included details on the development of a RUM that was primarily developed for costing healthcare within economic evaluations, designed for retrospective data collection, and primarily designed for self-report by adults. Articles were not included if they were not published in the English language, not freely available via the University of Bristol library or published only as a conference abstract.

From included articles, I extracted information on RUM characteristics, development steps and psychometric assessment. RUM characteristics included year of publication, country/setting, population and administration mode. Development steps included item development and pilot testing. Psychometric assessment included validity (content, face, criterion, construct) and reliability. I performed a narrative synthesis for each development step and psychometric property.

2.4.2 Designing ModRUM

Items for the core healthcare module of ModRUM were identified in a Delphi consensus survey (71); however, this process did not establish how questions should be formulated from the items. The aim of the research conducted in **Chapter 4** was to:

1. review existing RUMs, to identify how questions are formulated and what terminology has been used in existing RUMs and
2. formulate prototypes for the core and core and depth healthcare modules of ModRUM.

All questions that related to one of the ten core items (**Table 4.1**) identified in the Delphi study were extracted from RUMs stored within DIRUM that were designed for use with adult participants. For each RUM, details on each item were extracted and recorded on a spreadsheet in Microsoft Excel. Information extracted included question wording (e.g. GP description), layout (e.g. skip logic) and formatting (e.g. bold font). The number and percentage of RUMs using each option were estimated.

The research team were sent a summary of all options and the number of RUMs using them. For each detail, members of the research team independently indicated whether they had a preference. I then led team meetings where I presented collated preferences and the team agreed on question details and identified details which required patient input. Based on agreed details, I drafted the core, and core and depth modules of ModRUM, following questionnaire design principles, such as keeping questions concise without losing meaning and avoiding jargon (79). To keep the core module concise, the initial version was kept to one-page and no questions were included which required free text responses. The core and depth module was designed to offer flexibility to the end user and allow them to capture more detail for more accurate costing. This included questions which required free text responses (e.g. clinic type). An iterative process followed, which involved presenting the modules at team meetings, agreeing on revisions, re-drafting the modules, and taking them back to the team.

Once the team agreed that the modules were ready for feedback from beyond the research team, I sent them via email to two independent researchers, based at the University of Bristol, who were experienced in PROM development. Informal face-to-face meetings were arranged, where I met with each researcher individually to ask for feedback. I summarised the feedback and presented it to the research team. Final revisions were agreed and subsequently implemented prior to formal testing.

Input was also sought from an external design agency. Based on the specifications I provided, the agency prepared a logo and enhanced the aesthetics of ModRUM. An iterative approach was followed until the research team and I were satisfied with the design of ModRUM and the logo.

2.4.3 Qualitative interviews with health economists

The aim of the research reported in **Chapter 5** was to:

1. test the face and content validity of ModRUM and
2. assess whether ModRUM was suitable for costing purposes in economic evaluations.

Qualitative interviews were conducted with UK-based health economists, who were identified as co-authors on NIHR reports, which detailed economic evaluations conducted alongside RCTs. A purposeful sampling strategy with maximum variation was used to recruit experienced 'information rich' (particularly knowledgeable and experienced in the area of interest) health economists with a range of characteristics, based on funding stream, workplace, geographical region, research project setting, and disease area/condition/preventative intervention (100). I sent invitations via email and arranged telephone interviews with health economists who agreed to participate.

During semi-structured qualitative interviews, I asked participants pre-defined questions from a topic guide, which I prepared prior to interviews. The topic guide included questions on the relevance, clarity, conciseness, and omission of items and whether ModRUM appeared to measure what it was intended to measure (79, 101). Using a responsive interviewing approach, I was able to be "flexible and adaptable" and gain greater understanding of participant responses (102). Interviews were audio-recorded and transcribed. Analysis drew on methods of constant comparison to identify themes (98, 103). I presented findings, grouped under key aspects of feedback with relevant illustrative quotations, to the research team, and adaptations to ModRUM were agreed.

2.4.4 Patient ‘think-aloud’ interviews with retrospective verbal probing

The aim of the research reported in **Chapter 6** was to:

1. test the content validity of ModRUM and
2. test the acceptability of the content, length and layout of ModRUM;

by performing cognitive interviews with patients.

Patients were recruited via their primary care organisations (PCOs). A purposeful sampling strategy was used to recruit patients who were active users of healthcare services (80, 100). Maximum variation sampling was used to recruit a diverse sample; based on sex, age, ethnic group, number of long-term conditions, age on leaving full time education, number of recent healthcare contacts; that reflected the wide range of patients that could complete ModRUM in an RCT context (104). Interviews consisted of an established warm-up exercise, the think-aloud exercise and a semi-structured interview, where I asked participants questions to clarify issues that occurred while completing ModRUM and questions on prespecified areas of interest from a topic guide I developed (88, 105).

Analysis first involved a data scoring activity, whereby ModRUM questions were scored for errors in comprehension, retrieval, judgement and response (85, 89). Struggles were also scored when a participant appeared to struggle with a question but were able to reach the correct answer (106). Following independent rating, raters (SH, JT and KG) met on four occasions to compare scores and reach consensus on a final score. Inter-rater agreement was assessed using Gwet's agreement coefficient (107, 108). The second approach involved a qualitative analysis including constant comparison, to continually compare and contrast new data with existing data and codes (97, 98).

Sampling, interviews and analysis were performed concurrently and in rounds. Following each round I summarised the findings, with illustrative quotations, into analytic accounts (95). I then presented my findings and made suggestions on whether ModRUM should be revised to the research team. Agreed changes were implemented prior to further interviews. Interviews continued until sufficient data was collected to enable identification of issues with completion and important themes.

2.4.5 Piloting with health economists

The aim of the research reported in **Chapter 7** was to:

1. test the feasibility of adapting ModRUM for use in an RCT, and
2. test the suitability of ModRUM for capturing healthcare utilisation data that is appropriate for costing purposes in economic evaluations;

by recruiting health economists to adapt ModRUM for a recently funded RCT and asking them to provide feedback in an online survey. I devised this study once it was clear that the study piloting ModRUM with patients would be delayed due to the Covid-19 pandemic.

Health economists were identified as co-applicants on recently funded NIHR grants. Email invitations were sent to health economists, who had the option to take part themselves, forward the invitation to another health economist working on the grant, or participate together. I developed a user guide for ModRUM, and this was sent with ModRUM to health economists who agreed to participate. Respondents were asked to adapt ModRUM as if they were going to use it to capture resource-use data for their grant and complete a brief online survey that I prepared in Online Surveys (109). The survey included closed questions, with responses captured as ‘yes’ or ‘no’, or on 5-point Likert scales (110). Open questions followed closed questions to gain further reasoning for responses (111).

Respondents’ trial-specific versions of ModRUM were reviewed and I summarised the adaptations which had been made. Simple descriptive statistics were used to present quantitative data. Data was also uploaded to NVivo 12, where techniques of constant comparison were used to identify whether common themes arose between respondents (100). I convened research team meetings, where I presented the results and agreement was reached on how the findings should inform further development.

2.4.6 Piloting with patients

The aim of the research reported in **Chapter 8** was to:

1. test the feasibility and acceptability of completing ModRUM,
2. test the construct and criterion validity of ModRUM, and

3. test the feasibility of identifying and applying unit cost data to ModRUM responses;

by recruiting patients to complete the core module or core module with depth questions and comparing ModRUM responses with primary care medical records.

Patients were recruited from the PCOs that took part in the cognitive interviews. Postal invitations were sent to patients who had had a consultation with a clinician in the past month. ModRUM, the EQ-5D-5L and a patient characteristics form were included with the invitation and patients who agreed to participate were asked to complete and return the documents and a consent form. Participant reported data was entered into a Microsoft Access database. Data from participant's primary care medical records was obtained on their primary care consultations and prescribed medications used during the three-month recall period they completed ModRUM for. All data was uploaded to Stata 17, where I performed data cleaning and analyses (112).

Cleaning the data, and sourcing and applying unit costs to ModRUM data provided an indication of the feasibility of using ModRUM for costing purposes. Simple descriptive statistics were estimated for question completion rates, response rates and participant-reported completion time to provide an indication of feasibility, acceptability and respondent burden (11). Construct validity was assessed via hypothesis testing. With reference to published literature, pre-specified hypotheses were formulated for the correlation between ModRUM data and validation items, including patient characteristics and health-related quality of life. Criterion validity was estimated as the level of agreement between self-reported resource-use data captured in ModRUM and data from primary care medical records. To compare binary reporting of healthcare utilisation, sensitivity and specificity were estimated with medical record data considered the 'gold-standard' (113). For continuous data, Lin's concordance correlation coefficient was estimated (114).

2.5 Conclusion

In this chapter, the desired properties of a measurement instrument are defined, including an indication of their relevance to RUMs; literature relevant to the development and assessment of a RUM is summarised; and an overview of this

thesis is provided, including summaries of the objectives and methods employed in each chapter. In the next chapter, a detailed description of the review of existing RUMs is provided.

Chapter 3 A review of the development and psychometric assessment of RUMs

3.1 Chapter overview

This chapter reports on a review of the development and psychometric assessment of existing RUMs. Articles reporting RUM development and assessment were identified from: (1) a search of DIRUM (www.DIRUM.org) (62), (2) a bibliographical database search and (3) reference searching. Articles were included if details on the development of a RUM, including item development and/or piloting, were reported. For each RUM included, information was extracted from corresponding articles on item development, pilot testing, reliability, and construct, content, face, and criterion validity.

This chapter includes background information about existing evidence on the development and psychometric assessment of RUMs and the objective of this review. The methods section outlines the search strategy, how studies were selected, which data were extracted and how the data were summarised. The results section details the number of studies identified, screened, assessed and reviewed; the characteristics of included RUMs; and a narrative synthesis of the development steps and psychometric assessment of included RUMs. The discussion outlines the main findings of this review, strengths and limitations, how this review compares to existing literature and how this review can inform future research. A brief conclusion is provided at the end of the chapter.

3.2 Background

3.2.1 Resource-use measurement

Patient self-report has been recommended as the optimal method for resource-use data collection in economic evaluations due to resource coverage and availability (53). The strengths and limitations of using self-report to collect resource-use data are outlined in section **1.4.2.2**. Despite being a recommended source and the wide use of self-report RUMs for capturing resource-use data within RCTs, there is no standardised generic RUM that is relevant and well-utilised in a wide range of trials (9, 49). Instead, researchers often design bespoke RUMs or adapt an existing RUM

for each trial they conduct, which represents a duplication of research effort and inhibits the comparability of results (9, 71).

To reduce the limitations associated with RUMs, they should be developed in a way that minimises patient burden with respect to RUM length and comprehensibility, while enabling sufficient detail to be captured to allow for precise cost estimation of key cost drivers. A review of RUMs stored within the DIRUM found problems with the presentation and readability of existing RUMs, suggesting a need for additional effort during RUM development to minimise respondent burden (115). In the development of a new RUM, questionnaire design literature should be utilised which could enhance the presentation of the RUM and subsequently minimise respondent burden, improve data accuracy, and minimise partial (individual questions) and complete (RUM not returned) missing RUM data (115). Avoidance of missing data is particularly desirable in an economic evaluation as total cost is generated from the sum of the costs of individual resource-use items, meaning that one piece of missing data leads to the patient being dropped from the analysis (in a complete-case analysis) or imputation of missing data, which can lead to biased estimates if inappropriate methods are employed (58, 116). In a review of 52 trial-based CEAs published between 2013 and 2015, Leurent et al. identified that only five studies reported less than five percent of participants with missing data and across the included studies, the median proportion of patients with complete data was 63 percent (interquartile range 47-81 percent) (117). Leurent et al. collated recommendations from research papers that were focused on missing data in trial-based CEAs and reported that a common recommendation was to minimise missing data from the outset, by designing a RUM that is user-friendly (117).

3.2.2 Resource-use measurement reviews

An initial step in the development of a new measure is to review existing measures, including the methods undertaken to develop them (79). In order to provide context and identify the unique contribution of this review, I first studied six published articles, which I was aware of, that included reviews of the development and/or psychometric properties of existing RUMs (9, 50, 51, 56, 72, 118). A summary of the purpose of each of these reviews is provided in **Table 3.1**.

Four articles only reviewed evidence on validation from comparisons with alternative data sources (e.g. administrative data) (50, 51, 56, 72). Findings from these four articles included that it is often only a subset of a questionnaire that is validated (51), validation is uncommon for non-healthcare resources (118), self-report healthcare utilisation data is of variable accuracy (56, 119) and self-report is more accurate for less common, more memorable events (such as, hospitalisations) (51). While Noben et al. concluded that the exchangeability of self-report and administrative data can only be supported cautiously, with no obvious optimal source for resource-use data (50), Ridyard and Hughes reported good agreement for eight of the 12 studies included in their review (72).

Two articles included information beyond validation with an alternative data source (9, 118). Ridyard and Hughes (2010) reported on item identification, piloting and validation with alternative data sources in studies funded by the UK HTA programme (9). They found that less than 25% of studies provided any evidence of a systematic approach to the identification of resources and evidence of piloting was found for 37% of studies (9). The definition of piloting was wide, and a measure was considered to have been piloted if it was based on an existing measure that had been piloted elsewhere (9). Mayer et al. aimed to review the characteristics and psychometric properties of RUMs that included items capturing the impacts on the criminal justice and education sectors (118). As RUMs needed to include criminal justice or education items, many were designed for capturing resource use of children or adolescents, with proxy-completion by parents (118). The authors reported that most studies provided some information on RUM development, such as, the RUM was based on an existing RUM, and that nine RUMs had been piloted (118).

In the articles described above, the focus has predominantly been on reviewing studies which have validated self-report against alternative data sources. Only two articles included other aspects of developing RUMs and detail on these aspects was limited.

Table 3.1 A summary of the purpose of published reviews relating to resource-use measurement

Authors (year)	Study summary
Mayer et al. (2017) (118)	Through searches of DIRUM and bibliographical databases, the aim was to identify and provide an overview of RUMs which include items related to intersectoral costs and benefits (i.e. non healthcare sector) items capturing the impacts on the education and criminal justice sectors and review psychometric evidence, including validity and reliability, of included RUMs.
Leggett et al. (2016) (51)	Through a search of bibliographical databases, the aim was to identify validated self-report RUMs and summarise the attributes, including validation approach (e.g. with administrative records or patient diary) and validation results, of identified RUMs.
Noben et al. (2016) (50)	Through a search of NHS Economic Evaluation Database and a supplementary Medline search, the aim was to identify and review studies assessing the exchangeability of resource-use data collected via self-report and administrative records and assess the methodological reporting quality of these studies.
Ridyard and Hughes (2015) (72)	Through searches of Google Scholar, Web of Science and Scopus; the aim was to define the characteristics of RUMs stored within DIRUM via a review of publications citing use of a RUM stored within DIRUM. Characteristics included the use of each RUM, administration modes, items of resource-use most commonly measure via patient self-report, variations in resource-use measurement method in relation to availability of alternative sources, agreement between methods when multiple methods are employed.
Ridyard and Hughes (2010) (9)	Through a review of studies published in the HTA journal, the authors identified and reviewed the stages of the resource-use data collection and costing for studies funded by the UK HTA programme, including methods for item identification, piloting and validation of RUMs.
Bhandari and Wagner (2006) (56)	Through searches of BIOSIS, the Cochrane Library, Current Contents, Medline, PsycINFO, and Web of Science, and reference searching, the authors identified published articles that reported on the accuracy of self-report via validation with data from an alternative data source. A summary of factors associated with recall accuracy was also reported including patient factors (sociodemographic characteristics, cognitive impairment, quantity of utilisation) and RUM characteristics (recall period, resource type, administration mode, memory aids, questionnaire design).

3.2.3 Objective of this review

The objective of the research described in this chapter was to review and provide a narrative synthesis of the methods taken in the development and psychometric assessment of existing RUMs that were designed for capturing and costing healthcare use.

3.3 Methods

3.3.1 Search strategy

The review followed the four phases outlined in the PRISMA flow diagram for systematic reviews: [1] identification, [2] screening, [3] eligibility and [4] inclusion (99). To identify RUMs with details on their development and assessment, a review of all RUMs stored within DIRUM, an electronic bibliographic database search and reference searching were conducted. I considered that these three search methods combined would provide a comprehensive approach to identifying all eligible RUMs for inclusion in this review. An outline of each method is provided below.

3.3.1.1 DIRUM

The primary source for the identification of RUMs with details on their development was DIRUM (62). DIRUM is a repository of instruments for resource-use measurement, where researchers performing trial-based economic evaluations upload their instrument in an open-access database (62). Within DIRUM, for each instrument uploaded the document provider can provide information on instrument characteristics, primary references, and instrument qualities (including references to work on development, validation and adaptations) (62). Primary references and instrument qualities were reviewed for information on development and assessment. The review of DIRUM was conducted in April 2018 and updated in August 2020 to identify any instruments that had been added since the initial review.

3.3.1.2 Database searches

It was acknowledged that not all RUMs with information on their development would be included in DIRUM. For systematic reviews of PROMs, it has been recommended that at a minimum Medline and Embase should be searched, in addition to other relevant databases (73). To supplement the DIRUM search, in April 2018 and updated in August 2020, a search of Medline (1946 to 10th August 2020), Embase

(1974 to 2020 Week 32) and PsychINFO (1806 to August Week 1 2020) accessed via OvidSP was undertaken with no restriction on publication date. As this review focused on literature on the development of RUMs, I did not consider the PICO (population, intervention, comparator, outcome) approach to developing a search strategy appropriate, as it is usually employed for systematic reviews of interventional studies (99). To identify literature relevant to the objective of this review, the search strategy (**A1.1**, **Appendix 1**) employed across the databases included combinations of related terms for each of the following concepts: development, completion, outcome and measure. I formulated each concept to include synonymous terminology, alternative spellings (e.g. 'resource utili#ation' to capture utilisation and utilization), and inflected forms (e.g. using truncation of terms, such as 'develop*', which captured 'develop', 'developed' and 'developing'). The strategy was reviewed by the research team, who suggested additional related terms, which were subsequently added. Related terms were combined using a Boolean 'OR' and a search of text words in the title and abstract was performed for each concept (120). To identify the final sample, concepts were combined using a Boolean 'AND' (120).

3.3.1.3 Reference searching

Given that many RUMs are revisions of existing RUMs (such as the CSRI, which has been adapted many times (74)), forwards and backwards reference searching were performed to identify further RUMs with details on their development in August 2020. Forward reference searching involved a search of the 'cited by' list of the primary references of included RUMs in Google Scholar. Backwards reference searching entailed reviews of the reference lists of the primary references of included RUMs.

3.3.2 Study selection

Duplicates identified across the three databases in the database search were initially dropped in OvidSP. Articles were then downloaded and managed in Endnote, where further duplicates were identified and dropped. I screened the title and abstract of all articles identified through the database searches. Articles were excluded if it was apparent from the title and abstract that the record did not meet the inclusion and/or met the exclusion criteria described below. Title and abstract screening was bypassed for articles that were identified in DIRUM. Next, I performed the full text

eligibility assessment on the references identified from the DIRUM and database searches. For articles identified via DIRUM, the DIRUM page for each RUM was also screened. Articles were excluded if they did not meet the eligibility criteria outlined below. Articles which were identified from reference searching of included articles that met the inclusion and exclusion criteria, were added at this stage.

Inclusion and exclusion criteria

Articles were included if they provided:

- development details on a RUM that was:
    - primarily developed for costing healthcare within economic evaluations
    - designed for retrospective data collection
    - designed for use by adults
    - primarily designed for self-completion or interviewer-administered self-report
- information on RUM development, including item development and/or pilot testing (outlined in detail below in section **3.3.3**)
- original research (i.e. not a review of development methods for existing RUMs).

Articles were not included if they were:

- not published in the English language
- not freely available via the University of Bristol library
- published only as a conference abstract
- reporting development related to ModRUM (i.e. the Delphi study (71)).

3.3.3 Data extraction

For all included articles, I extracted details on RUM characteristics, development steps and psychometric assessment. For the initial searches conducted in April 2018, a second reviewer also assessed the extracted information of included articles under each development step, to determine whether they agreed that the extracted data provided sufficient evidence for each step (JT). The article/RUM characteristics extracted included year of publication, country/setting, mode of administration, and what population the RUM was designed to capture resource use for or whether it was designed for use in any population. RUM development steps extracted included

item development and pilot testing. Five psychometric properties, that were considered important for RUMs, were extracted including validity (content, face, criterion, construct) and reliability. A complete list of psychometric properties of health measurement instruments, and their relevance to a RUM, is provided in **Table 2.1 (Chapter 2)**. As the aim of this thesis was to develop the healthcare modules of a new RUM, this review predominantly focused on the development and psychometric assessment of healthcare items. Some of the included RUMs also captured resource-use data from beyond the healthcare sector.

3.3.4 Data synthesis

Literature on included RUMs was synthesised and is presented narratively for each of the development steps and psychometric properties listed below. Detailed definitions of the psychometric properties are provided in **Table 2.1**.

1. Item development – including information on the method(s) employed to identify and select items for inclusion in the RUM. Some authors stated that their RUM was an adapted version of an existing RUM (most commonly the CSRI); this alone was not considered informative enough to include under item development.
2. Pilot testing – including an indication that the RUM was piloted, tested within a feasibility study, or a description of work conducted that could be considered as piloting (e.g. pre-testing (121)).
3. Psychometric properties:
    a. For content and face validity, evidence for the property was included if authors explicitly stated that the RUM had content or face validity.
    b. Information on criterion validity was included if there was evidence of validating the RUM with an alternative data source, regardless of whether the authors stated they were performing criterion validation. This may be categorised by authors as a type of construct (i.e. convergent) validity if the alternative was not considered a ‘gold standard’. For the purpose of this review, unlabelled validation with alternative sources is reported under criterion validation.
    c. Evidence for construct validity was included if authors referred to the testing or evidence of validity (construct, cross-cultural, convergent, discriminant, divergent, known group) or hypothesis testing.

d. Evidence for reliability was included if authors reported testing any form of reliability, such as test-retest reliability.

3.4 Results

3.4.1 RUM selection

A flow diagram of the search for articles is presented in **Figure 3.1**. Based on the combined results from searches conducted in 2018 and 2020, 91 RUMs were identified within DIRUM and after duplicates were removed, including 3 articles that were already included from the DIRUM search, 850 articles were identified through database searching. Title and abstract screening led to the exclusion of 774 articles. Full-text eligibility was conducted for 167 articles. A total of 141 articles were excluded during eligibility assessment, articles were most commonly excluded for lacking information on RUM development (n=74). 13 additional articles were included from reference searching. The final sample included 39 articles relating to 34 RUMs.

3.4.2 RUM characteristics

RUM characteristics; including year of publication, country/setting, intended population and administration mode are presented in **Table 3.2**. As many of the included RUMs were used in UK HTA studies, most RUMs were developed for a UK setting (19 RUMs). RUMs were also designed for eight other settings, including Germany, The Netherlands and Europe. Of the 34 RUMs included, 33 were designed for use in a specific population, such as individuals with mental health problems. One RUM, the ACQP, was designed to capture resource use for any population, by self-report, for UK-based studies (76). As opposed to a formulated RUM, the article reporting the ACQP includes a bank of questions from which users can select relevant questions and formulate a trial-specific RUM from them (76). In the 34 included RUMs, a wide range of populations were covered and included cancer, musculoskeletal conditions, food allergies and virtual consultations. Most RUMs were designed for participant self-completion (24 RUMs), some were designed for interviewer-administration (five RUMs) and for other RUMs both administration modes were offered (five RUMs).

Figure 3.1 Flow diagram of RUM identification, screening, eligibility and inclusion

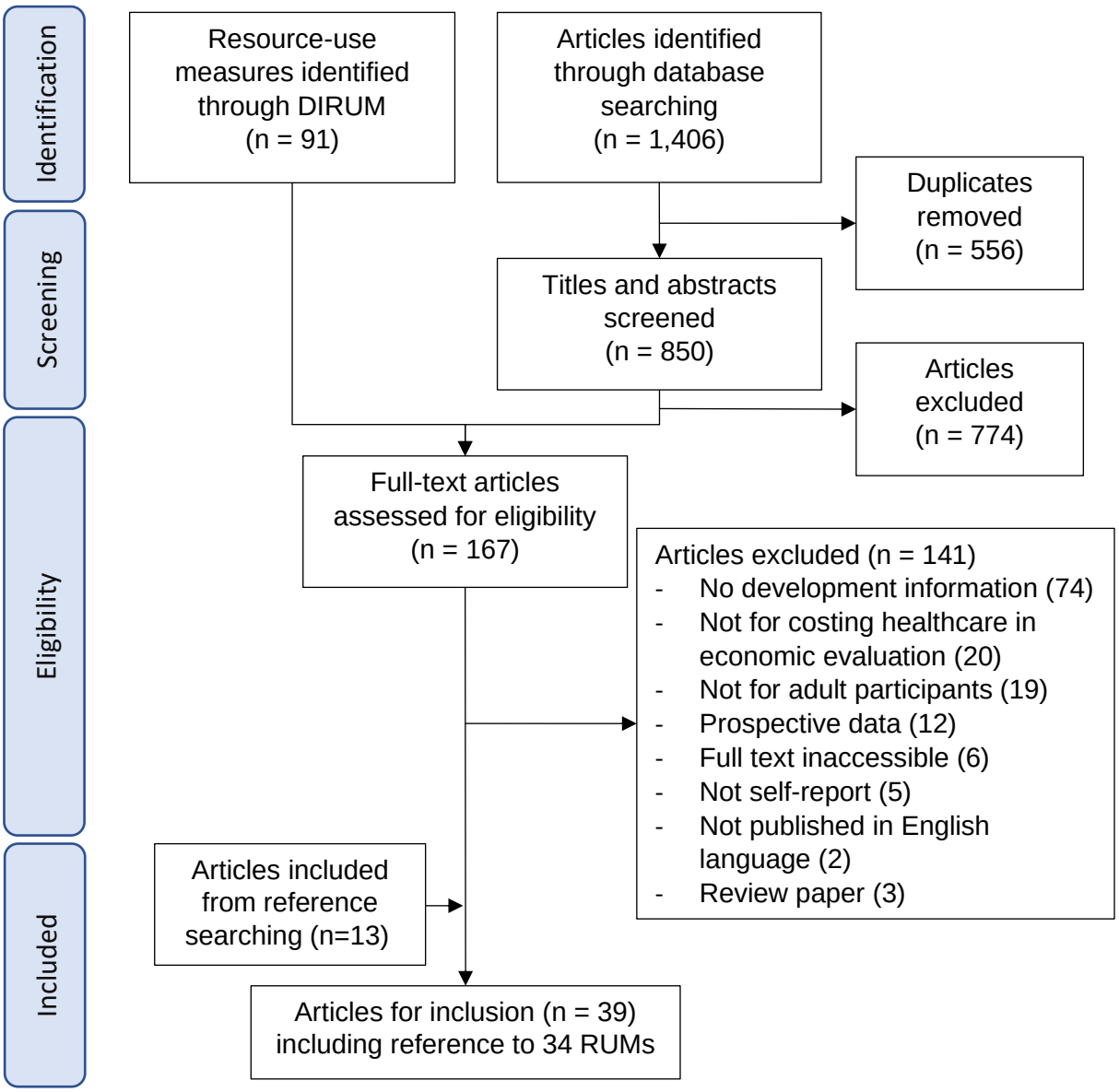


Table 3.2 RUM characteristics

Author(s)	Year of publication	Country/Setting	Intended population	Administration mode
Beecham and Knapp (74)	2001	United Kingdom	Mental health	Interview/self-complete
Beemster et al. (122)	2019	The Netherlands	Musculoskeletal Pain	Self-complete
Beresford et al. (123)	2019	United Kingdom	Reablement services	Interview/self-complete
Bouwmans et al. (124)	2013	The Netherlands	Mental health	Self-complete
Chernyak et al. (121), Icks et al. (125), Chernyak (126)	2012, 2017, 2011	Germany	Diabetes	Interview/self-complete
Chisholm et al. (75)	2000	European Union	Mental health	Interview
Chisholm et al. (127)	2004	Europe	Postnatal depression	Self-complete
Clarke et al. (128)	2016	United Kingdom	Parkinson's disease	Self-complete
Cooke et al. (129)	2009	United Kingdom	Ankle sprain	Self-complete
Cooper et al. (130)	2003	United Kingdom	Rheumatology	Self-complete
Fox et al. (131), Fox et al. (132)	2009, 2013	Europe	Food allergy	Self-complete
Gordon et al. (133)	2012	Australia	Colorectal cancer	Interview
Gray et al. (134)	2011	United Kingdom	Parkinson's disease	Self-complete
Griffin et al. (135)	2016	United Kingdom	Hip impingement	Self-complete
Guzman et al. (136)	1999	Canada	Low-back pain	Interview
Hirst et al. (137)	2008	United Kingdom	Uterine fibroids	Self-complete
Houchen-Wolloff et al. (138)	2018	United Kingdom	Cardiac rehabilitation	Self-complete
Mauch et al. (139)	2011	Kenya	Tuberculosis	Interview
Mirandola et al. (140)	1999	Italy	Mental health	Interview
Murray et al. (141)	2014	United Kingdom	Knee replacement	Self-complete
Ness et al. (142)	2020	Germany	Multiple Sclerosis	Self-complete

Table 3.2 continued

Author(s)	Year of publication	Country/Setting	Intended population	Administration mode
O'Dowd et al. (143)	2006	United Kingdom	Chronic fatigue syndrome	Self-complete
McIntosh et al. (144)	2016	United Kingdom	Parkinson's disease	Self-complete
Peek et al. (145)	2010	United Kingdom	Respiratory failure	Interview/self-complete
Pinto et al. (146)	2011	New Zealand	Osteoarthritis	Self-complete
Ruof et al. (147), Hülsemann et al. (148)	2004, 2006	Germany	Rheumatoid Arthritis	Self-complete
Russell et al. (149)	2019	United Kingdom	Depression	Self-complete
Sabes-Figuera et al. (150)	2012	United Kingdom	Mental health	Interview/self-complete
Schweikert et al. (151)	2008	Germany	Acute cardiac events	Self-complete
Thompson and Wordsworth (76)	2001	United Kingdom	General population	Self-complete
van den Brink et al. (152), van den Brink et al. (153)	2004, 2005	The Netherlands	Rectal cancer	Self-complete
Wallace et al. (154)	2004	United Kingdom	Virtual consultations	Self-complete
Williams et al. (155)	2016	United Kingdom	Ulcerative Colitis	Self-complete
Wyatt et al. (156)	2012	United Kingdom	Lysosomal storage disorders	Self-complete

3.4.3 RUM development and assessment steps

Evidence is summarised below under (1) item development, (2) pilot testing, and (3) psychometric assessment of existing RUMs. This information is also summarised in **Table 3.3** to indicate RUMs for which there was evidence for each development step and psychometric property.

3.4.3.1 Item development

Information on how items were developed was provided for 17 of the 34 RUMs. Most authors referred to selecting items with reference to existing RUMs and/or the literature (94%). Many used existing RUMs and/or the literature in combination with expert and/or patient input (71%), including Ness et al. and Ruof et al. (142, 147), while one reported that they had used expert input alone (150). Patient input was sought for eight RUMs (123, 130, 131, 136, 142, 145, 147, 151), clinician input was sought for eight RUMs (126, 136, 141, 142, 145, 146, 150, 151) and health economist input was sought for one RUM (146).

Five authors collated items from existing RUMs to form a list of potentially relevant items for a new RUM (71, 76, 121, 136, 147). For each item, Ruof et al. also extracted information on the level of detail required, length, wording and psychometrics (147). For some RUMs, existing RUMs were revised to make them applicable for the condition of interest (122, 131) and relevant parts of existing RUMs were incorporated into new RUMs (74). Literature reviews were used to inform main cost drivers (141) and cost-generating categories (130). In the development of one RUM, resource-use data were obtained from multiple sources including guidelines, studies and administrative data (142).

For the eight RUMs soliciting clinician feedback, a description of the approach taken was often limited, however, interviews were reported for two RUMs (150, 151) and focus groups for one RUM (150). Some authors reported gaining input from a range of clinical professionals (136, 142). For example, in the development of a RUM designed for capturing resource-use data from patients with low-back pain, the draft was sent to a rheumatologist, an occupational physician, a family physician, and a chiropractor for comment (136). For some RUMs, specialist clinical feedback was

sought for condition-specific RUMs to gain expert knowledge on the resources likely to be used by a specific population (121, 150, 151). The impact of clinician feedback on item formulation was rarely reported; however, one author reported revising missing, ambiguous, or redundant items or concepts (136).

Methods to elicit patient feedback included informal feedback via patient and public involvement groups (123) or acquaintances with patient experience (136), to more formal qualitative research including focus groups (130, 131) and interviews (147, 151). While some authors reported involvement of patients, the extent to which they were involved or in what capacity was sometimes unclear (142, 145). Several authors reported developing topic guides and/or questionnaires based on existing literature and using them as a foundation for discussions with patients (130, 131, 147). Patients provided feedback on the relevance (131), comprehensiveness (136) and appropriateness of items (147), which led to changes to the RUMs including the addition of further questions and varying levels of question aggregation for controversial areas (147).

RUM design was also informed by the expected cost of the items with, for example, more detail included for high cost items (74). The salience and frequency of items also impacted RUM design, with one study using patient interviews and plausibility to determine salience and frequency of items, which then informed recall periods and question order, with more memorable items given longer recall periods and questions grouped by recall period in the questionnaire (151).

Table 3.3 RUM development steps and psychometric properties

Authors	RUM development		Psychometric assessment				
	Item development	Piloting testing	Content validity	Face validity	Criterion validity	Construct validity	Reliability
Beecham and Knapp (74)	Y	Y	Y	Y	Y	-	-
Beemster et al. (122)	Y	Y	-	-	-	-	Y
Beresford et al. (123)	Y	Y	-	-	-	-	-
Bouwmans et al. (124)	-	Y	-	-	Y[1]	(Y)[1]	Y
Chernyak et al. (121), Icks et al. (125), Chernyak (126)	Y	Y	-	-	Y	-	-
Chisholm et al. (75)	-	Y	-	Y	-	Y	-
Chisholm et al. (127)	-	Y	-	-	-	-	-
Clarke et al. (128)	-	Y	-	-	-	-	-
Cooke et al. (129)	-	Y	-	-	-	-	-
Cooper et al. (130)	Y	Y	-	-	Y	-	-
Fox et al. (131), Fox et al. (132)	Y	Y	-	-	-	Y	-
Gordon et al. (133)	-	Y	-	-	Y	-	-
Gray et al. (134)	-	Y	Y	-	-	-	-
Griffin et al. (135)	-	Y	-	-	Y	-	-
Guzman et al. (136)	Y	Y	Y	Y	Y	-	-
Hirst et al. (137)	Y	Y	Y	-	-	-	-
Houchen-Wolloff et al. (138)	-	Y	-	-	-	-	-
Mauch et al. (139)	Y	Y	-	-	-	-	-
Mirandola et al. (140)	-	Y	-	-	Y	-	-
Murray et al. (141)	Y	-	-	-	-	-	-

Table 3.3 continued

Authors	RUM development		Psychometric assessment				
	Item development	Piloting testing	Content validity	Face validity	Criterion validity	Construct validity	Reliability
Ness et al. (142)	Y	-	-	Y	-	Y	Y
O'Dowd et al. (143)	-	Y	-	-	-	-	-
McIntosh et al. (144)	-	Y	-	Y	-	-	-
Peek et al. (145)	Y	Y	-	-	-	-	-
Pinto et al. (146)	Y	Y	-	-	Y	-	-
Ruof et al. (147), Hülsemann et al. (148)	Y	-	-	-	Y	-	-
Russell et al. (149)	-	Y	-	-	Y	-	-
Sabes-Figuera et al. (150)	Y	Y	-	-	Y	-	-
Schweikert et al. (151)	Y	-	-	Y	Y	Y	-
Thompson and Wordsworth (76)	Y	-	Y	-	-	-	-
van den Brink et al. (152), van den Brink et al. (153)	-	Y	-	-	Y[1]	Y	-
Wallace et al. (154)	-	Y	-	-	-	Y	-
Williams et al. (155)	-	Y	-	-	-	-	-
Wyatt et al. (156)	-	Y	-	-	-	Y	-

[1]Comparison of RUM data with administrative data was labelled as construct/convergent validity by some authors, to aid comparison across RUMs, I have included it under criterion validity, not construct validity, for this review.

3.4.3.2 Pilot testing

Some evidence of pilot or feasibility testing was found for 29 RUMs; however, the amount of detail provided varied, with some authors providing minimal detail. For 12 RUMs, authors did not elaborate beyond stating that the RUM was piloted, piloted with patients, or piloted in another trial. Settings included pilot studies (136), feasibility studies (135, 138, 149, 155) and within trial pilots (128). Methods to elicit feedback from patients included interviews (124, 150), more specifically cognitive interviewing (with concurrent and follow-up verbal probing) (121) and structured interviewer administered questions (136); qualitative research (135) and postal debriefing survey questions (130). Others did not elicit patient feedback directly but used their patient pilot to review responses (149) and revise questions (155).

Interviews to obtain feedback on the RUMs were conducted with between 5 (140) and 43 patients/service-users (121). In the development of one RUM, self-complete and interviewer-administered versions of the RUM, designed to capture resource-use from patients with diabetes, were tested with 43 patients in cognitive interviews with behaviour coding to identify issues (121). Scripted probes explored comprehension, information retrieval and answer confidence ratings, while unscripted probes allowed exploration of insufficient responses (121). For another interviewer-administered RUM, piloting was conducted with 80 patients and interviewers followed RUM completion with eight questions covering difficulties experienced, acceptability and length (136). The interviewer also scored comprehension and ability to recall on a five-point Likert scale (136). For a patient-completed RUM sent by post, the survey procedure was piloted by 24 patients who were subsequently sent debriefing questions where feedback on the RUM could be reported (130).

Details on the purpose of piloting were provided for 12 RUMs. Piloting was used to identify difficulties with RUM completion (75, 123, 150) and included the identification and assessment of issues with comprehension and interpretation (75, 121, 130, 131, 136). Others explored ease of completion (130, 135) and ease of information retrieval (121, 136). Patient acceptability was assessed for four RUMs (130, 135, 136, 150) and included assessment of the length of the RUM (130, 136). Patient burden was considered by asking patients to report the length of time it took them to

complete the RUM (136). Piloting was also used to assess and refine the content of RUMs. Content assessment included consideration of item relevance (136, 146), suitability of questions and mode of administration (130) and clarity of questions (146). Beresford et al. used piloting to identify the optimal way to frame questions which minimised burden, maximised completion and avoided double-counting (123). Content refinement included identification of redundant (131, 149) and additional questions (149). Data completeness, response rates and/or missingness were often estimated during piloting and used as an indicator of RUM acceptability (123, 130, 135, 146). Piloting was also used to construct a list of commonly used medications (155) and assess RUM generalisability across geographical settings (130).

Adaptations following piloting were reported for 10 RUMs; however, detail on the adaptations was often omitted or limited (121, 123, 131, 136, 150, 156). Piloting resulted in the addition and/or omission of some questions (136, 156). Adaptations included shortening and simplifying questions (such as adding skip logic) (131, 150), adding instructions (121) and reformatting questions (121). Beresford et al. provided detailed information on the changes made, including changes to avoid double counting (e.g. removal of consultant appointments in GP practices from the outpatient question), increased clarity to minimise ambiguity (e.g. clarification that face-to-face and telephone appointments should be reported) and extension of the recall period (123).

3.4.3.3 Psychometric assessment

Within the included articles describing RUMs, evidence was provided for the assessment of content, face, criterion and construct validity, and reliability.

Content and face validity

In this section and under content validity in **Table 3.3** the summary of content validation is limited to only those who explicitly report it. Evidence for, or assessment of, content validity was reported for five RUMs. Most commonly, evidence of content validity consisted of statements that the RUM was developed with expert input or based on existing evidence including relevant RUMs or literature (76, 136, 137). Although not specified as content validation, some of the methods employed in the

development, and in particular item development, of other included RUMs suggest that they may have content validity.

Evidence for, or assessment of, face validity was reported for six RUMs, with information on how it was assessed provided for five RUMs. For two RUMs, face validation involved a judgement on the completeness and consistency of the RUM by clinicians (151) and in expert consensus meetings including clinicians and health economists (142). For one RUM, judgement on the face and content validity was made by a range of clinicians and resulted in revisions to items and concepts that were judged missing, ambiguous, or redundant (136). Involvement of patients in face and content validation was reported for one RUM and involved focus groups including clinicians, social care workers, informal carers and service users (75). In the focus groups, participants considered the content included and language used in the RUM (75).

Criterion validity

Validation of the RUM or a subset of questions from the RUM with an alternative source was reported for 14 RUMs. Bouwmans et al. and van den Brink et al. labelled validation with an alternative source as an assessment of construct or convergent validity; however, given its similarity to what others describe as criterion validity it was included here to aid comparison (124, 152). Validation was described with varying levels of detail. For some it was unclear which items were validated, while for others they reported that it was a subset of the RUM (124, 136, 149, 151). Alternative sources included payer records (147), insurance data (125, 151), healthcare provider (e.g. GP/hospital) records (130, 135, 146, 152) and 'administrative' data (140). Some reported retrieving data from multiple sources (146, 150). The number of participants ranged from 10 (135) to 432 (125).

Multiple statistics were often estimated to compare RUM results with alternative sources. Four studies presented Kappa statistics for categorical variables (e.g. visited/not visited a GP) (130, 133, 136, 147). Pinto et al. estimated sensitivity (the proportion indicating resource use when resource use is observed in the records) and specificity (the proportion indicating no resource use when no resource use is observed in the records) (146). Four studies reported use of Bland-Altman plots to

visualise the difference between data reported in alternative sources (125, 146, 147, 151). For the assessment of continuous variables (e.g. number of GP visits) the following statistics were used: t-tests (150, 151), concordance correlation coefficients (150), intraclass correlation coefficients (ICCs) (133, 151), Lin's concordance correlation coefficients (140, 146) and Spearman's rank correlation coefficient (124, 147). Concordance was always estimated at the resource use level, and sometimes also estimated the cost level (140, 146, 150). Comparisons were also presented based on raw data including absolute and percentage agreement (124). Information was generally lacking on whether RUMs were refined following validation.

Reporting of agreement between self-report and other records was variable. Several studies found that there was good agreement (124, 135, 140, 151), while one reported that agreement was poor, with agreement rates no higher than 61 percent (149). Some found that agreement varied by resource (130, 136, 147). For outpatients, agreement was reported as limited (147) and moderate to good (130), yet for inpatients substantial (147) and perfect (130) agreement were reported. For medications, discrepancies were found in one study, with 15% of participants reporting medications not included in physician records and 8% of participants not reporting medications that were included in the records (136). Some found that resource-use was lower in self-report than other sources, including insurance data (125) and administrative data from the statutory sickness funds (151), while the opposite was found in another study, where hospital case notes were used as an alternative source (150).

Icks et al. reported that no healthcare use (specificity) was reported more accurately than use of resources (sensitivity) (125). They also explored how agreement between RUM responses and health insurance data differed between a three- and six-month recall period version of their RUM. Results suggested there was under-reporting in both versions; however, the difference between data sources was larger in the six-month version. The authors highlighted that underreporting and overreporting should be estimated and assessed independently, as comparisons of average resource-use do not reveal issues if over- and under- reporters cancel each other out in the average. When the authors accounted for this, they reported that

accuracy of reporting was similar at the individual level and they could not conclude whether a shorter recall period leads to more accurate results.

Construct validity

In the development of one RUM, convergent and discriminant validity were assessed by estimating correlations with associated patient-reported outcomes and known-group analyses (142). For other RUMs, convergent validity was assessed via comparison with prospective cost diaries (151, 153). For one RUM, participants completed both the RUM and diary and higher mean values were estimated for most items in the prospective diary data, with ICCs averaging at 0.72 (151). For the second RUM, to avoid bias when completing the RUM due to completing the diary, separate samples completed the RUM and the diary (153). Repeated measures analyses found no significant difference between RUM and diary for most resource-use items, with the exception of several items including number of hospital days, where questionnaire respondents reported significantly more days, and number of other health-care worker contacts, where diary respondents reported significantly higher utilisation (153). Cross-cultural validation was performed for two RUMs (75, 132). For one the clarity, comprehensiveness and relevance of the questions in translated versions were analysed using item completion and questionnaire response rates (132). For the other, focus groups were conducted with healthcare professionals, service users and carers, and the content and language of the RUM was considered (75).

Reliability

Test-retest reliability analyses were conducted in the development of three RUMs (122, 124, 142). The time between test and retest RUM completion was reported for two RUMs and was on average 17 (124) and 20 (122) days. While retest responses were excluded for reliability testing of one RUM if they were completed more than one month following initial RUM completion, as the gap was judged too long (124), for another RUM reliability testing was conducted over a 3-month retest period, with a subset of patients who were stable and not expected to have varying resource-use over the period (142). Binary responses were assessed with percentages for absolute agreements and Cohen’s kappa coefficients for chance correlated

agreements (122, 124). ICCs were estimated to indicate reliability of total costs (142), total healthcare use (122) and individual resource-use items (124).

Testing of 1192 stable patients showed good reliability (ICC: 0.83) for total costs (142). For total healthcare use among 52 stable respondents an ICC of 0.81 was estimated (122), while for another RUM, tested with 99 respondents, estimated ICCs for individual resource-use items were generally considered good with the exception of some items for which an ICC could not be estimated due to low usage (e.g. hospital admission duration) (124). Kappa values were mostly considered moderate to satisfactory (0.49-0.84) for one RUM (124), while for another they ranged from fair to perfect (0.11 to 1); perfect agreement was only reported for 'dietician' and this service was used by few participants (2%) (122).

3.5 Discussion

3.5.1 Main findings

The objective of this review was to establish what methods have been employed in the development and psychometric assessment of existing RUMs, to inform the development of ModRUM. 34 RUMs were identified and information on their development and psychometric assessment was synthesised. RUMs were most often designed for capturing resource-use data for a specific group of patients (e.g. mental health, cancer), for use in a UK setting and for self-report administration only.

Labelling of development and assessment was not consistent, which meant assumptions, defined in **3.3.4**, were required to categorise development and assessment. A range of methods were employed in the development of items, including adapting existing RUMs, reviews of literature from a variety of sources, and expert and patient interviews and focus groups. Piloting methods included testing within feasibility studies, within-trial pilots, interviews, and postal and interviewer-administered RUM completion. The purposes for piloting included: the identification of difficulties, assessment of the ease of completion and information retrieval, acceptability, content assessment and refinement, and estimation of response rates and missingness. Despite some information being provided on item development and piloting, detailed descriptions were rarely provided and information on subsequent adaptations to RUMs was mostly limited or omitted.

Criterion validity was the most commonly assessed psychometric property, with most authors estimating the agreement between some RUM data and data reported in electronic medical records, payer records or insurance data. There was disparity in criterion validity findings, with some finding higher resource use reported in administrative data, while others found the opposite. Adaptations subsequent to criterion validation were not reported. For the remaining psychometric properties; including content, face and construct validity and reliability; there was evidence for the assessment of each property for less than a quarter of RUMs.

3.5.2 Strengths and weaknesses of the review

In this review I searched multiple sources to maximise the chance of identifying all relevant RUMs. The review demonstrated that a wide range of methods have been used in the development and psychometric assessment of existing RUMs. The review also demonstrated that no RUM currently exists that is designed for use in a wide range of trials, and that has gone through a thorough development and validation process. For researchers developing new RUMs, the review has summarised the processes that have been taken and were feasible in the development and assessment of existing RUMs.

In line with other reviews, the primary source of articles related to RUM development was DIRUM (72, 118). Supplementary searches of bibliographical databases were conducted with the aim of identifying RUMs not stored in DIRUM; however, there is the potential that some relevant articles were missed. Owing to the lack of standardised terminology for RUMs, I developed a search strategy. In developing the search, I considered the relative value of expanding the search to ensure comprehensiveness against precision of the search, which was expected to be low due to the variety of terms used to describe development and validation, and the lack of specific search terms. For example, ‘design’ was initially included as a search term, but it resulted in a considerable number of articles, so to increase precision ‘questionnaire design’ and ‘design a questionnaire’. A limitation of this review was that further expansion of the search may have identified additional relevant articles. For example, other terms for ‘questionnaire’ could have been included alongside ‘design’ in the previous example, and alternative terms for ‘questionnaire’ could have

been included, such as inventory or schedule, with the former likely to have identified adapted versions of the CSRI (74). A further limitation of this review was that I solely undertook all screening, eligibility assessment and extraction of data. Despite a second reviewer reviewing some extracted information, having a second reviewer at all stages of the identification, screening, eligibility and inclusion process, may have reduced data errors and increased confidence that all relevant articles were identified and included, with all relevant information extracted and categorised under development steps correctly.

In addition, due to the vast number of references, full details of screening were not recorded during reference searching. For example, at the time of searching, there was over 1000 forward citations for the publication reporting the development of the CSRI (74). This may mean that articles reporting RUM development were missed; however, as this was a supplementary search, I believe detailed reporting was not warranted in comparison to the number of articles that would be identified through this approach, and in the context of a considerable number of articles that were reviewed more stringently from DIRUM and database searches. Over half of the RUMs included were designed for UK-based studies. This may be due to the search strategy and inclusion criteria, such as DIRUM being the primary source of RUMs which is a repository developed by UK-based researchers, and only including articles published in English language. It could also be due to less usage of self-report RUMs in other healthcare systems or the use of alternative terminology which was not included in the database search strategy.

RUMs were only included if they were designed for self-report via self-completion or interview administration. This decision was intentional as RUMs designed for proxy completion are conceptually different, and the aim of this thesis is to develop a RUM for self-completion. However, had the search been widened and included RUMs designed for proxy completion, other RUMs may have been included. For example, in the refinement of an adapted version of the CSRI, designed for proxy completion by bereaved relatives, ‘think-aloud’ interviews were conducted with concurrent and retrospective verbal probing to identify comprehension issues (157). A three-step development and testing process was conducted in the development of another RUM, which was designed for caregiver completion to capture use of rehabilitation

care of people with disorders of consciousness in long-term care in Germany (158). To develop the first version of the RUM, reviews of existing guidelines and existing RUMs, and a Delphi survey with healthcare professionals were conducted (158). In the second stage, semi-structured telephone interviews with three expert physicians led to RUM refinements (158). The final stage involved cognitive interviews with four caregivers with techniques including think-aloud, probing, observation of participant behaviour and participant-reported confidence ratings in their answers (158).

The review was also limited to the primary references for each RUM, as the main objective was to summarise the original development of existing RUMs, with the aim that it could potentially inform the development of ModRUM. Although for most RUMs this will encompass all the information relevant to their development and assessment, for the minority there will be other published articles, potentially by research groups independent of the RUM developers, that provide further evidence on the psychometric properties of the RUM. Most notably, the CSRI was found to have been cited 143 times in a review reported in 2015 (72). However, as many of these studies reported using adapted versions of the CSRI, the applicability of any psychometric evidence for the original CSRI could be questioned (72). To also meet the aim of this review, RUMs were only included if information on item development or piloting was provided. As such, despite information being provided on the psychometric assessment of some RUMs, they were omitted as no development information was provided.

The analysis was limited to a narrative synthesis, with no attempt made to score the quality of the development and assessment of each RUM. For systematic reviews of PROMs, the COSMIN Risk of Bias checklist has been developed for quality assessment; however, this could not simply be applied to RUMs as it includes properties that are not relevant to RUMs, including structural validity (relevant to multi-item PROMs, concerning the dimensionality of the construct) and internal consistency (interrelatedness of items) (69, 159). Furthermore, for PROM development content validity is deemed to be the most important psychometric property (73). For systematic reviews of PROMs, reviewers are guided to not consider other measurement properties of a PROM if there is strong evidence that the content validity of a PROM is insufficient (73). In this review, it has been

demonstrated that RUM developers rarely report on the content validity of their RUM while criterion validation is much more regularly assessed. This further emphasises that PROM guidelines may not be applicable for the assessment of RUMs.

Despite quality scores not being estimated, the extent of development information was judged at the eligibility stage of identifying RUMs, with RUMs being omitted if there was insufficient information on item development or piloting. For item development, this required a more subjective judgement than for piloting. For example, RUMs were not included if item development information was limited to stating that they were developed with reference to another instrument, without further elaboration. RUMs were also excluded if information was limited to item formulation (i.e. how they were structured/formatted), without reference to how items were selected.

3.5.3 Comparison to existing literature

While there is some overlap in the RUMs included in this review and other reviews, as the most recent review to have been undertaken was by Mayer et al. in 2017 and each review has a unique set of inclusion criteria and objectives, some of the RUMs and their development and assessment included in this review have not been previously been reviewed (118). This review included any RUM which included items on healthcare resource-use, Mayer et al. included only RUMs which incorporated items on education and criminal justice sectors (118), Ridyard and Hughes (2010) included only studies funded by the UK HTA (9) and Ridyard and Hughes (2015) only included RUMs stored within DIRUM (72).

In this review, multiple sources were searched in the identification of relevant RUMs. In addition to Medline, Embase and PsycINFO, Leggett et al. also searched the Health and Psychosocial Instruments Database and Cumulative Index to Nursing and Allied Health Literature (CINAHL) (51), while Noben et al. limited their search to NHS Economic Evaluation Database and Medline with DIRUM used to test the sensitivity (50). Mayer et al. took a similar approach to this study, searching DIRUM and reference searching; however, their search of seven databases, also included searches of Social Science Citation Index, Econlit, Education Resources Information Centre and CINAHL (118).

The objective of Ridyard and Hughes' (2010) review was broad as they considered all resource-use data collection and costing methods, with the aim of generating data collection and costing guidelines for future studies (9). As such, their review was not limited to retrospective RUMs designed for self-report; they included all HTA studies that conducted economic evaluations (95 studies), of which 63 studies involved patient-completion, proxy-completion or interviewer-administered RUMs (9). As a result, limited detail was summarised on the development of existing RUMs and consideration of psychometric properties was limited to criterion validity. They did however create guidelines for item development, piloting and validation (9). For the latter, they recommend validation with an alternative source of resource-use data (9). In developing items, they reported that health-care professionals should be consulted and evidence from the literature and pilot studies should be considered to identify the main cost drivers (9). They suggested pilot testing should also be conducted to test clarity, ease of completion and completion rates (9).

Evidence of validation with another source was the primary objective of the reviews by Leggett et al. and Noben et al. (50, 51). Leggett et al. reviewed 15 studies and concluded that validated RUMs are available, but it is usually a subset of the RUM that is validated and resource-use data are more accurately reported for some items than others (51). In contrast to our study, Noben et al. assessed the quality of included studies using the Methodological Reporting Quality (MeRQ) tool, which they created using existing tools designed for other purposes (50). The tool considers three parts which include: [1] clearly specified aims, methods, rationale, procedures, population and recall period, [2] assessment of quality of evidence including statistical (uncertainty) methods, validation estimates and interpretation, and [3] additional information (50). Based on evidence comparing self-report and administrative data of six studies considered to be of adequate quality the authors suggested that neither self-report or administrative data is optimal and researchers should be cautious when considering them as exchangeable (50).

The review by Ridyard and Hughes in 2015 appeared to focus more on the implementation of RUMs rather than their development and assessment (72). They included 146 records relating to 25 RUMs listed in DIRUM. However, as 143 records

cited the CSRI, the results in this review predominantly related to how the CSRI had been administered in other studies (72).

Similar to our review, Mayer et al. reviewed psychometric evidence beyond criterion validation (118). To identify the evidence, for RUMs identified in DIRUM they reviewed information contained in DIRUM under 'instrument qualities' and for RUMs identified using alternative methods they conducted secondary database searches (118). They found evidence for the psychometric assessment (including test-retest reliability, concurrent validity, construct validity) of seven RUMs and piloting of nine RUMs; however, within the scope of a journal article detail on the methods employed was limited (118). As a result, I believe the description of the psychometric assessment of existing RUMs within this review extends beyond any current evidence in this area.

3.5.4 Implications for research practice

This review has demonstrated that reporting of the development and psychometric assessment of RUMs has been limited to date. Where it has been performed there is a lack of clarity in reporting. When selecting a RUM, health economists should consider what evidence is available on the development or assessment of the RUM, and the implications of a lack of evidence on the quality of data retrieved. If health economists decide to develop a new RUM, psychometric assessment should be considered and the development and assessment should be clearly reported to increase transparency. For decision makers, who are interpreting evidence from economic evaluations, consideration should be given to information provided on RUM development and assessment, and how this information, or lack thereof, may impact the quality of the results of obtained.

3.5.5 Implications for this research

The lack of a precedent for developing and assessing RUMs meant that the methods employed to develop ModRUM were informed by literature from a range a fields (**Chapter 2**). It was evident from this review that in the RUMs reviewed, criterion validity was the most commonly assessed psychometric property, with an established method of comparing RUM data with healthcare provider/administrative data. This may be an indication that criterion validity is considered the most

important property to assess among RUM developers or it could be due to more pragmatic reasons. For example, criterion validation is performed post RUM implementation, whereas content and face validity are assessed prior to implementation, where time is often limited prior to starting a trial. As this research was not limited to the time constraints of a trial, which was the case for the majority of RUMs included in this review, in this thesis, the development of ModRUM went beyond criterion validity, to assess several psychometric properties.

3.5.6 Unanswered questions and future research

This review is purely descriptive; however, information on item development and pilot testing was judged at the eligibility stage to decide whether RUMs should be included. Checklists have been developed in other fields to assess the quality of information reported on development and psychometric properties (e.g. the COSMIN Risk of Bias checklist for PROMs) (159). For RUMs, a criteria has been developed for judging the methodological reporting quality of studies reporting on the exchangeability of self-report and administrative healthcare resource-use data (50). Future research could involve the development of a checklist for assessing the quality of the entire development and psychometric assessment process for RUMs. Such a checklist would help researchers to identify what steps and methods are necessary during the development process and it could also increase consistency and transparency in reporting of development and assessment.

3.6 Conclusion

The objective of this review was to establish what methods have been employed in the development and psychometric assessment of existing RUMs to inform the development of a new standardised RUM for use across different patient populations. From the plethora of RUMs that have been developed, through systematically searching multiple data sources, I identified only 34 RUMs with articles reporting details on their development, which were often fairly limited. A range of methods were used in the development of items and piloting. Psychometric assessment of RUMs was generally limited. Detail on RUM refinements implemented following piloting and psychometric assessment was largely limited or omitted, indicating a need for increased transparency in the development of RUMs.

This chapter confirmed that there is currently no generic RUM available, that has been designed for capturing healthcare utilisation data in a wide range of trials, and for which a thorough development and validation process has been reported. The next chapter describes how items identified in the Delphi consensus survey were formulated into a RUM (71).

Chapter 4 Designing ModRUM

4.1 Chapter overview

This chapter describes how the first prototype of ModRUM was constructed. In the background section, I set out the rationale and objectives of this chapter. The methods section outlines the activities that were undertaken to design ModRUM. The results section describes how questions were formulated from items and how decisions were made about other aspects of ModRUM including the instructions, layout and formatting. It also describes input from PROM developers and a design agency. In the discussion, I provide a chapter summary, strengths and weaknesses of this chapter, a comparison to existing literature and implications of this chapter for the research conducted within this thesis. A concluding section completes the chapter.

4.2 Background

There does not exist a RUM, that is generic and designed for self-completion, that has been well-utilised in UK-based trials. The Delphi consensus survey which preceded this thesis suggested there is an appetite for such a measure, in particular a concise generic RUM with a modular format (71). It is recommended that prior to developing a new measure, a questionnaire developer should consider whether an existing instrument can be used or modified (79). However, as evident in **Chapter 3**, most existing RUMs are not generic, but are designed for a specific patient group. In the Delphi study, the first stage of development of a new generic RUM was conducted (71). Ten items were identified that could form the core healthcare module of the new modular RUM (71). While the Delphi study informed which items should be included in the core module, recommendations for the wording, formatting and layout were not made (71).

It has been demonstrated that there is an overlap of items in existing RUMs (115). When developing a RUM, existing RUMs can indicate the range of terms used for specific items, and the types of formatting used for such items. Questionnaire design literature should also be followed as it provides guidance on how to produce good quality questions (85, 101). For example, questions should be clear and

unambiguous, while avoiding overly lengthy questions (101). The importance of clearly written instructions should also not be overlooked (85). Instructions orientate patients to the topic under consideration and the period of interest, which are crucial to avoid confusion and incorrect responses (85). The main topics covered in the instructions of existing RUMs can be utilised to inform a checklist of topics to cover in the instructions of a new RUM. Presentation of response options also requires consideration at the development stage, to ensure respondents understand how to record responses (85).

The aim of the research described in this chapter was to develop a prototype of ModRUM from items identified in a Delphi study (71).

4.3 Methods

This chapter describes how information was extracted from RUMs stored within DIRUM to inform the wording, formatting and layout of ModRUM (62). It also reports on the roles of experienced PROM developers and a professional design company in the development of ModRUM.

4.3.1 Extraction from RUMs stored in DIRUM

As described in **Chapter 3**, DIRUM is an open-access database of RUMs where researchers can submit for inclusion the RUMs they have developed (62). In the Delphi survey, a long list of items for potential inclusion in the new RUM were extracted from RUMs stored within DIRUM (71). For this reason, and the ease of access to RUMs, I utilised RUMs stored within DIRUM to inform the terminology, formatting and layout of ModRUM.

All questions that related to one of the ten core items identified in the Delphi study, presented in **Table 4.1**, were extracted from RUMs stored within DIRUM that were designed for use with adult participants. For each RUM, details on each item were extracted and recorded in Microsoft Excel spreadsheets. Details extracted related to question wording (e.g. GP description), question layout (e.g. use of skip logic, defined in section **1.4.2.2**) and formatting (e.g. use of bolding or italics). Once I had extracted all relevant information, the number and percentage of questions using each detail were estimated and also recorded on the spreadsheets.

Table 4.1 Items identified for inclusion in a new RUM

Item
1. Number of hospital admissions (inpatient stay or day case)
2. Length of stay (e.g. dates or number of nights)
3. Number of hospital outpatient appointments
4. Number of visits to Accident and Emergency
5. Number of admissions to hospital, after Accident and Emergency
6. Number of appointments at a GP surgery or health clinic or other community setting
7. Type of professional seen at a GP surgery or health clinic or other community setting
8. Number of health care professional visits at home
9. Type of health care professional seen at home
10. Name/class of medication

4.3.2 Designing a prototype of ModRUM

Spreadsheets were sent via email to other members of the research team, who between them have extensive experience of designing RUMs for trials. The research team were asked to independently report their preferences for ModRUM question design. While considering the options, the team were asked to take into consideration the design principles that the team had previously agreed upon for ModRUM (**Table 1.1**, **Chapter 1**). For each detail, a preferred option or options could be selected, or team members could state they had no preference. The research team could also include comments for each detail, such as whether they thought the decision should be informed by patients. Once complete, I collated the research team's preferences in one spreadsheet. At team meetings, I presented the combined spreadsheet and used it to lead a discussion regarding preferences for the prototype of ModRUM.

Following team meetings, I created first drafts of the questions based on the results of the Delphi study (71), the exercise described in the previous paragraph, discussions at team meetings, the design principles of ModRUM (**Table 1.1**,

Chapter 1) and questionnaire design literature (79, 85, 101). To abide by questionnaire design principles I aimed to avoid ambiguous questions and using jargon (79). I also aimed to keep questions concise, without losing clarity (79). An iterative approach was followed, where I presented drafts to the team, the team provided feedback and I revised ModRUM.

4.3.3 Input from experienced PROM developers

Once the research team and I were satisfied with the first prototype of ModRUM, I sought informal feedback from independent researchers, based at the University of Bristol, who were experienced in developing PROMs. Researchers were invited via email. Those that agreed were sent a brief summary describing ModRUM and the first prototypes of ModRUM core module and ModRUM core and depth module. While reviewing ModRUM, I asked researchers to focus on the questions below.

- Do you think any of the questions or terms will be problematic for patients, and if so, why?
- Do you have any feedback on the order of questions?
- Do you think the instructions are clear and is there anything missing from the instructions?
- Do you have any comments and/or suggestions on the formatting and layout of ModRUM?

I arranged to meet each researcher, individually, on one occasion, where they were able to provide feedback verbally. They were also invited to send any further feedback via email following the meeting. Following meetings with each expert, I summarised the feedback they provided and adapted the prototype to visualise the suggested changes. I presented the feedback and revised questions to the research team, and the prototype was revised until the team agreed that the prototype was ready to present to health economics experts (**Chapter 5**).

4.3.4 Input from a professional design company

Funding was obtained by one of my PhD supervisors (JT) for an external company (Dirty Design, a Bristol-based design agency) to design a logo for and enhance the aesthetics of ModRUM. Along with two members of the research team (JT and SN), I met with the design company to describe what input was required. The design agency did not input to the content of ModRUM, their role was to enhance the design

of ModRUM, so that it was more visually pleasing and cohesive with the logo they designed. An iterative approach followed with the design company, until the research team and I were satisfied with the design of ModRUM and the logo. Following design company input, the design of ModRUM continued to be revised based on participant feedback in **Chapter 6**, **Chapter 7** and **Chapter 8**.

4.4 Results

4.4.1 Review of RUMs and DIRUM prototype design

In July 2018, there were 54 RUMs stored in DIRUM that were designed for use with adult participants. Question details were extracted from questions for each core item identified in the Delphi study (71). For each RUM, not all details were extracted from all items, as some details were consistently used throughout the RUM (e.g. skip logic was generally used for all items, or not at all). The research team met on two occasions (August 23rd and September 6th, 2018) to discuss the wording, layout and formatting of ModRUM. In **Figure 4.1**, the first prototype of the ModRUM core module is presented, which includes items that made it through the final round of the Delphi study (71). In **A2.1** (**Appendix 2**), the first prototype of the core and depth module is presented. As respondents may run out of space in tables, continuation tables were developed for each table included in ModRUM. The continuation table developed for the core module is presented in **A2.2**. The sections below describe how these prototypes were constructed by the research team.

Figure 4.1 ModRUM core module prototype (first version developed with the research team)

Resource use questionnaire

We would like you to answer some questions about the health care you have used [because of XXXXX/for any reason] [in the last X months/since XXXXX]. We only want you to include health care you had as an NHS patient.

It is important for us to find out what health care you have and have not used, so please answer all of the questions, even if your answer is zero. If you are unsure of an answer, please write your best guess.

1. How many times have you been to a hospital Accident and Emergency (A&E) department? ☐
2. How many times have you been to hospital for an outpatient appointment? ☐
3. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)? ☐
4. How many times have you been to hospital for an overnight stay? ☐

For each overnight stay, please write down the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital
1	☐ nights
2	☐ nights
3	☐ nights
4	☐ nights
For any other overnight stays in hospital, please complete Table 1 on page X.	

5. How many times have you visited a doctor (GP) at a GP surgery or health clinic? ☐
6. How many times have you visited another health care professional (e.g. a nurse or physiotherapist) at a GP surgery or health clinic? ☐
7. How many times has a doctor (GP) visited you at your home? ☐
8. How many times has another health care professional (e.g. a nurse or health visitor) visited you at your home? ☐
9. How many medications have been prescribed for you? ☐

Core module, version 6, 16/01/19

4.4.1.1 Question layout

For the layout of questions, options included standalone questions by resource type and a table with multiple healthcare resources. While opinions varied on which option was most preferable, for example, a table potentially being preferable if there is a large list of questions, the group agreed that standalone questions would be the most appropriate option for a brief questionnaire. It was also agreed that questions would be drafted using Arial font, in font size 12.

4.4.1.2 Skip logic and instructions

Around half of the RUMs reviewed used skip logic. The team had mixed preferences over whether skip logic should be used in ModRUM core module. It was acknowledged that the use of skip logic may overcome ambiguity of whether missing values were truly missing or zero. However, it was agreed that using skip logic would both extend and potentially overcomplicate ModRUM, as respondents would be required to answer more questions and additional instructions would be needed to navigate respondents through the questions. The group agreed that skip logic would not be appropriate for a short paper version of ModRUM; however, for an online version it may be easier to embed. To minimise the missing data problem, where respondents do not include zero answers, the group agreed that clear guidance should be given in the instructions, that respondents should still provide an answer when they have not used a resource. In the instructions, to minimise missing data, the team also agreed that respondents should be guided to provide an estimate or best guess, even if they were unsure of the answer.

4.4.1.3 Recall period

For questions asking about GP contacts, the most common recall period in existing RUMs was three months (24%); however, there was a wide range of recall periods, from two weeks (one RUM) to 12 months (4 RUMs). Some RUMs also used anchor points, which included: since the participant last completed the RUM, since a certain date and since an event occurred, such as an operation. The group agreed that setting a recall period at this point would not be appropriate. To align with ModRUM design principles, there was a preference for the recall period to be adaptable by the end-user. However, to minimise patient burden, it was agreed that the same recall period should be adopted for the whole RUM rather than different recall periods for

each resource, which could hinder comprehensibility. The group decided that the recall period would be discussed with patients in interviews (**Chapter 6**) and different recall periods would be tested in the patient pilot study (**Chapter 8**). The latter evaluation did not ultimately happen as the Covid-19 pandemic meant that the study design changed to focus solely on validity rather than testing different recall periods.

To keep questions concise, it was agreed that information that could be repeated in each question, including the recall period, the reason for resource use (e.g. condition-specific) and who funded the resources (e.g. NHS resources only), would only be included in instructions at the beginning of ModRUM. It was agreed that the effectiveness of this (that is, whether patients remember these details) would be reviewed during testing. The team also discussed whether emphasis, such as using italics, underlining or bolding, should be used ModRUM. Within existing RUMs, 98 percent used some form of emphasis. The team agreed that emphasis should be used, but that it should be kept consistent (e.g. bold fonts only) and limited to key terms in the questions (e.g. the healthcare service, such as outpatient appointment). The reason for emphasising certain terms is to avoid errors by distinguishing important details for respondents, when questions are otherwise relatively similar. In the initial prototype of ModRUM it was decided that emphasis should be left out, so that issues with the content could be explored first. The team agreed that patient input on acceptable emphasis would be valuable in interviews (**Chapter 6**).

4.4.1.4 Inpatient and day case stays

Most existing RUMs asked for details on both inpatient and day case stays, and this was equally split between those who asked for the information together and those who asked for the information separately. As the unit costs usually differ between inpatient and day case stays, the research team decided that inpatient and day case stays should be included as separate questions. As the cost of an inpatient stay could vary substantially based on length of stay, the research team also agreed that a table should be included in the core module for inpatients stays, where the number of nights could be reported for each stay. It was also agreed that continuation tables would be drafted so that the nights for additional stays, beyond the number of rows included in the main table, could be recorded.

4.4.1.5 Outpatient appointments

While most questions in existing RUMs asked for the number of visits (88 percent), others asked for the frequency of visit (e.g. weekly visits) or number of visits with categorical answer options. The research team agreed that for costing purposes the most appropriate format was to ask the exact number of visits. There were multiple options for 'number of times'; the group agreed that the initial version of ModRUM would include a frequently used option from the existing questions, and this would be taken to patients to confirm whether it is appropriate. The research team agreed that the wording should be consistent across questions and that starting each question with 'How many times' was the most appropriate option.

'Professional seen', 'speciality' and 'reason for your visit' were included in some of the existing questions; however, as they did not make it through the final round of the Delphi survey (71), the research team decided they should be included in depth questions rather than the core module. Issues were raised around whether a patient can distinguish between a consultant and a non-consultant. It was also acknowledged that in the National Schedule of NHS Costs, which is commonly used to cost resource-use data, the costs are driven by whether an appointment is consultant-led or not; so, while a patient may not have seen a consultant, the appointment could have been consultant-led, which is information the patient would not know (48). The research team thought that 'speciality' would be jargon for patients and decided 'clinic type' would be an appropriate term to use instead and explore with patients in interviews (**Chapter 6**). The research team thought that the use of tables would be more appropriate for depth questions, where respondents would record one appointment per row. Also discussed was how to phrase 'reason for your visit', as ideally this question would elicit what interventions, tests or procedures were performed. The research team decided that the wording should initially come from existing questions and that this question should be explored with patients.

For outpatient appointments, existing RUMs also included the following items: name of hospital, duration of outpatient visit and cost; however, as they were not included in the Delphi study, the research team agreed that they would not be included in the either version of ModRUM (71).

4.4.1.6 Emergency care

The group agreed that the number of visits to Accident and Emergency (A&E) and the number of visits to A&E that led to an inpatient admission should be collected as this followed the items that were identified for inclusion in the core module in the Delphi study (71). The most common wording of A&E from existing RUMs was used in the initial version of ModRUM, with agreement that comprehension would be explored in patient interviews (**Chapter 6**).

4.4.1.7 Care at a GP surgery or health clinic or other community setting

Within the existing RUMs there were several terms used to describe a GP, including doctor, family doctor, primary care physician, general practitioner and GP. In addition, the word preceding GP also varied, for example questions included ‘a GP’, ‘your GP’ and ‘any GP’. The research team agreed that the question should capture visits to any GP and therefore the word ‘your’ should be avoided. It was also agreed that patients should inform the wording of GP, but as a starting point the initial draft of ModRUM should include the most frequently used terminology from existing RUMs which was ‘General practitioner’ and ‘GP’. The research team also discussed a paper version glossary of terms but concluded that as the aim is to develop a short questionnaire, a terminology sheet would be an inappropriate extension of ModRUM.

The research team discussed which types of healthcare professionals should be prespecified in ModRUM in questions on care at a GP surgery or health clinic or other community setting. The group agreed that which professionals to prespecify should be driven by the difference in costs between the professionals and by which professionals patients are able to tell the difference between. Supervisors involved in the Delphi survey (71) explained that the original vision was that there would be a GP question and an ‘other’ question to incorporate all other healthcare professionals. The research team agreed to continue with the original vision as they believed that it met the cost difference criteria.

Details were also extracted from existing RUMs on whether questions asked what happened during the appointment, the duration of the visit, who paid for the visit and who provided the visit. Few existing questions (less than 14 percent) asked for any

of these details. The research team agreed that none of these details should be included in the core module as they did not make it through the final round of the Delphi survey (71).

4.4.1.8 Healthcare at home

Based on the Delphi survey there was consensus that healthcare at home should be separate from healthcare in the GP practice or health clinic (71). The wording of the question for the initial version was informed by the most popular wording in existing questions which was ‘at home’. The group also decided that to follow the ‘care at a GP surgery or health clinic or other community setting’ questions, the initial version would include GP and other healthcare professional questions, with examples of other healthcare professionals included in the question.

4.4.1.9 Medications

Existing RUMs captured information on both prescribed and over-the-counter medications; however, as the core module of ModRUM is designed to capture NHS resource-use only, it was agreed that medication questions would only ask for prescribed medications. Prescribed medication questions in existing RUMs included numerous details, such as dose, form and cost; however, the results of the Delphi suggested that only name/class of medication should be captured in the core module (71). In the Delphi study medications did not meet prespecified criteria for inclusion, but ‘name/class of prescribed medications’ was included based on discussions at a final item selection meeting (see **Box 2.1** for more information) (71). The research team considered including ‘name/class’ but concluded that it would be preferable to have number of prescribed medications, to avoid adding a question that requires a free-text response in the core module. For the depth questions, the team agreed to include a question asking whether respondents pay for their prescriptions. This was included so that NHS expenses could be separated from personal expenses.

The team also discussed whether medication names should be prespecified in ModRUM depth questions. Experience from one team member, who had gained feedback on this from members of patient and public involvement groups previously, was that prespecified names were preferable. The research team agreed that prespecified names could be added to the question by trial research teams using

ModRUM in the future, so that the most relevant medications for the patient group could be included.

4.4.2 Feedback from experienced PROM developers

On January 23rd, 2019, I met with the first experienced University of Bristol-based PROM developer and on March 11th, 2019, I met with the second PROM developer. In consultation with the research team, I revised ModRUM based on the feedback following each meeting. A summary of the feedback given by PROM developers is presented below in **Table 4.2**. Changes made as a result of PROM developer feedback included: (1) using tick boxes for response options, (2) repetition of the recall period at the beginning of each section and (3) a new heading at the beginning of ModRUM to indicate to respondents the purpose of the questions. Following the changes, the research team agreed that the revised version of ModRUM was ready for testing with health economists in qualitative interviews (**Chapter 5**). These versions are presented in **Appendix 3**, **A3.3** and **A3.4**.

Table 4.2 PROM developer feedback and subsequent actions

Feedback and suggestions	Agreed actions
PROM developer 1	
Consider numbering the pages and allow them to be adapted by the end user, if they plan to include ModRUM in a larger pack of questionnaires.	This suggestion was implemented.
Add navigation instructions following continuation tables e.g. return to question X, page Y after continuation tables.	This suggestion was implemented.
Consider including continuation tables as standard, whereby researchers can remove them and related instructions if they are not required.	This suggestion was implemented.
Consider adding 'in the last X months' to each question. In their experience patients forget the recall period as they progress through a questionnaire.	No change was made. The research team had already discussed this and agreed to leave it out of each question, to keep questions concise. Further testing was needed to see if this was an issue for patients.
The response boxes are of most concern, as respondents may miss them or use them to write free text. This issue is unlikely to be revealed in cognitive interviews with patients, as respondents act differently in interviews and are less likely to leave missing answers. Other options should be considered, such as those presented in **A2.3**.	The research team agreed that this was a concern and given the potential that this issue would not be revealed until piloting (**Chapter 8**), the team decided to take a proactive approach and adopt suggestion 1 (**A2.3**), which included tick boxes for '0' to '4' and a 'other' box for responses more than 4. This design meant that for most questions, one tick would be required.
Depth questions 2 and 10, which do not appear in the core module, could be moved to the end of the core module with depth questions version, so that question numbers relate to the same resource in both versions of ModRUM.	No change was made. It was agreed that this would be more of an issue for researchers (i.e. matching corresponding questions), but for patients it is better to group related questions together.
Consider asking the total number of inpatient nights rather than nights per stay as this would eliminate the need for the table in the core module and a continuation table for the core module.	No change was made. It was agreed that nights per stay was more desirable, so that patients do not need perform addition of nights for multiple stays.
Consider navigational instructions around the inpatient question for participants who have zero inpatient stays.	No changes were made. It was agreed further testing was needed to see if an issue materialised.

Table 4.2 continued

Feedback and suggestions	Agreed actions
The medication question is concise but unclear. It is unclear whether it asks for the number of different medications or number of prescriptions written.	Additional explanation was added "e.g. if you have been prescribed Medication A once and Medication B twice, your answer should be 3"
PROM developer 2	
Consider reformatting the core module inpatient table to formatting more consistent with other questions. It changes the format of ModRUM when respondents have just become familiar with the style.	I drafted an alternative option using tick boxes that were similar to other questions. However, deviations from other questions were required (e.g. no zero-tick box was necessary for number of nights per stay), so the team agreed to keep the table and test it further.
The recall period should be repeated at least at the start of every section. It could also be emphasised using bold or larger font.	As repetition of the recall period was also raised by PROM developer 1, the team agreed to repeat the recall period for each new section. I also suggested that condition-specific, or all-cause resource-use should be repeated too, and the team agreed.
Include a name for ModRUM that indicates to patients what they will be asked in the questionnaire (e.g. healthcare use questionnaire or your use of healthcare services)	This suggestion was implemented, with 'your use of healthcare services' replacing 'resource use questionnaire'.
Change the order of sentences in the introduction so the sentence containing the most important information is presented first.	This suggestion was implemented.
Make it more obvious, potentially using bold font or two-part questions, that GP questions are only for reporting GP contacts, not other healthcare professionals at a GP practice.	I drafted both of the suggested options and the research team agreed that two-part questions were preferrable, so this change was implemented.
Consider revising instructions to go to continuation tables to 'If you need more space, please use Table X on page Y".	This suggestion was implemented.
If respondents do not know the name of a medication, they may leave the question blank. Consider including changing to 'Name (if known), or type (e.g. painkiller)'.	It was agreed that the following should be added: 'if you don't know the name, please put the type of medication e.g. painkiller'.

4.4.3 Aesthetic enhancements and logo design from the design agency

Following a face-to-face meeting with the design agency (November 26th, 2019), they provided revised versions of the modules and potential logos for ModRUM. I liaised with the agency, over email. Several iterations to the logo and modules were made until it was signed off on February 28th, 2020. The logo is presented in **A2.4.** This version was ready and used in the third round of qualitative interviews with patients (**Chapter 6**), and the aesthetic changes can be seen in **Figure 7.1** (core module) and **A5.3**, **Appendix 5** (core module with depth questions).

4.5 Discussion

4.5.1 Chapter summary

In this chapter, I have described how the first prototype of ModRUM was developed, how the prototype was revised based on feedback from PROM development experts and the input from a design agency to develop a logo for ModRUM and improve the aesthetics of ModRUM.

4.5.2 Strengths and weaknesses

Utilising a range of materials, including existing RUMs and questionnaire design literature, I was able to develop drafts of ModRUM from items identified in the Delphi study (71). Existing RUMs offered a range of alternative options for formulating questions. Development was conducted in consultation with the research team, which included researchers with health economics backgrounds, with significant experience of using RUMs in economic evaluations. I was also able to obtain valuable feedback from experienced PROM developers on the first drafts of ModRUM. Consistent with health measurement instrument development guidelines, I followed an iterative approach which involved drafting questions, gaining feedback on them, revising them, and eliciting further feedback.

The use of existing RUMs and a research team including experienced health economists meant that consideration to how resources would be costed was at the forefront. ModRUM includes a range of healthcare metrics, which is driven by how the healthcare would be costed. For example, secondary care questions generally require the participant to report what service they have used (e.g. inpatient stay) and primary and community care questions generally ask about service contacts (e.g. GP

consultation). Healthcare interventions are covered in the medications and depth secondary care questions (e.g. tests and procedures). While healthcare interventions may allow for increased costing precision, they may also be more difficult to recall and analyse. However, if more detailed information, i.e. at intervention level, is likely to impact cost-effectiveness results, it should ideally be captured.

Cross-cultural validation of ModRUM, for use internationally, is beyond the scope of this thesis, as I am developing ModRUM in the context of the UK healthcare system. However, I utilised RUMs stored in DIRUM to inform question development, which includes RUMs designed for non-UK-based studies. While only five RUMs (nine percent) were included from non-UK-based studies, they presented options that were not relevant for a RUM designed for UK-based studies. For example, 'physician' was extracted as a synonym for GP; however, it is not commonly used in the UK. While caution was needed for terminology, reviewing the layout and format of RUMs designed for non-UK-based studies was still informative for layout and format options for ModRUM. Using existing instruments when developing a new instrument is advised in questionnaire design literature (79); however, as demonstrated in **Chapter 3**, many existing RUMs have not undergone psychometric testing, so while I considered the use of existing RUMs a sensible starting point, extensive validation was still required.

Involving PROM developers in the development of ModRUM was valuable as they had insight into potential problems that could occur with the design of a new instrument. Feedback from these experts was informal and advisory, rather than a formal step in the research process. While it may have been valuable to conduct a qualitative study with questionnaire developers, health economists who were interviewed in the next study (**Chapter 5**) were recruited specifically for their experience of leading trial-based economic evaluations, which will have involved the development of RUMs.

4.5.3 Comparison to existing literature

During the development of the prototype, I took into consideration the design principles that are specified in **Chapter 1** (**Table 1.1**). One of the aims of developing a well-validated RUM is to improve the quality of self-reported data on healthcare

utilisation. This requires a balance between the level of detail that is required for accuracy in costing and burden on participants. For example, for medications the prototype requires minimal information from participants, and I believe the design of the depth medication question was enhanced by PROM developers who suggested adding instructions indicating to respondents that they should write the type of medication if they did not know the name. I believe this decision is supported when considering it against previous research, such as literature on the development of the TiC-P (124). The developers stated that four details were required for costing medications, namely, *"the name of the medication, dose per intake, the daily dose and the number of days that the medication was used"* (124) (pg.5). However, they found that at least one item was missing for 29% of cases, which was most often the name or dose per intake (124). They concluded that it may be possible to reduce the number of details requested (e.g. using daily defined dose instead of participant report), or use an alternative source, such as medical records, if more detailed data are required (124). Including less-detailed information on resource use is also supported in an article by Ruof et al., where in interviews with patients and a study comparing self-report to administrative data, they found patient-preference and increased accuracy for questions that required less detailed responses (147). Heslin et al. found support for less resource-intensive approaches to costing medications when medications contribute less to total costs, which would also placate the need to capture as much detail from trial participants (160).

4.5.4 Implications for this research

Formulating ModRUM in consultation with the research team and experienced PROM developers highlighted multiple areas where input from health economists and patients was required in further testing. The research team agreed that while selecting terminology for healthcare resources from existing RUMs was a practical starting point, comprehension would need to be tested with patients as it is unlikely that existing RUMs would have undergone thorough testing based on the findings in **Chapter 3**. Both discussions within the research team and feedback from PROM developers highlighted that patient input was required on the appropriateness of the recall period and how often it should be repeated within ModRUM. Potential issues with response options highlighted by feedback from the first PROM developer led to

the inclusion of tick boxes for response options. I identified that patient feedback was required on the formatting of response options in patient interviews (**Chapter 5**).

Future testing, following the formulation of ModRUM was important to identify whether any questions were particularly problematic. For example, despite keeping the prescribed medication question concise, it may be more difficult for both respondents recalling the data and health economists sourcing unit costs, when compared with other questions. Other questions that may be more problematic are other healthcare professionals seen at the GP surgery or home in the core module. These questions are more cognitively challenging than other questions as they require respondents to recall different healthcare professionals seen and sum the number of contacts with each professional. Although in a different context (magazine readership and television viewing research), research on questionnaire design has also found underreporting of items that are not explicitly mentioned or referred to as 'other', when compared with items that are clearly defined (161).

4.6 Conclusion

In this chapter, I have described how I constructed ModRUM from items identified in a Delphi study. The process involved reviewing existing measures stored within DIRUM, presenting options to the research team and gaining feedback from researchers experienced in PROM development. The chapter also describes how a design company was involved in the development of ModRUM, to enhance aesthetics and design a logo. In the next chapter, I describe qualitative interviews with health economists, where I tested face and content validity of ModRUM and the suitability of ModRUM for costing purposes.

Chapter 5 Qualitative interviews with health economic experts

5.1 Chapter overview

This chapter reports on qualitative interviews with UK-based health economic experts who had recent experience of undertaking trial-based economic evaluations. Experts were asked to review and provide feedback on ModRUM. The face and content validity, as defined in **Chapter 2**, of ModRUM were tested and experts assessed whether ModRUM captured resource-use data that were suitable for costing purposes within economic evaluations.

This chapter includes background information on the purpose of this work and details on the identification and recruitment of health economists. Qualitative methods for interviewing are described more generally in **Chapter 2**; in this chapter the methods used for expert interviews are described in more detail. A results section first details expert identification and recruitment. It also includes a descriptive summary of the feedback with illustrative quotes. The final part of the results section describes how the feedback was used to refine ModRUM and to inform the topic guide for interviews with patients, described further in **Chapter 6**. The chapter finishes with a conclusion.

5.2 Background

Questionnaire development literature suggests that experts should be consulted to review the initial pool of items for a new questionnaire (101). Experts can improve the content validity of a questionnaire by advising on: 1) the relevance of items, 2) the clarity and conciseness of items, and 3) whether relevant items have been omitted (101). Partial content validation was undertaken during the Delphi study that preceded this work, where experts were given the opportunity to rate the relevance of items for inclusion in a core standardised set of resources use items and suggest items that were omitted from the Delphi survey (71).

In the work described in this chapter, the content validity of ModRUM was tested with health economics experts, who provided feedback on: [1] the relevance of items included in ModRUM, [2] the clarity and conciseness of the questions, and [3]

whether relevant items were omitted from ModRUM. This research study extended the assessment of content validity undertaken in the Delphi study (71), as health economists were asked to provide feedback on resource-use questions, in the context of a formulated RUM, as opposed to judging items without knowledge of what other items would be included in the core set of resources-use items. Testing the face validity of a questionnaire requires a subjective judgement by experts; experts were therefore asked to state whether ModRUM appeared to measure the desired qualities (79). Experts were also asked whether the questions would generate answers that are suitable for valuation purposes within economic evaluations. That is to say, unit costs from appropriate sources, such as the Unit Costs of Health and Social Care, can be applied to responses from ModRUM to estimate the cost of healthcare resources (66).

5.3 Methods

5.3.1 Health economist identification and recruitment

Health economic experts were identified using a purposeful sampling strategy to ensure that 'information-rich' experts who had recent experience of using RUMs in trial-based economic evaluations were recruited (100). A list of senior health economists, excluding members of the supervisory team, who were co-authors on recently published (between November 2015 and October 2018) economic evaluations alongside RCTs funded by NIHR HTA and Public Health Research (PHR) programmes was formed. Being a senior health economic co-author on a recent NIHR report indicates they have recent experience of working with RUMs and can be considered subject experts, able to provide judgement on the face and content validity of a RUM designed to capture healthcare resource-use data in a UK-based RCT.

Maximum variation sampling was used to ensure that a wide range of experts were recruited based on several characteristics including: the funding stream of the report (HTA/PHR), workplace, geographical region, research project setting, and disease area/condition/preventative intervention (104). These characteristics were included so that the range of trial-based economic evaluations that may be conducted in the future, with ModRUM included to collect healthcare utilisation data, could be represented. The characteristics and the name and job titles of experts were shared

with the research team and the team identified experts with experience of working on a wide range of trial-based economic evaluations, who were approached to participate. Experts were recruited from around England, Wales, Scotland, and Northern Ireland with the aim that they would report on any geographical variations in healthcare terminology or differences in valuation methods. As experts had already informed the content of ModRUM in the Delphi study, a sample size of 12 was selected, as this covered each region of the UK and the research team felt this would be sufficient to identify any major issues with the questions and ModRUM more generally.

I approached experts by email. The email briefly described the aim of the study, why the expert was being asked to participate and what would be involved if they decided to participate. Attached to the email was an information sheet (**A3.1**, **Appendix 3**), which provided more detail on the aims of the research, the role of the expert and details on how ModRUM had been developed. Literature suggests that people are more likely to participate in research if they feel a personal connection (102); for this reason, all supervisors were copied into the email as the experts were likely to have an existing professional link to at least one of the supervisors. If an expert did not reply within one week of the first email, a follow-up email was sent. If the expert did not respond to the follow-up email, another expert from the same region was invited to participate. Prior to deciding on whether to take part, experts had the opportunity to ask questions about the research and their potential role within it.

5.3.2 Data collection

For experts who were interested in participating in the study, a date and time for a telephone interview to provide feedback on ModRUM was organised. Experts were advised the interview would take approximately 30 minutes. Experts were sent a consent form (**A3.2**) which they were asked to sign and return prior to the interview. Experts were asked to give consent for the interview to be audio-recorded, which allowed me to concentrate on the interview and capture expert opinions accurately. Each expert was sent drafts of ModRUM core module (**A3.3**) and ModRUM core and depth module (**A3.4**) via email, so that they could review and become familiar with them ahead of the interview.

Ethical approval for the research conducted in this chapter, **Chapter 6** and **Chapter 8** was provided by South Central - Berkshire B Research Ethics Committee (REC reference 19/SC/0244).

5.3.3 Qualitative interviews

During the interview, I asked experts pre-defined questions from a topic guide (**A3.5**) which included questions on the relevance, clarity, conciseness, and potential omission of items in ModRUM. The aim of the interview was to establish whether ModRUM questions captured what they purported to measure, and whether the questions generated answers that were suitable for valuation purposes within economic evaluations. A responsive interviewing technique was used to allow me to be "flexible and adaptable" in the research design, meaning that future interview questions could be framed on the answers to previous interview questions and quick changes could be made to the line of inquiry when unexpected topics arose (102).

5.3.4 Analysis of expert feedback

I transcribed the audio-recordings taken during each interview. Transcriptions were then uploaded to and coded in NVivo 12 Pro (103). The transcriptions were read line-by-line and expert feedback was coded under categories and sub-categories. Categories included higher level concepts representing more general themes (e.g. the 'outpatient' question is inadequate for costing), while sub-categories were more specific and represented lower-level concepts which showed the variation in themes under each category (e.g. 'speciality' should be captured for increased precision) (90). Analysis was conducted concurrently to interviews. The analysis drew on methods of constant comparison where feedback provided by experts was continually compared to identify the common themes (98). Attention was paid to any contrasting feedback to ensure that the opinions of all experts were considered.

As a next step, I presented the research team with the findings from the interviews, grouped under key aspects of feedback with relevant illustrative quotations. The research team discussed each aspect of feedback, and with reference to the objectives of ModRUM, decided whether ModRUM should be adapted or whether suggested changes should be explored further with patients. The objectives of ModRUM are defined in more detail in **Chapter 1 (Table 1.1)**, but in short include

that it should be generic (include questions relevant to all healthcare RCTs, particularly in the core module), precise (for increased accuracy in cost estimates of highly utilised resources or key cost drivers), comprehensible (to respondents), flexible, concise, consistent and transparent. When considering changes, both the respondent and analyst burden were considered. Following research team discussions, I revised ModRUM, and the research team subsequently reviewed the revisions to check that they adequately reflected the agreed changes.

To validate the changes, as a form of respondent validation, the updated modules and a summary of the feedback were sent to experts (162). While not specifically asked for, experts were advised that they could share any further feedback on the updated modules or summary document.

5.4 Results

5.4.1 Identification of experts

Between November 2015 and October 2018, 68 senior health economists from 28 different workplaces, predominantly academic institutions, were named on HTA and PHR reports that included trial-based economic evaluations. Of these, 59 were authors on HTA reports only, 4 were authors on PHR reports only and 5 were authors on both HTA and PHR reports. Studies were undertaken in a range of settings including primary, secondary and community care, and school settings. There was also a wide range of disease areas, conditions and preventative interventions studied including cancer, dermatology, diabetes, mental health, obesity and physical activity.

5.4.2 Recruitment of experts

Twelve senior health economists, one from each region of the UK, were initially invited to participate. Of these, ten health economists were available and agreed to provide feedback on ModRUM and two (17%) did not respond. Two further health economists were invited from the regions I did not recruit from. Of these, one agreed to participate and the other could not commit due to timing. Of the 11 interviews organised, 10 went ahead as planned and one was cancelled due to the health economist no longer being available. The participating health economists were from 10 different workplaces, from around the UK. Participant characteristics are provided

in **Table 5.1** and NIHR HTA or PHR project characteristics are provided in **Table 5.2**. The health economists had considerable recent experience of trial-based economic evaluations with 29 HTA and five PHR publications between them. The studies reported in the publications were undertaken in a range of settings with a wide range of disease areas, conditions and preventative interventions studied.

Table 5.1 Participant characteristics

	Mean	(IQR)
Number of reports	3.4	(2-4)
	n	**(%)**
Journal		
Health Technology Assessment	6	(60)
Public Health Research	2	(20)
Both	2	(20)
Job role		
Professor	7	(70)
Reader	1	(10)
Associate Professor	1	(10)
Health Economist	1	(10)
Region		
East of England	1	(10)
London	1	(10)
North East	1	(10)
Northern Ireland	1	(10)
Scotland	1	(10)
South England	1	(10)
South West	1	(10)
Wales	1	(10)
West Midlands	1	(10)
Yorkshire and the Humber	1	(10)

Table 5.2 Project characteristics

	n	(%)
Journal		
Health Technology Assessment	29	(85)
Public Health Research	5	(15)
Disease area/condition/ preventative intervention		
Pregnancy	6	(18)
Mental health	5	(15)
Body weight	2	(6)
Back pain	2	(6)
Smoking	2	(6)
Orthopaedics	2	(6)
Alcohol	1	(3)
Blood borne viruses	1	(3)
Brain injury	1	(3)
Cancer	1	(3)
Dementia	1	(3)
Dental	1	(3)
Dermatology	1	(3)
Epilepsy	1	(3)
Heart	1	(3)
Immune disorder	1	(3)
Learning disability	1	(3)
Liver disease	1	(3)
Lungs	1	(3)
Substance misuse	1	(3)
Diabetes	1	(3)
Research project setting		
Secondary care	14	(41)
Community	8	(24)
Primary care	5	(15)
Schools	4	(12)
Primary and secondary care	1	(3)
Ambulance services	1	(3)
Dental practices	1	(3)

5.4.3 Descriptive summary

Health economic experts provided feedback during interviews conducted in May and June 2019. Interviews were on average 44 minutes long (range: 29 to 58 minutes). This section includes a descriptive summary, with illustrative quotes of the feedback provided. General feedback is summarised first, then item-specific feedback, and feedback on the omission and overlap of items is provided. No further feedback was provided by experts upon receipt of the revised version of ModRUM. Pseudonyms have been used to protect each expert's identity, with 'E' indicating expert. Actions agreed by the research team in response to feedback is provided in section **5.4.4**.

5.4.3.1 General feedback

Introduction

Most experts provided positive feedback about the introduction text, suggesting that it was clear and concise. However, E4 and E9 suggested that alterations could be made to the text 'healthcare you have used as an NHS patient' to enhance clarity about what should be included. E4 also said that they usually specify that trial-related visits should be excluded, within the introduction.

> E5: *"I liked the introduction, it was short, I like that you said best guess, I like that you said answer even if your answer is zero."*
>
> E9: *"… 'care you have received as an NHS patient'?... It could be interpreted differently by different people. Maybe it could be 'your use of NHS services and not private or things you have paid for'."*
>
> E4: *"… maybe missing from the pre-text, in a lot of the studies we ask them to exclude trial related visits, to avoid double counting."*

Two experts suggested that when using ModRUM in a study the user may want to alter the wording of the introduction, so that it fits better with the language used in other study documentation and is suitable for the study population.

> E4: *"… it's hard to plonk a questionnaire which is like this, which is quite wordy into something else, and the language changes… I don't think you would change the language a lot necessarily, but it just depends, you don't want it to seem suddenly out of kilter with the way you have designed it here."*
>
> E6: *"… you have to take into account the population, age, reading level… make sure it is written in a lay and active tense."*

Question order

Experts agreed that questions should be grouped by type of care (e.g. primary, secondary). Four experts provided positive comments on the order of questions.

> E1: *"I think this makes sense because you are talking about hospital first and they tend to be the most costly services."*
>
> E3: *"It's good to have medications at the end as people can get bogged down with that."*

However, over half of experts suggested alternative ordering. Several experts said they would start with the key cost drivers or resources that are more likely to be used, so that if the patient does not complete ModRUM, the most important information is captured first. E6 said they sometimes order questions as a journey through the care patients receive, starting with primary care.

> E4: "…I would have started with something that everyone is likely to have had… Going to a GP, I would have thought that is the one most people would report something on. Just to get them engaged in the questionnaire…"

Response options

While four experts provided positive comments on the response options, E2 and E9 said they would prefer one box where respondents could enter a number instead of using a tick box.

> E5: *"I do like the use of the zero box, it is quite clear for people have zero to tick zero."*
>
> E2: *"…there are no instructions on the numbers, it looks like 0 to 4 could be construed as being for the researchers/office use... Personally, I prefer a box with a number, but it is a matter of preference, it wouldn't stop me using it…"*

Section headings

Several experts said that they would usually include section headings for each resource category, which may help to avoid double-counting, where some services are provided in both community and secondary care.

> E2: *"We tend to split them up into 'these are hospital-based service' and 'these are services delivered in the community'. For example, you get hospital and community physiotherapists."*

E7: *"Maybe it would help to flag each section 'hospital services', 'primary care/care in your community'. See if patients find it helpful or if they even notice it. If patients see clear heading they can work through each section, it focuses them and provides structure, otherwise a list of questions may seem overwhelming."*

Continuation tables

While one expert thought the continuation tables were adequate, six experts provided negative comments, with the main themes being that the tables were too long, tables on different pages would create confusion for patients and the tables would not be well-received by trial teams, who would be considering the cost implications of printing and posting additional pages that may not be used. Several experts had mixed feelings about the continuation tables. While most of these experts said that the tables were quite bulky, they also acknowledged that most patients would not need them, so they are fine to include.

E2: *"I think you are going to struggle getting anything completed when you have tables that are continued on lots of different pages. People get confused and get fed up of doing it."*

E3: *"It makes the forms quite bulky. There's not many people that will have more than the number in the tables in the module… If someone is having more contacts you want to record it and you don't want it recorded in the margin of the questionnaire so I think they are fine."*

Four experts agreed that a general comments box at the end of the depth module could be used as an alternative; however, it was acknowledged that providing a box would allow patients to include unrelated free text.

E7: *"A general comments box at the end could be used to capture anything that doesn't fit in the tables. More open to patients writing unrelated free text though."*

Implementation of ModRUM

Several experts said they would need guidance on how to implement ModRUM and what assumptions should be employed for costing and analysing ModRUM responses. A guidance document that would be published alongside ModRUM was suggested. E5 said that guidance should be provided on how to cost the information captured in ModRUM to avoid double-counting at the costing stage.

> E6: *"give them [health economists] some rules, that some of this will be assumptions"*
>
> E10: *"I wonder whether given that we do have now, quite solid NHS reference costs and the PSSRU, if it would be nice to see something to link resource use and unit cost."*
>
> E5: *"when you are talking about procedures… making sure that you are not double counting time in hospital and operations twice. That's about how you handle the data and what number you assign to activity."*

Two experts discussed what guidance could be included in the document. Both thought that more detail should be collected when the resource is likely to be highly utilised or a cost driver.

> E5: *"I might go for more granularity where I think there are going to be major cost drivers."*
>
> E11: *"If you are going to get a small number of responses you could look at an average cost. If you are going to get hundreds of positive responses, then you may want to cost them by the areas they are treated in."*

Two experts discussed how users would implement ModRUM. E6 suggested an item library of validated questions, akin to what is already being used for some PROMs (163). E11 discussed how users would formulate a trial-specific version of ModRUM from the modules.

> E6: *"I think one of the things PROMs have moved towards is, once they are validated, they are moving towards more of an item library. You have core items and then other things that can be added in, but the questions are all written, and you can tailor them to your study."*
>
> E11: *"One way if you had it on a computer, and you had all of these questions in a menu that you could drag across and make your own questionnaire."*

E5 suggested that those implementing ModRUM may choose a level of depth between the core and the depth module.

> E5: *"… if you think of the most detailed as your longer questionnaire, and the least questions the core module, at some point in between … an analyst may have something in between the two, but they will have to justify why ..."*

Generic examples were included in ModRUM (e.g. nurse under other healthcare professional). E6 suggested that the end-user should be able to adapt the examples so that they are relevant to the population being studied.

> E6: *"I guess there would be some tailoring. For example, if it was an early years intervention then maybe you could put different words"*

Other general feedback

Other general feedback focussed on ways to ease completion and reduce the burden of completing ModRUM for patients. E6 suggested that important words, such as the resource type, should be in bold or highlighted.

> E6: *"…whether things should be in bold or highlighted. I read that one been to hospital, oh it's a day case. I think you just need to grab some words out."*

E11 expressed concerns that ModRUM may be asking for too much detail and suggested this could hinder the amount and quality of data returned, they also suggested that free-text fields should be kept to a minimum to avoid the burden of analysing free-text data. E9 suggested looking at how the precision in estimating unit costs changes as the level of detail asked changes.

> E11: *"I certainly wouldn't include more. When patients see this questionnaire, some will recoil in horror… there is a strong correlation between how big your questionnaire is and what you get back… I spoke to some of the researchers here… In terms of clinic type, reason, test, etc., they said I really wouldn't want to analyse this data as it would take a long time to apply costs… The main thing is the free text fields… and the difficulty of interpreting them."*
>
> E9: *"See if the level of detail you are asking for changes the unit costs, and if it doesn't, then I would keep it simple."*

E9 said that numbering each appointment, in the first column of each table, is unnecessary as it may give patients the impression that appointments need to be sequenced when the order is insignificant for the health economic analysis.

E9: *"… listing out the outpatient appointments 1 to 6, giving it an order like that is unnecessary because people might get into the mindset that there is an importance of the ordering."*

When probed about repetition of the recall period and resource type, E6 stated that the amount of repetition is fine, providing it is consistent, although they suggested that it might differ between studies.

E6: *"Yes, as long as it is consistent. Get the balance right on what suits your study."*

5.4.3.2 Item-specific feedback

Feedback in this section relates to specific core questions (CQs) and depth questions (DQs). CQs and DQs that were sent to experts can be seen in **A3.3** and **A3.4**, respectively.

A&E (CQ1, DQ1)

Although it was acknowledged that more detail could be captured on A&E visits, most experts thought that CQ1 was adequate for capturing the required information in a core module and could be used for costing.

Expert 7 (E7): *"If this is a minimum you can capture because of patient burden, then I think as a cost driver, is adequate… I could cost this. I have asked this question before and used an average unit cost."*

Several experts said that the cost of A&E varies dependent on how a patient presents at A&E, whether by ambulance or by going to A&E themselves. While E7 and E9 thought this additional detail could be added to the depth module only, E3 thought that given the cost difference between a patient that arrives by ambulance or using another mode of transport, more detail should be captured in the core module.

E3: *"Some people will go to A&E by ambulance and some of their own accord and there is a massive cost difference between them so you may want to split out of the two."*

Only E11 thought DQ1 may be inadequate for capturing A&E in a depth module. They thought that more detail may be needed to cost A&E attendances more accurately when utilisation of A&E attendances is expected to be high. However, as they did not have experience of a trial where high usage of A&E was anticipated, they did not have suggestions for the extra details that could be captured.

Paramedic care (DQ2)

Almost half of the experts thought that DQ2 would be adequate for most trials. However, E1, who had experience of collecting paramedic data through administrative sources, thought that DQ2 may not capture enough detail for costing, as different variations of paramedic care have different costs. This expert did however acknowledge that DQ2 could be useful in some contexts.

> E1: *"I don't think it is sufficient for costing in an economic analysis… As these are self-complete questionnaires, the question you have could be helpful in certain contexts."*

While several experts thought that DQ2 could be costed, several other experts expressed uncertainty in how they would cost this question, with E7 suggesting that there may be different unit costs for 'see and treat' and 'see, treat and convey'. When probed on whether paramedic care should be split into 'see and treat' and 'see, treat and convey', E6 said that it would depend on the unit cost differential. E5 also suggested that there is a difference in the cost of an ambulance and an ambulance car. They agreed that DQ2 could be used for most studies, but in other studies, it may be important to capture and cost more detail.

> E7: *"Not sure how I would cost this as there may be a different cost for see and treat and take to hospital."*
>
> E6: *"… you would do a weighted cost because you could have a paramedic who turns up and treats, or turns up and scoops you off to A&E… I would look and see what the differentials would be."*
>
> E5: *"I need to make some assumptions about who it is that is attending. Is it an ambulance or is it an ambulance car?... I think they are conflated in the reference costs."*

Several experts questioned whether 'paramedic' was suitable terminology, with 'ambulance' suggested as an alternative.

> E9: *"Received care from a paramedic, do you not just mean have you called an ambulance?"*

Outpatient appointments (CQ2, DQ3)

Most experts thought that CQ2 was adequate for capturing outpatient appointments in a core module, with several experts commenting that a unit cost could be applied, although it would be an average unit cost across different hospital outpatient departments. However, several experts thought that more information was needed for costing, including specialty and reason.

> E1: "*I think the precision will be far greater if you ask about the different types and then you can attach different cost."*

Five experts thought that 'outpatient' may be a problematic term for patients. Experts suggested that this term could be clarified by providing examples or using alternative terminology. Experts also suggested that CQ2 should be shown to patients to ensure the wording is meaningful.

> E9: *"I often wonder if people understand what an outpatient appointment is, it is a term we use a lot but if I ask my relatives what an outpatient appointment are, they would go 'what do you mean?'."*
>
> E7: *"Outpatient/clinic is what I put in; I think it would make it clearer. It depends on the age of the person and how the hospital is set up to call it."*

Several experts thought that DQ3 adequately captured information on outpatient appointments for a depth module, and several also thought the question could be used for costing. However, five experts thought that DQ3 was inadequate. Experts provided mixed comments on the third column, 'Main test or surgical procedure performed'. While E1 thought that all tests and procedures should be captured, E5 said that it may not be necessary to capture 'Main test or surgical procedure' for every study and several other experts thought that 'Main test or surgical procedure' was unnecessary for costing, with 'clinic type' alone being sufficient. Several experts suggested patients may struggle to complete the third column, including E4 who suggested adding 'if applicable' to the column heading, as respondents may not know how to complete the table if no tests or procedures were performed.

> E7: *"I would think the clinic type is enough as you can get a unit cost for that."*
>
> E5: *"For the third column, I'm not sure I would go to that level of detail always and I think it would depend on context."*
>
> E2: *"… the main test or surgical procedure column can be quite problematic because people aren't always sure what test or procedure they have."*

Day case admissions (CQ3, DQ4)

While four experts thought that CQ3 was adequate for capturing day case admissions in a core module and most experts thought it could be costed, two experts thought more detail should be captured due to differences in unit costs. E1 suggested differences would be driven by specialty and E3 suggested there would be a difference based on whether surgery was performed.

> E3: *"As it stands, I would probably always add the depth day case question. As a minimum I would split out day case and day surgery."*

Most experts were positive about the terminology used in CQ3. Positive comments were generally provided on the definition 'used a bed, but did not stay overnight' for 'day case'. However, two experts disagreed, with E5 stating that a bed could be used for an outpatient appointment, and E10 stating that beds are not used for all day case admissions. They suggested alternative wording which included 'day patient' and 'admitted', but acknowledged that 'admitted' may be problematic for patients.

> E5: *"I can imagine you have been lying in a bed for a little while, but it was an outpatient appointment… We've sometimes said, 'being admitted to hospital', but that requires someone to understand what admitted means. I would be interested to find out what patients think"*
>
> E10: *"You are giving the impression you get a bed and I don't think that is always the case... we did it as 'Has your baby been admitted to a hospital as a day patient?'."*

While four experts thought that DQ4 was adequate for capturing day case stays in a depth module, E1 thought it was inadequate as all tests and procedures should be collected for costing purposes. Although several experts thought the information captured was suitable for costing, several others agreed with E1 and suggested that more or alternative detail was required. They suggested 'specialty' was more important to capture than 'reason', which contrasted with feedback from E10, who said they would always capture 'reason' in order to link answers to Healthcare Resource Group (HRG) codes, which are often used for costing secondary and emergency healthcare services.

> E5: *"You have 'reason for visit' and 'main test or surgical procedure', which I think is useful but halfway house may be just asking specialty… You are greatly increasing the burden all through the study with the open text."*

Inpatient admissions (CQ4, DQ5)

Most experts thought CQ4 captured adequate information for a core question on inpatient admissions; however, E10 thought 'level of care' and E1 thought 'specialty', 'ward' and 'surgical procedures' should always be captured. While most thought CQ4, which included number of stays and nights per stay, could be used for costing, E5 thought that CQ4 would allow you to estimate hotel charges but not the procedure cost.

> E10: *"I always try to cost hospitalisations by level of care. So, if it was an ICU, high dependency, special care"*
>
> E5: *"The number of nights will allow you to get to hotel charges, what it does not allow you to get to… is procedures, so you don't know why they have been there and that raises questions of how you would work out cost…"*

Several experts thought capturing 'number of nights per stay' was adequate; however, E9 said they would prefer to omit 'number of stays' and just ask the total number of nights across all stays to reduce respondent burden. E3 thought that using 'nights', rather than 'dates', may mean that respondents include parts of stays that fall outside the recall period. E5 suggested adding several prespecified reasons with tick boxes to provide some additional information about the visit in the core.

> E9: *"just ask for the total number of nights… the majority of people over the course of 3 or 6 months will probably only have had one or two stays… the number of stays and the number of nights per stay, that it is asking for quite a lot from people..."*
>
> E3: *"In mental health, for example, patients have very long lengths of stay which may start before the recall period, some patients may exclude this, and some may include the whole duration. Dates would allow you to pick out the bit of interest."*
>
> E5: *"you could have the option that if you know something about the context of why people might be admitted. You could have some tick boxes. You might still allow writing after say 4 choices but hopefully it will be the minimal amount of writing required from anyone"*

Six experts thought DQ5 was adequate for capturing inpatient stays in a depth module, including E7 who said that while respondents may not complete all the information asked for, this may not be an issue as partially completed data could be

sufficient for costing. Several experts suggested that more or alternative detail could be captured, including E1 who suggested all reasons for a stay should be captured, instead of just the main reason, and E9 who suggested 'specialty' is preferable to 'reason for stay'.

> E1: *"Reason for stay, I wonder whether you could allow for more than one reason, so add (s)."*
>
> E9: *"If you put reason for stay you are going to get loads of things. You could maybe be quite clever and just provide the top 5 or 6 specialities. I just think reason for stay, what are you going to do with it... It's a more tenuous link to getting specialty."*

GP surgery or health clinic visits (CQ5, DQ6)

Most experts said that the GP part of CQ5 was adequate for capturing GP appointments at the GP practice or health clinic, and that they would be able to apply a cost.

> E7: *"For GP visits, yes, as a basic measure I think that's fine… For GP visits, I would know how to cost this."*

Three experts said that the other healthcare professional part of CQ5 was adequate, but five experts thought it was inadequate, and E9 provided mixed views. Of those that thought CQ5 was inadequate, four suggested pre-specifying several of the most common healthcare professionals with an 'other' option where patients can specify other healthcare professionals.

> E4: *"I don't have a problem with it being grouped. You're going to have to make an assumption for any of them for the band and they are all on the same NHS pay banding."*
>
> E5: *"You've argued that a nurse could be the most common and everything else quite rare… I think that is a reasonable case for most situations… it is not worth going to the extra effort of granularity, the increase in respondent burden does not add sufficient detail to my work and accuracy to my work to compensate…"*
>
> E1: *"other healthcare professionals, each of those would carry a different unit cost… you could break that down by type of healthcare professional… You can select the most common, and then other, please specify."*

Two experts said they could use an average unit cost of non-GP healthcare professionals to cost CQ5, but E7 said that health economists are likely to use different unit costs.

> E7: *"For other healthcare professionals, when costing this, it would differ between health economists as to what they think is appropriate to use for the average unit cost."*

When asked about the terminology used to describe a GP surgery or health clinic, E4 said that rather than 'health clinic', 'health centre' was the correct terminology in their locality. They also suggested that for less intensive trials, where less questions are permitted, they would ask a GP question that incorporates all modes.

> E4: *"Maybe 'health centre' around here is what people would call [a health clinic]"*

Over half of experts thought that DQ6 was inadequate for capturing visits to the GP practice. When asked about how the question could be improved, all experts said DQ6 should be expanded, with two experts stating that DQ6 should capture more detail.

> E7: *"I think this is where you can expand the healthcare professionals. You're allowing it to expand for telephone/online, you need to let it expand here."*

Two experts thought that the order of DQ6, DQ7 and DQ8 was confusing and said that rather than ordering by mode of visit, it should be ordered by healthcare professional seen.

> E2: *"I think what you have got in there is right, but I would order it differently. I would group it by healthcare professional rather than type of visit."*

Home visits (CQ6, DQ7)

Several experts thought that CQ6 was inadequate for capturing home visits, as it was not sufficiently disaggregated. E7 also thought that patients may include home visits in both CQ5 and CQ6 which would lead to double-counting.

> E7: *"Similar to the visit at the practice question, too broad and would not know how to cost… Patients may include home visits in the visits at the practice question."*

Consistent with feedback for DQ6, experts said that DQ7 was inadequate for capturing home visits and could be expanded. E7 also suggested that the formatting of questions DQ6, DQ7 and DQ8 should be consistent with one another.

Remote access care (DQ8)

As this study was conducted prior to the Covid-19 pandemic, feedback was provided in the context of a healthcare system where remote access care was considerably less well-utilised. Remote access care questions were only included in the depth module, as they did not meet the requirements to be included in a core set of resource-use items in the Delphi study (71).

Although two experts thought DQ8 adequately captured information on remote contacts, several thought that DQ8 was ambiguous. They questioned whether the table for DQ8 only related to remote contacts and E1 also queried whether DQ8 only related to GP-based professionals. Experts also thought that telephone and online appointments may carry different costs, so should possibly be separated in ModRUM. However, E11 thought they are not well utilised enough at present to warrant separate questions. E11 also questioned whether unit costs were available for telephone and online contacts, and stated that if they are not available, it will not be possible to cost them, so they should not be collected.

> E1: *"The telephone appointments and the online, might carry different costs so you may not want to group them… The types of healthcare professionals this question relates to, you need to be much more specific, so that there is no double counting."*

E11: *"If you have different unit costs then you may want to separate them, but at the moment I don't think they are prolific enough for this… If you don't have a cost for it, you won't use it, so you shouldn't collect it."*

E5 suggested including NHS 111 as an example if the aim was to capture NHS 111 within question DQ8.

E5: *"NHS 111, would that be captured by DQ8?... I suggest you probably want to put that in as an example. Depending if you are going to use this in Scotland and any of the other devolved nations, whatever their service is."*

Prescribed medications (CQ7, DQ9)

Two experts thought that CQ7 was adequate for capturing prescribed medications and E4 thought it could be costed using an average prescription cost in a core module, but the majority of experts thought CQ7 was inadequate and could not be costed due to the number of prescriptions alone being insufficient information for costing. When asked about whether a question on prescribed medications should be optional and excluded from the core module, all experts agreed.

E9: *"It will give you a descriptive analysis about whether number of medications used has changed. I don't really know how useful it is."*

E7: *"…shouldn't be a core question, if the health economist thinks it is important then they will try to capture it and capture it well."*

Most experts thought it was adequate to collect just name and duration for prescribed medications in a depth module, but E7 thought the opposite.

E3*: "…just knowing the type of medication and how long they took it for is sufficient."*

E7: *"Too minimal. Would require lots of assumptions. Different health economists may make different assumptions."*

Two experts suggested that the question prior to the medications table was unnecessary as it meant the same question was asked twice in different formats which adds an unnecessary validation check. The first question asked, 'how many medications have been prescribed for you?'. In the table, respondents were asked to provide the 'prescribed medication name' and the 'number of days you used the medication' for each medication prescribed.

E2: *"Why are you asking the same question twice. You ask the number of medications and then you have the number down the side anyway."*

There was a variety of comments on the table in DQ9. E5 thought that it would be difficult to complete and to analyse. Several experts said that the third column should ask for the 'number prescribed' not the 'number used', as the number of medications prescribed is the important question for costing purposes. While E2 thought 'number of days' could be captured, several others thought it may be problematic.

E5: *"…DQ9 is an example of something that would, and I'm sorry to say, be horrific to complete, data entry, code, cost and analyse."*

E9: *"This is all about adherence and compliance. Whether they take it or not for the full time is irrelevant, as we just want the cost and whether they take it is a clinical question."*

E4: *"I don't know if I would use number of days here… as I'm not sure how accurate that information would be, but I would probably ask the number of times they have had the prescription or something like that."*

E5 suggested several alterations to avoid the patient and analyst burden of free-text medication names, including grouping and prespecified medications or including guidance notes to state that certain medications should not be included. E6 agreed that cheaper medications could be grouped and pre-specified, and more detail could be captured for high-cost medications.

E5: *"Possibly giving instructions to not worry about some things. Such as aspirin, paracetamol when compared to expensive drugs for specific conditions"*

E6: *"You could group those together because you are going to get names of everything and then just do a weighted cost of pain medications for example. But if there is a specific high cost drug, so in asthma that would be important because severe asthma patients are on a lot high cost stuff and that is where you would want to be specific."*

Guidance notes were provided for DQ9 and stated the number of days respondents should report if they took the medication daily or weekly for the recall period. However, all feedback on the guidance notes was negative, with experts stating that it was too complicated. Instead of the guidance notes, E6 suggested patients could be asked the number of days or whether they took the medication continuously. E9

suggested that patients could instead indicate whether they were taking the medication throughout the entire recall period (ongoing), or whether they were only taking the medication for a short duration. If patients indicated the latter, they could then be asked to report the number of days the medication was used for.

> E7: *"Additional guidance notes, I think that is hard, patients won't use that."*
>
> E9: *"I would have a simple column, 'ongoing medication'. And then 'short duration', 'specify the duration'."*

5.4.3.3 Omitted and overlapping items

Experts were asked to report whether they thought any important NHS resources had been omitted from ModRUM. Suggested missing items included chiropodist, dentist (two experts), district nurse, equipment and adaptations (two experts), level of inpatient care (two experts), optician (two experts), out-of-hours care, social care, transfers and walk-in centres. Experts were also asked specifically about whether they thought items were missed from the core or depth modules. Experts did not report any missing resource items for the depth module, but several experts thought that primary care consultations conducted over the telephone should be included in the core module. E5 thought big procedures and E6 thought out-of-hours care should be added to the core.

> E1: *"I would veer towards incorporating other types of primary care consultations in the core module, even if it was just telephone, I think then you will capture the 3 most common types."*
>
> E5: *"…maybe something around big procedures because they could be main cost generating events."*

Five experts did not think that there were overlapping questions within ModRUM, other than those discussed previously under individual items. However, E11 thought there may be overlap between A&E visits and inpatients stays, where patients are admitted via A&E.

> E11: ***"****The only one is A&E and then referral to a hospital bed. There may be a few complications there. We used to try and do this with arrows in the questionnaire and ask them where they went afterwards. But then it got complicated because there was an overlap in inpatient stays and overnight stays in A&E."*

5.4.4 Research team meeting

The research team met on two occasions (July 11th and 24th, 2019) to discuss the feedback provided by experts. The research team agreed on changes to be implemented, issues to explore with patients and where no changes were required based on the feedback. ModRUM was redrafted and the research team reviewed changes at one final meeting (August 14th, 2019). A summary of the main themes and actions agreed at the meetings are provided in **Table 5.3**.

Table 5.3 Main themes and subsequent actions

Theme	Action
Themes and actions emerging from general feedback	
The introduction was generally adequate.	No changes were required.
In general, the order of questions was adequate, but user flexibility could be considered.	No changes made as allowing users to change question order would reduce consistency in the implementation of ModRUM
Response options were mostly well-received.	No changes were required.
Section headings may be helpful for respondents.	Usefulness of section headings was subsequently informed by patients (**Chapter 6**).
The continuation tables are too long and are likely to confuse respondents.	A free-text box replaced continuation tables in the depth module.
A user guide was needed to describe how to implement ModRUM and cost ModRUM data.	A user guide was developed for ModRUM **(Chapter 7).**
Themes and actions emerging from item-specific feedback	
Questions on A&E, outpatient, day case, inpatient and GP were adequate for capturing detail and applying costs in a core module.	No changes were required.
The A&E question was adequate for capturing detail and applying costs in a depth module.	No changes were required.
Opinions differed on the adequacy and level of detail captured in core questions on other healthcare professionals at the GP practice and home.	No changes made to the core questions, but depth questions adapted to allow for more detail to be captured.

Table 5.3 continued

Theme	Action
Opinions differed on the adequacy and level of detail captured in depth questions on paramedic, outpatient, day case, inpatient and prescribed medication.	More detail captured for the paramedic question and in the outpatient, day case and inpatient tables. 'Number of prescriptions' captured instead of 'number of days' for medications.
The core prescribed medication question was inadequate and prescribed medications should not be included in the core module.	The question on prescribed medications was removed from the core module.
Questions on other healthcare professionals at the GP practice and home and remote access care were inadequate for capturing detail and applying costs in a depth module.	NHS 111 was added as an example.
Non-hospital-based questions should be ordered by healthcare professional rather than mode.	Non-hospital-based care were ordered by healthcare professional.
'Paramedic', 'outpatient' and 'health clinic' may be problematic terms.	'Paramedic' changed to 'ambulance service', 'health clinic' changed to 'health centre' and understanding of 'Outpatient' was explored with patients (**Chapter 6**).
Formatting of non-hospital-based depth questions should be consistent.	The formatting of non-hospital-based questions depth questions was made more consistent.
Themes and actions emerging from feedback on omitted or overlapping items	
Several items were felt to be missing in the modules.	No changes were made. Many items were beyond the scope of the modules or already captured within the questions.
No overlap between questions.	No changes were required.

5.4.4.1 Changes implemented

Ordering and formatting changes

Experts suggested that primary and community care questions should be grouped by healthcare professional rather than mode of appointment due to potential double-counting. The research team agreed, and the core and depth questions were reordered by healthcare professional. Several experts said that it was unclear what question the table under DQ8 (online and telephone appointments) related to. The table in DQ8 was removed and the formatting was made more consistent with other questions to alleviate ambiguity.

Changes to how ModRUM is implemented

To ensure the questions are relevant to each population ModRUM is used in, the research team agreed that while generic examples of healthcare resources are included in the questions, the user can adapt them to trial-specific examples. For example, in question 5 and 6 of the core module (**A3.3**), the revised version included square brackets around ‘nurse or physiotherapist’ to indicate that they can be updated to trial-specific examples. Although this reduces standardisation, it should encourage uptake of ModRUM and ensure that examples are applicable to each study population.

Several experts said that a user guide should be created which explains how ModRUM should be implemented, costed and analysed. The research team agreed, and development of the ModRUM User Guide is described in **Chapter 7**. The research team also agreed that as ModRUM evolved through subsequent research studies, consideration would be given to potentially having a level of depth between core and depth questions. For example, for inpatients stays, this could involve capturing ‘reason’ in addition to the core module detail, but not ‘department’ and ‘tests or procedures’, which are included in the depth module. Based on the feedback, it was also agreed that users could include individual depth items rather than choosing between the core and core and depth modules in their entirety. This change makes the inclusion of depth questions less prescriptive and more flexible, which could encourage uptake.

More detail captured in ModRUM

For DQ2 (paramedic care), it was suggested that it could be split into “see and treat” and “see, and treat and convey”. The unit costs in the NHS reference costs (2017/18) were compared, with the unit cost for ‘see and treat’ at £192 and for ‘see, treat and convey’ at £252 (164). As more detail was captured in ModRUM for resources where the cost difference is lower, the research team decided to separate paramedic care into these two categories.

For DQ3 (outpatients), while several experts thought that ‘main test or procedure performed’ was unnecessary to collect for costing, the research team agreed to keep it in for the version that was tested with patients, as more expensive procedures are increasingly being performed in outpatient clinics. During the patient pilot (**Chapter 8**), the additional time it takes to cost more detailed information was considered. One expert said that ‘main’ should be removed from ‘main test or procedure’, as all tests and procedures are important for costing. The research team agreed to remove ‘main’ and trial the question with patients, to see what they include. The research team also agreed to add ‘if applicable’ to this column, as a test or procedure will not be performed at all outpatient appointments. Both suggestions were also included in DQ4 (day case stays) and DQ5 (inpatient stays). The research team also agreed to add a ‘reason for visit’ column, so that the outpatient table corresponds with the day case and inpatient tables. While this goes against feedback suggesting that the amount of free-text data collected should be minimised, the depth questions are designed to capture all relevant information for estimating costs more precisely.

There were mixed opinions on CQ5 (GP practice appointments) and CQ6 (home visits). While all experts agreed that the ‘doctor (GP)’ part was acceptable, many stated that having one option for non-GP healthcare professionals was inadequate. Once the rationale for grouping other healthcare professionals was explained and it was suggested that this question could be expanded upon in the depth module, most experts agreed that CQ5 and CQ6 could remain the same in the core module. The revised depth questions have options for prespecifying healthcare professionals most relevant for the study and an ‘other’ option where respondents can also report any other types of healthcare professionals seen.

Less detail captured in ModRUM

As most experts thought the CQ7 on prescribed medications was inadequate and prescribed medications could be captured in an optional question only, the research team agreed that prescribed medications should be omitted from the core module. The research team felt comfortable omitting prescribed medications from the core module as questions on prescribed medications did not make it through the second round of the Delphi consensus study (71). An item on prescribed medications was added during the final selection meeting, as the group thought that prescribed medications would be relevant to participants in most trials, but agreed more research was required to decide whether it should remain in the core module or form a separate module (71).

As there was clear preference from experts to remove the continuation tables, the research team agreed to exclude them from the revised depth module and replace them with a free-text box at the end of the depth module. Other alterations included adding a heading to and increasing the size of tick boxes, and increasing the line spacing within and between questions.

Alternative detail captured in ModRUM

The research team agreed with experts and the duplicated question asking the number of medications used was removed from DQ9 (prescribed medications). The question was replaced with a binary question asking whether medications had been used. This will allow the analyst to tell whether a blank medications table is due to missing data or no medication use. The research team also agreed to ask for the 'number of prescriptions' as opposed to the 'number of days'.

Terminology changes

The research team agreed with experts who suggested that 'paramedic' was not a suitable lay term. 'Paramedic care' was changed to 'ambulance service'. Patient understanding of 'ambulance service' was subsequently explored with patients during patient interviews (**Chapter 6**). When asked about the terminology used in CQ5 'at a GP surgery or health clinic' one expert suggested 'health centre' would be more relevant than 'health clinic'. The research team agreed to change the terminology, with the hope that 'health centre' would encourage respondents to

include other resources such as ‘walk-in-centre’ and ‘urgent care centre’ which some experts reported as missing from ModRUM.

5.4.4.2 No changes implemented

No changes due to the existing questions being adequate

The following questions were left unchanged as most experts thought the questions were adequately capturing what they purported to measure and were suitable for costing purposes: CQ1 (A&E), CQ2 (outpatients), CQ3 (day cases) and CQ4 (inpatients).

No changes as changes would be informed by patients

Experts suggested that some of the terms in these questions, including ‘outpatient’ and ‘day case’, may be problematic for patients. It was agreed that if during patient interviews (**Chapter 6**), patients appeared to struggle with terms in ModRUM, alternative terms would be provided by experts or examples would be included. For CQ4, one expert suggested collecting total nights instead of nights per stay. It was also agreed that if patients appeared to struggle with CQ4, using total nights as an alternative would be explored.

Although experts suggested that telephone appointments could be included in the core module, the research team agreed to keep them out of the core module as they did not make it through the final stage of the Delphi study. However, as the Covid-19 pandemic meant that substantially more appointments took place remotely, telephone/online questions were added prior to the patient pilot study (**Chapter 8**). As it was suggested that home visits are not very prevalent, it was planned that the proportion they contributed to total costs would be estimated and their inclusion in the core module reconsidered during the patient pilot study. However, as the pilot was undertaken during the Covid-19 pandemic, when home visits were less likely, deciding on whether home visits should be omitted did not seem sensible, so this was not undertaken. Although it was suggested that free text on ‘big procedures’ could be included for inpatient stays, this went against the design principles for the core module, where to keep it concise, no questions which require free-text responses were included. In the pilot study with patients (**Chapter 8**), I explored the impact on cost estimates of including more detail from the depth module.

While the research team agreed not to make changes to the introduction as most experts were happy with it, the team agreed that understanding of ‘used as an NHS patient’ would be explored with patients in interviews (**Chapter 6**). The research team also agreed to explore several of the other suggestions within patient interviews including whether numbers should be omitted from the tables to avoid extra effort of remembering resources in order and whether patients would like section headings to differentiate resources.

No changes as the suggestions did not meet the objectives of ModRUM

While several missing items were suggested by experts, ModRUM was not altered as the suggested items were either beyond the scope of ModRUM healthcare modules (e.g. social care) or could be captured within questions already included in ModRUM (e.g. district nurse in other healthcare professional home visits).

Several experts suggested that more adaptability should be permitted in the introduction; however, the research team believed that the adaptability within the introduction was sufficient, while not jeopardising the ability of creating a standardised and validated introduction. Free adaptation of the introduction by the user will be discouraged.

Several experts suggested the order of questions should be altered, so the most pertinent items to each trial can be collected first. The order of questions was guided by questionnaire design principles, to include more salient, less frequently used items first. Adaptation of the question order will not be recommended to ensure ModRUM is used consistently.

5.5 Discussion

5.5.1 Main findings

Extensive feedback on ModRUM was provided during semi-structured qualitative interviews with ten experienced health economists. While there was disparity of opinion between health economists, their feedback provided evidence for the content validity of ModRUM; they confirmed that the core module items, with the omission of prescribed medications, clearly and concisely captured all the key NHS resources,

with depth questions allowing the flexibility to capture more detail when relevant to specific trials. Face validity was confirmed as health economists stated that ModRUM measures the desired qualities of a patient-reported RUM. While there were differences of opinion on how precise estimates would be, health economists agreed that the questions would generate answers that are suitable for valuation purposes within economic evaluations.

5.5.2 Strengths and weaknesses

The purposeful sampling strategy ensured that health economists with considerable, recent, and varied experience of conducting economic evaluations alongside UK-based RCTs were interviewed. Credible conclusions could be drawn as experts were independent and well-informed to provide feedback on ModRUM and make comment on its validity (11, 102). This study provides evidence of content validity as the quantity of detailed feedback allowed me to explore a range of contrasting and comparable opinions. This study also complemented the Delphi study that preceded this work in terms of the assessment of content validity (71). In the Delphi study, content validity was established with respect to what items to include, while in this study assessment of the relevance, clarity and conciseness of items (content validity) was made in the context of a formulated RUM. In additional interviews, I would have been able to probe further into the feedback generated from earlier interviews and could have allowed alternative ideas to emerge; however, as the development of a measure is not a single-stage process, further feedback was sought and refinement to ModRUM made in subsequent research studies (**Chapter 6**, **Chapter 7** and **Chapter 8**).

Despite some concern that telephone interviews may not offer the same engagement and rapport as face-to-face interviews (102, 165), I chose to conduct interviews via telephone for pragmatic reasons, as it allowed me to reach health economists located around the UK. I found that engagement from experts was high, both in terms of the percentage of experts agreeing to take part and within the interviews themselves. The average duration of interviews was longer than anticipated. Experts also often used ‘we’ when providing feedback, suggesting their feedback related to their research team or health economists more generally. Sending ModRUM prior to the interviews proved useful, as experts had the

opportunity to review the questions in advance. One expert stated that they had shared ModRUM and solicited feedback on it from their research team prior to their interview. Framing follow-up questions on previous answers allowed me to gain a deeper understanding of the answers provided, and also demonstrated to the expert that I had a genuine interest in their feedback and helped to build rapport during the interview (165).

Audio-recording of interviews avoided distractions from extensive note taking during the interview; it also meant that the opinions of experts could be recorded and cited accurately to support any conclusions drawn from the data (102). The semi-structured nature of the interview meant that pre-defined content was covered to ensure that I could make statements about the face and content validity of ModRUM, and allowed the interview to divert from the scripted questions when new ideas emerged. The responsive interviewing technique and concurrent interviews, transcription and coding allowed the research design to continually be improved to generate rich and detailed data (102).

Feedback from experts was compared and contrasted, allowing the development of common themes and identification of divergent opinions. In research team meetings, we were able to decipher which feedback related to personal preferences (e.g. response option tick boxes), which deviated from the aims of ModRUM (e.g. adaptable question order), which could be explored with patients (e.g. terminology) and which related to problematic areas of ModRUM that required adaption (e.g. core prescribed medication question). In the latter case, common themes on problematic areas of ModRUM allowed me to be confident that alterations implemented were justified and made with a solid evidence base to ensure the validity of ModRUM.

In comparison to health-related quality-of-life measurement, where the EQ-5D is a widely accepted measure for use in economic evaluations (23), recommendations on the measurement of resource use are not well-defined and there is no widely accepted standard. In this study, I found that parts of ModRUM could be adapted to rectify issues, but it was clear that a fixed, off the shelf measure may be too restrictive for health economists measuring resource use. For example, tensions emerged between the use of open and closed questions, and the level of detail

needed to ensure accuracy of cost estimates while minimising patient burden. Feedback provided informed the decision to increase adaptability (e.g. allowing individual depth questions to be added to the core) and reinforced the need for a modular questionnaire that allows for adaptable granularity of resources collected.

5.5.3 Comparison to existing literature

The development of existing RUMs is described in detail in **Chapter 3**. The review highlighted that psychometric testing, including assessment of the content and face validity, of existing RUMs is sparse. Where it is provided, detail is limited, which inhibits the ability to make comparisons between the methods used to assess face and content validity in this study, with methods for face and content validation of existing RUMs.

Health measurement guidelines state that judgement on content validity should be made by experts (12). For the development of ModRUM, experienced health economists were considered as experts who were best placed to inform item generation (71) and review the prototype of ModRUM in interviews, where they could comment on the appropriateness of items and suggest alternative ideas (166). This deviates from patient-reported outcome development where patients can be considered experts who should judge the relevance of items for their patient population (12). It also deviates from existing RUMs where, most often, evidence on the validity has been reported as the measures have been developed with input from healthcare professionals, as opposed to health economists (75, 136, 151). However, as many RUMs are condition-specific, clinicians can be considered experts who are in the best position to identify and comment on the range of healthcare resources used in a specific population. For ModRUM, which is a generic RUM, I considered experienced health economists to be the most appropriate experts, as they have considerable experience of collecting resource-use data in trials of a wide range of conditions and an understanding of what level of detail is required for costing purposes in economic evaluations.

Several authors report that there is evidence for the face (75, 136, 151, 167) and/or content validity (167-171) of their RUM. Where experts were involved in the validation, it was unclear for many RUMs whether experts were independent to the

development team, which would avoid bias in the judgement (11). In addition to judgment by healthcare professionals, several RUMs used alternative methods to provide evidence for face and/or content validity. For the CSSRI-EU, judgement on face validity was made in focus groups with healthcare professionals, care workers, informal carers, and service users (75). Face validity of the REFLUX questionnaire was assessed by reviewing questionnaire results graphically (172). Evidence for the content validity of the PDMED questionnaire is reported as the content was created from an expert panel review and pilot (168). For the CESAR questionnaire, the authors state identification of resources from similar studies provides evidence for content validity (170), and in their report, the authors also stated clinical expert input in resource-use item selection (145). The content validity of a RUM to capture resource-use data from patients with lower back pain was judged by experienced healthcare professionals who commented on missing, ambiguous, or unnecessary items and concepts (136). Cognitive interviewing of patients to assess content validity was only reported for the Cardiff Cardiac Ablation PROM, which primarily captures health outcomes but also includes several healthcare use and productivity questions (173).

While the majority of RUMs referred to were developed for a specific condition, the ACQP shares similarities with ModRUM as it includes a standardised set of questions that can be adapted and used to capture resource-use data from patients in economic evaluations across a range of conditions (76). Evidence for the content validity of the ACQP is reported as a checklist of resource-use items for the ACQP was formulated from existing RUMs provided by health economists (76, 171).

Although they did not state that face or content validity were being assessed, literature on the development of other RUMs report involvement of experts in the development process (121, 141, 146). For example, the Osteoarthritis Costs and Consequences Questionnaire was developed based on existing measures and with health economics, public health and clinical expert input (146), while a RUM designed to capture data from patients with diabetes was developed based on existing measures and with clinical expert input (121).

5.5.4 Implications for this research

To ensure the longevity of ModRUM, it is important to make sure that it is future proofed to be able to adapt to changes in healthcare services. For example, in January 2019, the NHS indicated a plan to increase access to telephone and online consultations within five years (174). During this thesis, this plan was accelerated due to the Covid-19 pandemic. As a result, the feedback on the relevance of remote access care items in this study should be interpreted with the caveat that it was provided pre-Covid-19 and it is likely that experts would consider remote access care much more important to capture in the core module based on healthcare utilisation trends following the onset of Covid-19. In **Chapter 9**, further consideration is given to future changes to healthcare provision and how ModRUM can remain relevant despite these changes.

Based on feedback provided by experts, ModRUM was revised. Experts suggested several aspects of ModRUM that required input from patients including some of the terminology. Based on expert suggestions, the topic guide for use in patient interviews was revised. Comments made on how to implement ModRUM, suggested that ModRUM would benefit from an accompanying user guide. The guide was subsequently developed (**Chapter 7**). Future research studies allowed for the revised version of ModRUM to be tested with patients (**Chapter 6, Chapter 8**) and health economists (**Chapter 7**).

5.6 Conclusion

The aim of this work was to generate evidence for the content and face validity of ModRUM. Experts were asked to comment on the relevance of items, clarity and conciseness of questions, omission of relevant items and whether ModRUM appears to measure what it is intended to measure. This study provides evidence for the content and face validity of ModRUM as experts either said that questions were adequate for their purpose and could be costed, or they provided suggestions on how to make the questions fit for purpose. Where necessary, questions were amended to ensure that ModRUM is relevant and adequate for capturing healthcare resource-use data in economic evaluations alongside RCTs. A study designed to test the acceptability and content validity of the revised version of ModRUM in cognitive interviews with patients is described in the next chapter.

Chapter 6 Cognitive interviews with patients

6.1 Chapter overview

This chapter reports on cognitive interviews with patients to test the content validity and acceptability of ModRUM. A purposeful sampling strategy with maximum variation was used to recruit patients from primary care to participate in “think-aloud” interviews with retrospective probing. Participants verbalised their thought processes as they completed ModRUM, which allowed errors (issues with completion) and struggles to be identified. Participants were asked follow-up and probing questions to investigate errors, struggles, clarity and acceptability.

Interviews were audio-recorded and transcribed verbatim. Transcripts were scored by three independent raters (KG, JT, SH) to identify errors in comprehension, recall, judgement and response; and struggles. Independent raters met to agree on final scores. I analysed interview transcripts qualitatively using techniques of constant comparison, to identify common themes and ideas for improvement. Data collection, analysis and revisions to ModRUM were performed concurrently.

This chapter includes background information describing the rationale and objectives of this study. The methods section outlines the study design; site identification; patient identification, sampling and recruitment; data collection procedures and the data analysis plan. The results section outlines site and patient recruitment, participant characteristics, scoring results and qualitative results. The discussion includes a summary of the main findings, strengths and limitations of this study, how this study compares to existing literature, implications for research practice and this research, and unanswered questions and future research. A brief conclusion is provided at the end of the chapter. The research described in this chapter has been published in BMC Health Services Research (175).

6.2 Background and objectives

When developing a new instrument, it is important to first demonstrate that it has content and face validity, as defined in **Chapter 2**, before testing other measurement properties (11). Evidence for the content and face validity of ModRUM has been

provided as the items were informed by health economists (71), and in qualitative interviews with health economic experts (**Chapter 5**), feedback was garnered on the relevance, clarity and conciseness of items, and whether relevant items had been omitted. Cognitive interviewing with patients provides the final opportunity to test the content validity of a new instrument before testing other measurement properties in a larger quantitative study (85). Interviews can be used to assess patients' comprehension and to evaluate the comprehensiveness of a questionnaire to ensure that questions are capturing the information they are expected to capture (85). The assessment of content validity also involves generating evidence to state that the instructions, recall period and response options are relevant, and patients find them comprehensible and acceptable (85, 166). To ensure that interviews provide evidence of content validity, it is important that they are conducted with a diverse sample of patients, with representation from different groups within the population, that covers patients who may have unique answers or perspectives and patients who are likely to experience more difficulty understanding or completing the measure (80, 85).

During the development of a new instrument it is also important to ensure that the instrument is well-received by respondents, as the acceptability of the questionnaire can impact upon data completion and response rates (81). The acceptability of an instrument can initially be assessed during cognitive interviews, by asking participants about their experience of completing the questionnaire (81). It can also be reassessed in a larger study where data completion, response rates and completion time can be analysed (81).

The aim of this study was to assess the content validity and acceptability of ModRUM with a wide range of patients. The think-aloud exercise and retrospective verbal probing allowed me to gain insight into patient understanding of the questions and to assess whether ModRUM measured what it was intended to measure. Retrospective verbal probing also allowed the elicitation of patient opinion on the layout and acceptability of ModRUM.

6.3 Methods

6.3.1 Study design

In this study patients participated in cognitive interviews, which encompassed a think-aloud exercise and retrospective verbal probing questions. The think-aloud exercise, which is an established technique for assessing the content validity of outcome measures (85), involved participants completing ModRUM while verbalising their thought processes (88). Retrospective verbal probing followed the think-aloud exercise to probe on areas where patients experienced issues and on pre-specified areas of interest which were outlined in a topic guide (88).

6.3.2 Site identification and recruitment

The study was advertised by the Clinical Research Network (CRN) West of England to PCOs, with a range of deprivation levels (The index of multiple deprivation for 2010 ranges from one, which is the most deprived, to 10 which is the least deprived), within the Bristol, North Somerset or South Gloucestershire (BNSSG) regions. PCOs were sent a Research Information Sheet for Practices (RISP) and asked to express interest via the CRN. The CRN provided me with details of interested PCOs and I contacted each PCO directly and provided them with further information about the study including an Organisation Information Document, Schedule of Events, the research protocol, and details of payment for participation. To increase the socioeconomic diversity of patients recruited, I aimed to recruit at least one PCO from the lowest two deciles of deprivation. For PCOs who wished to take part, a site initiation visit was arranged. At the visit I went through a site initiation checklist and asked the lead GP or Practice Manager to confirm the PCO's capability and capacity to take part in the research.

6.3.3 Patient identification

Each PCO was provided with two options for inviting patients to express an interest in taking part in the study. One option was for a receptionist to briefly introduce the study to adult patients checking-in at reception for an appointment. The second option was for a clinician to introduce the study to adult patients at the end of consultations. Where patients expressed an interest, they were provided with a patient information sheet (**A4.1**, **Appendix 4**) and a reply form (**A4.2**). The reply form contained questions on patient contact details and patient characteristics (sex, age

group, ethnic group, number of long-term conditions, age on leaving full time education, number of primary and secondary healthcare contacts in the last three months). This information was collected to aid recruitment and sampling decisions, which are described in detail in the section below. Patients were asked to return the reply form to the receptionist if they agreed to be contacted by me about participating in the study.

6.3.4 Patient sampling and recruitment

Sampling strategy

Patients were recruited to participate subject to the following eligibility criteria:

- aged 18 or over;
- able to understand written and verbal English;
- registered at one of the participating practices and
- capable of giving informed consent.

A purposeful sampling strategy was used to ensure 'information-rich' patients, who were active users of healthcare services, that based on their characteristics may have had different interpretations of content and purpose of the questions, were recruited (80, 100). With limited resources, this approach prioritises patients who are particularly experienced and/or knowledgeable about the subject of interest (100). To reflect the wide range of patients that could complete ModRUM in an RCT context, maximum variation sampling was employed to increase the diversity of patients recruited based on several patient characteristics including sex, age group, ethnic group, number of long-term conditions and age on leaving full time education, and use of primary and secondary healthcare services (104). Maximum variation sampling can allow for the identification of similarities and differences, with respect to the topic of interest, for a heterogenous group of individuals (100). Recruitment at later stages was informed by the characteristics of previously recruited participants, to maximise variation and to ensure that the sample included representation for each group and characteristic that was considered important for a generic RUM.

Priority was initially given to patients from groups that were considered harder to reach (male, non-white ethnic groups, lower age on leaving full time education) and patients who had used secondary healthcare in the last 3 months. Later, non-white

ethnic groups were prioritised as they were not represented among initial participants. It also became apparent that issues were more likely to be experienced, and subsequently elaborated on, if a participant had used a healthcare resource. Therefore, a theoretical approach was taken in later sampling to maximise representation from participants who had used healthcare resources, and who may also be more similar to RCT participants with respect to higher resource consumption (90). To achieve this, priority was given to patients who had used primary or secondary care in the last 3 months or had long-term conditions.

Sample size

Although there is no definitive guideline for the sample size required for cognitive interviewing studies, for patient comprehensibility of items of health measurement scales a sample size of seven to 10 participants has been suggested (79). However, it has also been acknowledged that sample sizes should be informed by the complexity of the measure and should include patients with characteristics similar to the population of interest (85). These points were taken into consideration when recruiting patients. Concurrent interviews and analysis allowed me to identify when 'data saturation' was reached, whereby additional interviews would not have identified any major new issues that had not already been considered and would not have resulted in further changes to ModRUM (97). At this point no further patients were invited to participate in the study.

Recruitment

I contacted patients identified as eligible to participate by telephone. During this contact, patient involvement in the study was described and the patient had the opportunity to ask questions. For patients who agreed to participate, a date, time and location for a face-to-face interview was organised. Based on the patient's preference, interviews took place at their home, at their PCO or another mutually agreeable location. At the beginning of the interviews, I explained the aims of the study to the patient, explained what they would be expected to do, asked whether they gave permission for the interview to be recorded and took written informed consent (**A4.3** [consent form]; **A4.4** [topic guide]).

6.3.5 Data collection

Face-to-face interviews were conducted as they allow the interviewer and participant to build rapport prior to undertaking the research activity (102). Good rapport with interview participants is considered important as participants are likely to be more open in their responses, which can lead to richer data (176). The think-aloud exercise was chosen for completion of ModRUM as it is considered an advantageous process where minimal input is required from the interviewer during questionnaire completion, which allows issues to be revealed while minimising interviewer-imposed bias (88). To help participants become familiar with the think-aloud process, I started each interview with an established think-aloud warm-up exercise, which involved visualising and counting windows in their home (105). Participants then completed the think-aloud exercise by answering (on paper) the core module (**Figure 6.10**) or the core module plus depth questions (**A4.5**) of ModRUM while verbalising their thought processes. Initially participants completed the core module only. Once I was confident that all core module specific issues had been identified and the research team agreed, subsequent participants completed the core module with depth questions version only. All questions referred to healthcare use in the last three months. I remained silent throughout unless the participant stopped verbalising their thoughts, in which case I prompted the participant to continue speaking aloud.

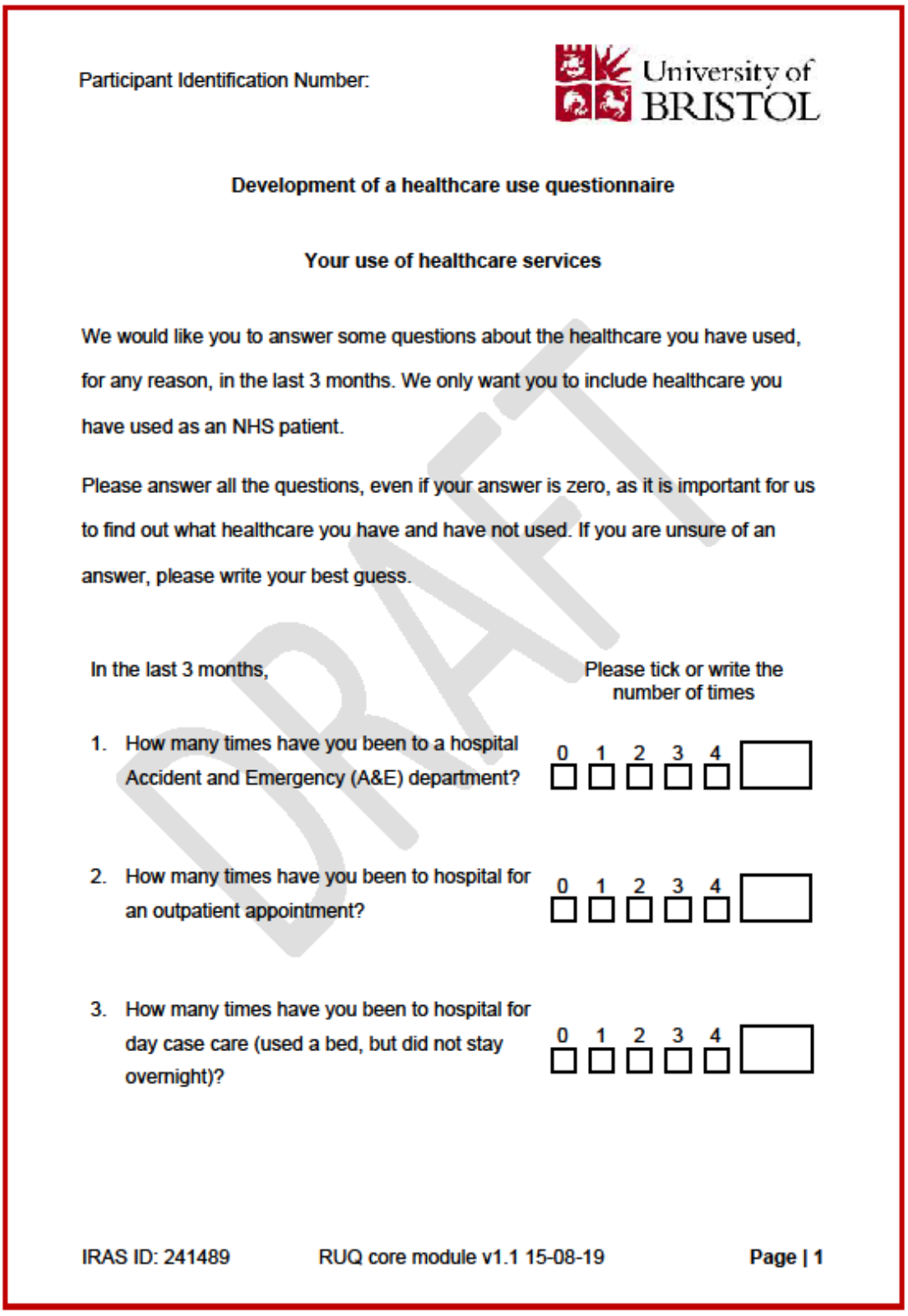
The think-aloud exercise was followed by a semi-structured interview, whereby I asked the participant questions to clarify any issues that occurred and on prespecified areas of interest, including content, terminology, ease of completion and acceptability of ModRUM, from a topic guide (**A4.4**). Participants were probed on their understanding of specific terminology, including ‘outpatient’ and ‘day case’, as health economists indicated that these may be problematic for patients (**Chapter 5**). Interviews were conducted in rounds, as this allowed issues that were identified, to be discussed with the research team and appropriate revisions agreed and implemented, prior to testing the revised version of ModRUM in further interviews (105). As mentioned in **Chapter 4**, a professional design company developed a logo and enhanced the aesthetics of ModRUM, these adaptations were added for testing in the third round of interviews.

Interviews were audio-recorded, and audio-recordings were transcribed verbatim. Transcription was performed by an external provider (UK Transcription). Data were managed and/or analysed in Microsoft Excel, Stata 17 and NVivo 11.

6.3.6 Data analysis

Analysis was performed concurrently to data collection, with iterative testing conducted so that findings from interviews led to adaptations to ModRUM and the topic guide before the adaptions were tested in further interviews (88).

Figure 6.1 ModRUM core module - pre-cognitive interviewing



Participant Identification Number:

University of BRISTOL

Development of a healthcare use questionnaire

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used, for any reason, in the last 3 months. We only want you to include healthcare you have used as an NHS patient.

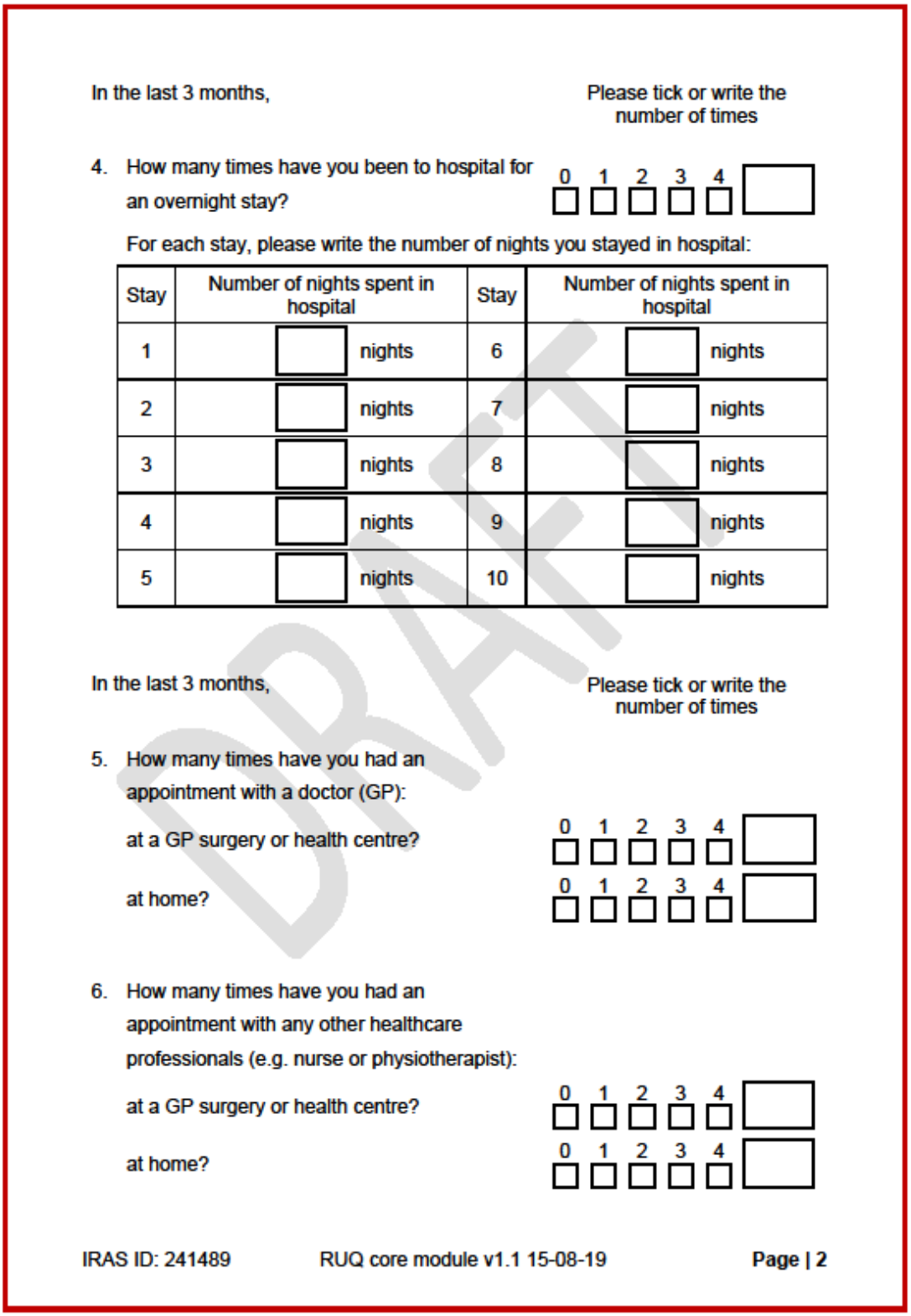
Please answer all the questions, even if your answer is zero, as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

In the last 3 months,	Please tick or write the number of times
1. How many times have you been to a hospital Accident and Emergency (A&E) department?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐
2. How many times have you been to hospital for an outpatient appointment?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐
3. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

IRAS ID: 241489 RUQ core module v1.1 15-08-19 **Page | 1**

DRAFT

Figure 6.1 continued



In the last 3 months,

Please tick or write the number of times

4. How many times have you been to hospital for an overnight stay?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

For each stay, please write the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital	Stay	Number of nights spent in hospital
1	☐ nights	6	☐ nights
2	☐ nights	7	☐ nights
3	☐ nights	8	☐ nights
4	☐ nights	9	☐ nights
5	☐ nights	10	☐ nights

In the last 3 months,

Please tick or write the number of times

5. How many times have you had an appointment with a doctor (GP):

at a GP surgery or health centre? 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

at home? 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

6. How many times have you had an appointment with any other healthcare professionals (e.g. nurse or physiotherapist):

at a GP surgery or health centre? 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

at home? 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

IRAS ID: 241489 RUQ core module v1.1 15-08-19 **Page | 2**

DRAFT

6.3.6.1 Data scoring

As recommended for the assessment of content validity within cognitive interviews, transcripts were analysed using a standardised classification scheme to identify response problems (85). The standardised classification scheme is based on the Tourangeau's survey response model which breaks down the cognitive process of answering questions into four actions (89). The actions include:

1. **comprehension** of the question in the intended way
2. **retrieval** of the appropriate information from memory
3. **judgement** of how the information should be used to answer the question and
4. **formatting** the information into a valid response (89).

For each participant, I provided SH and JT with the transcript, the participant-completed ModRUM and a core or core with depth scoring form (**A4.6**). Acting as raters we each independently scored responses by reporting for each question whether errors in comprehension, retrieval, judgement, or response occurred. Raters also noted when participants appeared to struggle with a question but were able to reach the correct answer (e.g. re-reading the question) (106). Error classifications were made in a hierarchical order; for example, if a comprehension error was scored then no further errors or a struggle were identified. Following independent rating, raters met on four occasions (following the first two interviews and at the end of each round) to compare scores. Where scoring differences arose, raters discussed the scoring until they reached consensus on a final score.

Inter-rater agreement was assessed using Gwet's agreement coefficient (107, 108). Gwet's agreement coefficient was chosen over other statistics for the following reasons: there were more than two raters; not all raters scored all of the questions; there were more than two scoring categories (uncategorised, comprehension, retrieval, formatting, response, struggle, no error); and the importance of disagreements in scores varied (107). User-defined weights were applied to account for these differences (**Table 6.1)**. A score of one was given for agreements and zero for complete disagreements (error versus no error). A score of 0.5 was given to disagreements in scoring category where both raters identified whether the answer was correct or incorrect (e.g. if one rater scored a judgement error and the other rater scored a response error, the raters disagreed but ultimately they both identified an incorrect response). A score of 0.25 was given to disagreements in scoring

category where both raters identified a problem, but did not agree on whether the answer was correct or incorrect (e.g. a response error would be an incorrect answer, whereas a struggle would indicate a problem but a correct answer). Strength of agreement of Gwet's agreement coefficient is usually considered to be substantial/excellent for agreement coefficient scores above 0.6 and almost perfect for scores above 0.8 (79).

Table 6.1 User-defined weights for inter-rater agreement

Scoring category	Uncategorised	Comprehension	Retrieval	Judgement	Response	Struggle	No error
Uncategorised	1						
Comprehension	0.5	1					
Retrieval	0.5	0.5	1				
Judgement	0.5	0.5	0.5	1			
Response	0.5	0.5	0.5	0.5	1		
Struggle	0.25	0.25	0.25	0.25	0.25	1	
No error	0	0	0	0	0	0.5	1

6.3.6.2 Qualitative data coding

I uploaded and analysed transcripts in a qualitative data analysis software package (NVivo 11) (177). NVivo was chosen to enhance the management and analysis of the qualitative data, including to generate an extensive coding structure, and to ease the process of applying codes to corresponding data.

Qualitative analysis focused on participants' reactions and views on the wording and presentation of ModRUM. Techniques of constant comparison were utilised to continually compare participants' comments on aspects of the RUM design, to develop key patterns and themes from participant responses and to enhance understanding of key issues experienced during RUM completion (97, 98). This approach was chosen as the iterative nature can enhance research rigour (178).

Prior to coding each transcript, the entire transcript was read alongside the participants' completed version of ModRUM and the accompanying scoring form, which allowed me to become fully immersed in the participant's experience (90). Analysis involved line-by-line coding of transcripts, with data organised into themes and assigned a representative code. A coding structure was developed and applied to all interview transcripts, with categories and sub-categories continually updated for new data. Categories represented higher-level concepts including more abstract themes (e.g. difficulty recalling prescribed medications) and sub-categories represented lower-level concepts which were more specific (e.g. difficulty due to medications being prescribed for different durations) (90). The lower-level categories represented the properties of each higher-level concept and showed the variation in themes under each category (90).The analysis of transcripts from earlier interviews focused on the identification of initial concepts. As interviews progressed the process evolved so that new data were compared and contrasted to concepts already defined and where applicable refinements were made to concepts.

Analytic accounts, which *"describe the data in context and make connections between categories and sub-categories"* were formulated in batches, at the end of each round, using results from data scoring, categories and subcategories, notes made while formulating these categories and corresponding quotes (95) (pg.105). Writing analytic accounts formed part of the analysis as it provided the opportunity to link issues identified in the scoring to the qualitative data emerging from the think aloud interviews. This linkage allowed me to provide context to the issues that occurred. Illustrative quotations were selected to provide support for the interpretations made (95).

6.3.6.3 Revisions to ModRUM

Following each round, I prepared a report, which included recruitment information, participant characteristics, responses to ModRUM, and the analytic account. These reports were then presented to the research team at group meetings. During the meetings, I presented my findings and made suggestions on whether ModRUM should be revised. I also explained any modifications that I thought should be made to the follow-up and probing questions in the topic guide. Following discussions, the

research team agreed upon which changes should be implemented prior to the next round of interviews or whether interviews should be concluded.

6.4 Results

6.4.1 Sites

Of seven sites who expressed interest in the study, five were recruited to participate. Of the two sites that did not participate, one site requested additional reimbursement information and subsequently declined the invitation due to lack of capacity. The other site did not respond to follow-up emails regarding their expression of interest. **Table 6.2** provides a description of each participating site.

Table 6.2 Description and characteristics of participating sites

Site	Location	Deprivation decile[1]	Registered persons[1]	Ethnicity estimate[12]
One	Bristol	2 (second most deprived decile)	19,903	3.1% mixed, 4.9% asian, 3.2% black
Two	Bristol	5 (fifth most deprived decile)	15,660	2.9% mixed, 2.5% asian, 3.0% black
Three	Bristol	4 (fourth most deprived decile)	8,250	1.9% mixed, 1.3% asian, 1.1% other non-white ethnic groups
Four	North Somerset	10 (least deprived decile)	18,504	1.1% mixed, 1.4% other non-white ethnic groups
Five	South Gloucestershire	10 (least deprived decile)	14,590	1.2% mixed, 1.3% asian

1: Figures were obtained between October and November 2019 (179),

2: Estimated proportion of non-white ethnic groups, ethnic groups with a proportion <1% are added to other non-white ethnic groups.

6.4.2 Patients and participants

A GP introduced the study and handed out information sheets and reply forms at four sites. At site two a receptionist introduced the study and if interested, patients took the forms into their consultation to discuss with the GP.

Of 58 reply forms returned, five patients did not tick the box confirming that they were willing to be contacted about the study and two telephone numbers provided were not in use. Of the remaining 51 patients, 39 patients, who responded to a telephone call within the timeframe of the study, were invited to participate, and 29 patients agreed to take part. Nine interviews were subsequently cancelled by the patient due to illness or the interview time no longer being convenient, leaving 20 patients who participated in the study.

A summary of participant characteristics is provided in **Table 6.3**. Half of participants were female. There was representation in all age categories; however, participation was slightly higher for patients aged between 31 and 55 (55 percent of all participants). No patients from non-white ethnic groups expressed an interest in taking part in the study; as a result, all participants were from the white ethnic group. Age on leaving full-time education was 18 or under for 11 participants. One quarter of participants were recruited from the site with the lowest deprivation score. Over half of participants (n=12) indicated that they had a long-term condition. Most participants had used primary care (n=19), while just over half had used secondary care (n=13) in the last three months.

Table 6.3 Summary of participant characteristics

Characteristic	Category	n	(%)
Sex	male	10	(50)
	female	10	(50)
	prefer not to say	0	(0)
Age group	18-30	1	(5)
	31-45	5	(25)
	46-55	6	(30)
	56-65	3	(15)
	66-75	3	(15)
	76 or over	2	(10)
	prefer not to say	0	(0)
Ethnic group	white	20	(100)
Age on leaving full-time education	16 or under	4	(20)
	17 or 18	7	(35)
	19 or over	9	(45)
	prefer not to say	0	(0)
Long-term conditions	none	8	(40)
	one	6	(30)
	more than one	4	(20)
	prefer not to say	2	(10)
Used primary care in last 3 months	none	1	(5)
	one or two times	12	(60)
	three or four times	4	(20)
	more than four times	3	(15)
Used secondary care in last 3 months	none	7	(35)
	one or two times	8	(40)
	three or four times	3	(15)
	more than four times	2	(10)
Site	One	5	(25)
	Two	3	(15)
	Three	4	(20)
	Four	2	(10)
	Five	6	(30)

A summary of participant healthcare use, as captured in ModRUM, is provided in **Table 6.4**. On average, participants reported 2.8 (SD 1.5) appointments with the GP at the GP practice over the last 3 months. All participants who completed the prescribed medication question reported that they had picked up or received prescribed medications. Outpatient appointments were the next most frequently reported healthcare used with 14 (70%) participants reporting appointments and an average of 1.4 appointments (SD 1.5). Few (less than 3) or no participants reported using the following healthcare: hospital day case and inpatient stays; GP, nurse and other healthcare professional home visits; care from the ambulance service and nurse telephone/online appointments.

Table 6.4 Healthcare use in the last 3 months

	Number who answered the question (%)		Number who used the service (%)		Mean number of contacts (SD)	
Questions in the core module and the core module plus depth questions (n=20)						
A&E	20	(100)	8	(40)	0.4	(0.5)
Outpatient	20	(100)	14	(70)	1.4	(1.5)
Day case	20	(100)	1	(5)	0.1	(0.2)
Inpatient	20	(100)	2	(10)	0.1	(0.3)
GP at the surgery	20	(100)	20	(100)	2.8	(1.5)
GP at home	18	(90)	0	(0)	0.0	(0.0)
Other HCP[1] at the surgery	19	(95)	7	(37)	0.6	(1.0)
Other HCP at home	16	(80)	1	(6)	0.1	(0.3)
Questions in the core module plus depth questions only (n=12)						
Ambulance see, treat & convey	12	(100)	0	(0)	0.0	(0.0)
Ambulance see & treat	12	(100)	0	(0)	0.0	(0.0)
GP phone/online	12	(100)	6	(50)	0.7	(0.8)
Nurse at the surgery	12	(100)	9	(75)	0.9	(0.7)
Nurse at home	12	(100)	1	(8)	0.2	(0.6)
Nurse phone/online	12	(100)	0	(0)	0.0	(0.0)
Other HCP phone/online	12	(100)	4	(33)	1.9	(5.7)
Prescribed medications	11	(92)	11	(100)	-	-

[1]HCP= healthcare professional

6.4.3 Interviews

I conducted interviews face-to-face at participants' homes (n=14), offices (n=3) or PCOs (n=3) between December 5th, 2019 and March 13th, 2020. Interviews, including the think-aloud exercise and semi-structured follow-up interview, lasted 26 minutes on average (range: 10–54 minutes). There were three rounds of interviews, with revisions to ModRUM made between and during rounds. In round one, five participants were interviewed and completed the core module. In round two, participants completed the core module until I was satisfied with feedback and modifications made to the core module and moved onto testing the core module plus depth questions. In round two, three participants completed the core module and six participants completed the core module plus depth questions. In round three, six participants completed the core module plus depth questions.

6.4.4 Scoring results

Agreement between the three independent raters, as estimated using Gwet's agreement coefficient, was 92% which represents almost perfect agreement.

Errors (labelled issues hereafter) and struggles are presented by participant in **Table 6.5**. Of the questions scored, 15 percent were judged as incorrect responses, 5 percent were judged as correct responses with a struggle and 80 percent were judged to be correct responses with no struggle. The number of issues as a percentage of questions scored increased from round one (core, 16%) to round two (core or core and depth, 22%) and then decreased in round three (core and depth, 9%). The most common issue was in comprehension, with 28 issues across 6 participants, but most comprehension issues (20 issues) were made by one participant. Of the four processes involved in answering questions, six participants experienced issues in comprehension, five participants experienced issues in retrieval, six participants experienced issues in judgement and seven participants experienced issues in response.

A summary of specific issues and struggles is provided in **Table 6.6**. The most common specific issue/struggle experienced was uncertainty of what healthcare to include. The number of issues/struggles scored by healthcare item is presented in **Table 6.7**. For the core module, issues were scored for the outpatient question most

often. For the core module plus depth questions, issues were most often scored for prescribed medication name and number of prescriptions. Changes made to ModRUM between and during rounds are presented in **Table 6.8**. Changes made included: (1) adding emphasis, mostly through bold font, to important information in the instructions (e.g. to only include healthcare used as an NHS patient) and questions (including the type of appointment e.g. outpatient appointment); (2) clarifying that family/dependant healthcare should not be included; (3) improvements to questionnaire design (e.g. including the recall period at the start of each question); (4) additional detail added to aid comprehension (e.g. examples added for different types of outpatient appointments).

Following the summary tables below, the results of the coding and the qualitative analysis are presented in further detail. Verbatim quotes are included to provide greater clarity and understanding of participants' reactions and views of ModRUM, and to illustrate themes that were identified in the data.

Table 6.5 Issues and struggles, by participant

Round	ModRUM version	Participant	Number of parts scored	Issues[1]						Struggles
				C	Ret	J	Res	U	Total	
1	Core module	1	9	-	-	-	2	1	3	1
		2	9	-	-	-	2	-	2	-
		3	9	2	-	-	-	-	2	2
		4	9	-	-	-	-	-	-	2
		5	9	-	-	-	-	-	-	-
		Total	*45*	*2*	*-*	*-*	*4*	*1*	*7*	*5*
2	Core module	6	9	-	4	1	-	-	5	2
		7	9	-	1	-	1	-	2	-
		8	9	1	1	-	3	-	5	1
		Total	*27*	*1*	*6*	*1*	*4*	*-*	*12*	*3*
	Core module and depth questions	9	31	-	-	2	-	-	2	1
		10	31	1	-	2	3	-	6	-
		11	31	-	3	-	-	-	3	-
		12	31	-	-	-	-	-	-	2
		13	31	20	-	-	-	-	20	1
		14	31	-	-	2	1	-	3	5
		Total	*186*	*21*	*3*	*6*	*4*	*-*	*34*	*9*
3	Core module and depth questions	15	34	-	-	-	-	-	-	2
		16	34	1	-	3	-	-	4	1
		17	34	-	-	1	-	-	1	1
		18	34	-	-	-	-	-	-	-
		19	34	-	-	-	-	-	-	-
		20	34	3	4	-	6	-	13	-
		Total	*204*	*4*	*4*	*4*	*6*	*-*	*18*	*4*
All		*Total*		*28*	*13*	*11*	*18*	*1*	*71*	*21*

[1]C=comprehension, Ret=retrieval, J=judgement, Res=response, U=uncategorised issue

Table 6.6 Issues and struggles, by issue/struggle category

Category	Round	Summary of issue/struggle	Number of issues/ struggles	Participant
Comprehension	1	Dependent/family use included	1	3
		'Outpatient' misunderstood	1	3
	2	Dependent/family use included	20	13
		Unintended HCP[1] included	2	8, 10
	3	Examples interpreted as question	3	20
		Frequency instead of number	1	16
Retrieval	2	Incorrect recall period	5	6, 8
		Recalled more in probing	4	7, 11
	3	Could not recall	3	20
		Recalled more in probing	1	20
Judgement	2	HC[2] included under unintended question	3	14, 9
		Included private healthcare	2	10
		HC judged irrelevant	2	6, 9
	3	Included test under clinic	1	17
		Included future appointments	2	16
		HC judged irrelevant	1	16
Response	1	Missing but was recalled	2	1
		Missing as zero use	2	2
	2	Missing but was recalled	3	8
		Missing as zero use	1	7
		Missing as did not see question	3	10
		Used table not as intended	1	14
	3	Missing as did not see the question	6	20
Missing	1	Missing with no indication of issue type	1	1
Struggle	1	Initially missed but returned and completed	2	3
		Uncertain of answer	3	1, 4
	2	Uncertain whether answer required	2	14
		Uncertain of what to include	6	6, 9, 12, 13, 14
		Struggle with recall	1	14
		HC term misunderstood	2	12, 8
		Uncertain if to include dependent/family use	1	6
	3	Uncertain of what to include	2	15, 16
		Hesitation as too similar to last question	1	17
		Uncertain if to include dependent/family use	1	15

[1]HCP=healthcare professional; [2]HC=healthcare

Table 6.7 Issues and struggles, by question

Question	Judged issues						Struggle
	C	Ret	J	Res	U	Total	
Core (n=8)							
A&E	1	0	0	1	0	2	1
Outpatient	1	2	0	2	0	5	1
Day case	0	0	0	0	0	0	2
Inpatient stays	0	1	0	1	0	2	1
Inpatient nights	0	1	0	0	0	1	1
GP at the practice	1	1	0	0	0	2	2
GP at home	0	0	0	2	0	2	0
Other HCP at practice	0	1	1	0	0	2	0
Other HCP at home	0	0	0	2	1	3	0
Total core module	3	6	1	8	1	19	8
Core with depth questions (n=12)							
A&E	1	0	0	0	0	1	1
Ambulance to hospital	0	0	0	0	0	0	1
Ambulance not to hospital	0	0	0	0	0	0	1
Outpatient number	1	1	1	0	0	3	0
Outpatient clinic	2	2	0	0	0	4	0
Outpatient reason	2	0	1	0	0	3	1
Outpatient tests/procedures	2	0	1	0	0	3	2
Day case number	0	0	0	0	0	0	0
Day case department	0	0	0	0	0	0	0
Day case tests/procedures	0	0	0	0	0	0	0
Day case reason	0	0	0	0	0	0	0
Inpatient number	1	0	0	0	0	1	0
Inpatient nights	0	0	0	0	0	0	0
Inpatient department	0	0	0	0	0	0	0
Inpatient tests/procedures	0	0	0	0	0	0	0
Inpatient reason	0	0	0	0	0	0	0
GP at the practice	2	1	1	0	0	4	1
GP at home	0	0	0	0	0	0	1
GP phone/online	1	0	0	0	0	1	0
Nurse HCP at practice	1	0	1	0	0	2	0
Nurse HCP at home	1	0	0	0	0	1	0
Nurse HCP phone/online	1	0	0	0	0	1	0
Other HCP at practice number	1	0	1	1	0	3	1
Other HCP at practice HCP seen	1	0	1	1	0	3	0
Other HCP at practice HCP number	0	0	0	1	0	1	0
Other HCP at home	1	0	0	1	0	2	3
Other HCP at home HCP seen	1	0	0	1	0	2	0
Other HCP at home HCP number	0	0	0	1	0	1	0
Other HCP phone/online	1	0	0	0	0	1	0
Other HCP at phone/online HCP seen	1	0	0	0	0	1	0
Other HCP at phone/online HCP number	0	0	0	0	0	0	0
Medication (yes/no)	1	1	0	1	0	3	0
Medication name	1	2	1	2	0	6	0
Medication number	2	2	0	1	0	5	1
Total core module with depth questions	25	9	8	10	0	52	13
Total	**28**	**15**	**9**	**18**	**1**	**71**	**21**

1: C=comprehension, Ret=retrieval, J=judgement, Res=response, U=uncategorised issue; 2: HCP=healthcare professional

Table 6.8 Changes made between and during interview rounds

	Reason for change	Change made	Time of change
Instructions	Emphasise instructions	Instructions put in a box	Post-round 1
		Important parts of instructions in bold font	Post-round 1
	Minimise missing responses	Message added to check answers at the end of the questions	Post-round 1
		Final text enlarged and put in bold font	Post-round 2
	Clarify that only own healthcare use should be included	'Yourself' added to instructions	Post-round 2
		Separate sentence to not include family/dependent care added	Post P17
Questionnaire design	Minimise missing responses	Added 'a' and 'b' to two-part questions	Post-round 1
		Multi-part questions separated	Post P8
	Reiterate the recall period	'In the last 3 months' added to the start of each question	Post P8
	Improve question navigation	Signposts added to the core module plus depth questions version	Post-round 2
	Improve labelling of response options	'Other' included over large box in response options	Post-round 2
		'How many' included over large box in response options	Post-round 3
Question design	Clarify that each row refers to one stay, appointment or visit	'Stay', 'appointment', 'visit' added before each number in the outpatient, day case and inpatient tables	Post-round 1
	Emphasise distinguishing details	Healthcare resource and mode of similar questions put in bold font	Post-round 2
		Location of healthcare resources put in bold font	Post-round 3
	Avoid confusion over what to include under 'test'	'Test' and 'Reason' columns switched in outpatient, day case and inpatient tables	Post-round 3
	Distinguish examples from questions	Examples in tables included on separate rows	Post-round 3
Outpatients	Aid comprehension of what to include	Examples added	Post-round 1
		'X-ray' added as a second example	Post-round 2
		Lay terminology included in the example	Post-round 3
Day cases	Distinguish from home care	'Day case care' changed to 'day case stay'	Post-round 2
	Distinguish from inpatient stays	'Day case stay' changed to 'day case'	Post round 3
Other healthcare professionals	Distinguish community from hospital physiotherapy	'Community' added before 'physiotherapist' example	Post-round 1
	Clarify to only include NHS resources	NHS added to other HCP	Post-round 2
	Clarify where to include walk-in centre	'Walk-in centre' added to 'GP practice or health centre'	Post-round 2
	Increase clarity of what is required in the table	'Number of times' added to the table	Post-round 2
	Distinguish between online and telephone	'NHS 111' now 'NHS 111 telephone call'	Post-round 3
Prescribed medications	Aid comprehension of what to include	Instructions, to include medications even if specific name unknown, made prominent	Post-round 2
		Eczema cream added as an example	Post-round 3

6.4.4.1 Comprehension

Comprehension issues occurred when participants misunderstood the meaning of a word or phrase. A summary of issues scored under each process is provided in **Table 6.6**. Most comprehension issues were due to participants including their dependant's/family's healthcare use in their response.

> Participant (P)3: *For myself, it would be none. For others, I think at least once. So, would you like- I'm just going to put one.*
>
> P13: *I'm the person responsible for them and it's a process that I'm heavily involved with. I am going to include my interactions on my kids' behalf.*

An attempt to clarify this issue was made following round two, by including 'healthcare you, yourself, have used' in the instructions, which has been suggested in questionnaire design literature to overcome confusion related to 'you' (161). However, in round three, P15 was also uncertain whether to include their dependant's healthcare, so an additional sentence was added to the instructions ('Please do not include any healthcare your family or dependants have used') and no further issues/struggles were scored for this reason.

There was also evidence of some comprehension issues surrounding the terminology used to describe healthcare professionals, with two participants reporting nurse consultations under GP consultations and one participant acknowledging when probed that they did not understand the term 'outpatient' resulting in an incorrect response.

P20 experienced issues with the depth table for outpatient appointments; they interpreted examples that were within table headings as part of the question. Table formatting was revised so that examples were clearly separated from table headings on a separate row.

6.4.4.2 Retrieval

Retrieval-related issues included being unable to recall information and including healthcare use that occurred outside of the recall period. Several participants referred to additional healthcare use during the probing questions that they did not include during the think-aloud exercise. Two participants suggested that this might have been due to the nature of the task and said they felt *"a little bit rushed… a little*

bit under pressure" (P11) while completing ModRUM, during the think-aloud exercise.

P7: *it's actually quite stressful, thinking, when somebody is there*

Several participants said they would have retrieved more information had an interviewer not been present during ModRUM completion.

P17: *put a best guess in because I would have otherwise had to have got up and disturbed the questionnaire*

The recall period was initially specified above each section as opposed to within each question; however, P6 and P8 included appointments outside the recall period. 'In the last 3 months' was added to the start of each question following the interview with P8 and no further problems occurred due to this issue. Some evidence of telescoping, where events outside the recall period are 'telescoped' in, was revealed during probing.

P6: *The last three months… I think it was actually before then, thinking about it.*

P15 said *"It was hard to decipher what events were in the last three month window"*. Several other participants also reported difficulty with this.

Further feedback on the recall period was sought in probing questions. Most participants thought that three months was acceptable and four participants said they could recall healthcare use from six months ago. Participants reported that irregular healthcare was easy to remember.

P15: *it's irregular for me, so it stands out in my mind*

Mixed feedback was provided on the ease of recalling regular healthcare use. P7 and P14 said seeing different healthcare professionals for different reasons makes it harder to remember regular healthcare use.

P7: *The GP is easy because that's long term, so every two weeks is actually quite easy to remember*

P14: *I think if you're a regular visitor to a GP surgery because of ongoing health issues you lose track of how many times… actually trying to remember in a three-month period that's quite hard.*

P4 said *"you always get a letter"* for secondary care appointments, so they are easier to remember, whereas for appointments at the GP practice they would have to *"look in my diary because I wouldn't be able to remember".*

6.4.4.3 Judgement

Issues in judgement occurred when participants deemed relevant healthcare irrelevant, irrelevant healthcare relevant or recorded healthcare use under an unintended question.

Three issues occurred due to healthcare being judged irrelevant; this included P16 who did not include eczema creams under medications and P9 who said they had an electrocardiogram at their cardiology appointment, but recorded 'N/A' under 'Tests or surgical procedures performed'. P9 said the example (removal of a skin lesion) led to them putting 'N/A', as the example *"seems quite severe"* and *"an ECG doesn't affect you… It's not invasive"*. P9 agreed a less invasive example could be included and 'x-ray' was added as an example following round two.

P10 included private physiotherapy under other healthcare professionals as the question did not specify to include NHS healthcare professionals only.

> P10: *'How many times have you had contact with any other healthcare professionals?' Now, they're healthcare professionals. She's a physiotherapist, a proper physiotherapist, that's why I filled it in*

While the instructions at the beginning of ModRUM specified that respondents should 'include healthcare you have used as an NHS patient', this response highlighted that further clarification, to include only include NHS healthcare, was required for questions on other healthcare professionals.

P14 recorded 'advanced nurse practitioner' under GP as opposed to nurse. They were unsure of where to include it, but judged advanced nurse practitioners to be more similar to GPs, with respect to the type of consultations they provide.

> P14: *I would class them more as a type of consultation I'd have with the GP.*

6.4.4.4 Response

Missing answers were the most common response issue and were due to the participant not seeing the question, due to the answer being zero or recalled verbally by the participant but not included in the questionnaire and without evidence of an intentional judgement to omit the information. The number of missed responses due to the answer being zero diminished to zero by round three. Following round one, an instruction to 'answer all the questions, even if your answer is zero' was inserted in bold font and an instruction was added at the end, asking participants to check that they have answered every question.

Four participants had missing answers for the second part of two-part questions, and indicated that they left the questions blank as they had not used the resource. When questioned about changing them to independent questions, P2 said it *"would probably work better"* and P7 said they would probably have been less likely to leave it missing. Two-part questions were separated into independent questions after the interview with P8.

Missing responses due to not seeing the question, appeared to be a feature of the questionnaire pages being stapled together. P20 said the questionnaire *"would be lot better if it was a booklet, when they turn over to the next page, then they know, it's, 'Oh, I've got two sides to fill in'"*.

6.4.4.5 Struggle

Struggles occurred when participants hesitated or displayed uncertainty about answering a question but appeared to reach the correct answer. This included uncertainty around where responses should be written (e.g. whether to write the type of test performed under test or reason for outpatient visit) and what should be included within a question (e.g. does a chiropodist come under another healthcare professional?). P12 struggled with the question on other healthcare professionals at home.

> P12: *That was the only one I wasn't quite sure what you meant… All the others, because you'd given an example, so I wonder whether, if you put an example*

Examples were added to this question (health visitor and NHS chiropodist) prior to round three and no further struggles were scored due to this problem for this question in subsequent interviews.

Several participants appeared to struggle in recalling information, including P4 who provided a guess for the number of GP appointments at the practice because they *"…would have to go to my diary… because I wouldn't be able to remember… so it has to be a guess"*. Guesses were not deemed as issues as participants were advised in the instructions to include their best guess if they were unsure of an answer.

6.4.5 Qualitative results

6.4.5.1 Terminology

Most participants suggested that they understood the term 'outpatient', with several referring to their own experience as outpatients. However, many participants displayed some lack of understanding as to what could occur at an outpatient appointment. Two participants thought A&E visits were included under outpatients, P2 described them as *"one-off"* appointments and P1 said *"you don't go in there for an operation"*. Examples of outpatient appointments were added to ModRUM following round one. Three examples were included as it was evident from the interviews that the examples needed to demonstrate different types of outpatient appointments (such as a consultant visit, diagnostic test and treatment). For example, 'hospital physiotherapy' was included as P2 said *"I don't think people would necessarily see that as an outpatient, because normally physio is like an ongoing thing".*

Following the addition of examples, participants' comprehension of the question appeared to improve.

> P15: *'outpatient', I'm not 100% clear on what that is. I just assumed, by the example you'd given- That sort of defined that, for me.*

In ModRUM, 'day case care' was accompanied by a brief definition ('used a bed, but did not stay overnight'), which many participants said was helpful. While three participants said they did not know the meaning of the term, most other participants described it as a hospital visit where a surgery or procedure is performed. 'Care' was

removed from ‘day case care’ following round two, as two participants thought the question was referring to home care.

Further uncertainty around what to include was expressed for other questions. For example, P12 said they were unsure of what other healthcare professional or healthcare service at home means but suggested that the inclusion of an example could help. Participants also highlighted areas where the same services could be delivered at different locations or double counting could occur. P13 included out-of-hours under GP practice/health centre and A&E as the out-of-hours clinic was usually run from the GP practice; however, on one occasion, due to Christmas, the service was delivered from A&E. P9 included the same visit to a walk-in centre under GP practice/health centre and A&E. To mitigate this issue, ‘walk-in centre’ was added to non-hospital-based questions; for example, for appointments with a GP the revised question asks ‘how many times have you had an appointment with a doctor (GP) at a GP surgery, health centre or walk-in centre?’.

6.4.5.2 Questionnaire instructions and design

Participants that read the instructions thought that they were acceptable and described them as *“straightforward”* (P1) and *“easy to understand”* (P4). P3 did not read the instructions which led to some uncertainty when answering the questions. P3 said they *“just usually scan it and just pick out the key words of what it says sometimes”* when reading instructions in general. Bold fonts were added to key points in the instructions following round one, as when asked if it would be helpful, P3 indicated *“Yes... Because I’ve got learning difficulties. That’s what helps me focus, picking out the key points of things.”*. Bolding of healthcare professionals was added throughout the questions following round two as other participants suggested highlighting important terms throughout.

> P15: *in the context of people trying to do it quite fast, it would be easy to miss a key point, and then your data is not accurate.*

While I tried to limit the amount of text that was included in bold font, to keep the emphasis on important text only, bolding of modes of healthcare (e.g. at home) was deemed necessary and added following P20’s interview. P17 struggled to differentiate how non-hospital based other healthcare professional questions differed

and indicated that the location/mode of appointments could be made *"a bit clearer"* by potentially putting it in a *"bold"* font.

There was mixed feedback on including section headings in the core module only. P5 said it could be useful as *"it clearly defines the two [sections]"*. However, other participants thought it was unnecessary, with P1 stating *"it's laid out like that anyhow"*. For the core module plus depth questions, P13 said it needed more signposting as it was unclear when answering earlier questions what healthcare would be captured in later questions. This resulted in them including midwife under nurse as opposed to other healthcare professional. P13 said: *"If I was doing the survey to do the survey for real, I would have to go back and fix it. It would take extra time, and I would find that quite frustrating."* Following round two, signposts, such as 'Questions 1 to 3 ask about emergency healthcare: A&E and ambulance', were added to each section. In round three, while P17 said *"I didn't feel I needed it"*, other participants found them helpful.

Response options for questions requiring numerical responses included tick boxes for 0-4 and a larger box for more than 4, with the instruction 'Please tick or write a number'. Most participants used the response options as intended and feedback was generally positive; however, there was some confusion about the large box. 'Other' was added above the large box following round two; however, P20 said *"I wouldn't put 'Other'"*, so this was updated to 'How many?' following round three.

One of the aims of participant testing was to identify areas where respondent burden could be minimised, while maintaining sufficient detail for precise estimation of costs. In the depth secondary care questions, tables capture detail on items including 'clinic type', 'reason' and 'tests or surgical procedures'. Participants were asked about combining 'reason' and 'tests or procedures'. As two participants said they would provide less detail and participants were generally positive about providing more information, items were not combined. However, they were reordered with 'tests or procedures' followed by 'reason' to rectify confusion about what to include under 'tests or procedures' when they had already included the test/procedure under 'reason'.

During the interviews, some participants retrieved information on their healthcare use (including diaries, calendars, hospital letters and medications) and others stated they would have retrieved them if an interviewer had not been present. P16 got up twice during the think-aloud to retrieve information, and agreed during probing that some instruction at the beginning, to have the information *"handy… would be a good idea"*. The research team agreed that in a ModRUM user guide, researchers could be advised to include this in a cover letter to accompany ModRUM.

6.4.5.3 Acceptability

Most participants provided positive feedback on the length, content and layout of ModRUM. When probed, most participants did not report difficulty completing the questionnaire, P19 said it was *"quite self-explanatory"* and P11 said they *"didn't find anything confusing"*. P20 had trouble as they *"don't have a very good memory"* due to a stroke. They also found examples unhelpful as *"at my age, thirties, downwards, wouldn't know what that means"*. Most participants who completed the longer version of ModRUM thought the length was acceptable.

> P7: *I was expecting it to be quite long, so that was actually quite easy.*

However, P14 and P17, who completed the core module plus depth questions version of ModRUM, indicated that it may not be acceptable had they used more resources in the past 3 months.

> P14: *I've had two really significant surgical procedures… had the three months fallen either side of those things I would have been writing forever I suppose.*

There was some indication completion of questions was easier when the answer was zero.

> P2: *'How many times have I been to hospital, accident and emergency department in the last three months?' Well, that's easy, I haven't been at all. So, that's a zero.*

Issues were more likely to be noted for resources with high utilisation. For example, while P16 retrieved their medication boxes and recorded six prescribed medications in ModRUM, they did not include two other medications that were subsequently recalled during probing and said they would have only recalled three if they had not been able to access their medication boxes.

6.5 Discussion

6.5.1 Main findings

Think-aloud and follow-up semi-structured interviews with patients recruited from primary care have generated evidence for the content validity and acceptability of ModRUM. Participant responses to questions were generally consistent with what the questions were intended to measure, and 80 percent were judged as being answered correctly without issue or struggle. Most participants reported that the content, length, and layout of ModRUM were acceptable. Issues identified were used to iteratively refine and enhance the comprehensibility and acceptability of ModRUM.

6.5.2 Strengths and weaknesses of this study

I was able to recruit 20 patients with a broad range of characteristics using a purposeful sampling strategy. While I attempted to recruit patients from harder to reach groups, I was more successful with some than others. I was unable to recruit any patients from non-white ethnic groups and, as a result, this study does not provide evidence for the acceptability and content validity of ModRUM for patients in these groups. Participation from younger patients (aged 18 to 30) was also lower than other age categories; few reply forms were returned from this group which may have been due to younger patients attending GP practices less frequently. Allowing the sampling frame to adapt as it became evident that some groups would be under-represented, for example by identifying groups for whom I had fewer reply forms returned, may have allowed me to reach these groups using an alternative approach (180). However, changing the approach would have required further planning and ethical approval, which is a time-consuming process. While I was able to recruit PCOs with a range of deprivation scores, differences in feedback based on geographical variations in terminology across the country may not have been identified as all PCOs were located in the South-West of England.

Interviews concluded approximately two weeks prior to the first UK coronavirus lockdown. Nearing the end of the recruitment period, some patients who had initially expressed an interest in taking part, decided not to proceed to an interview, with their decision not to participate in a face-to-face interview likely impacted by Covid-19. While recruitment would not have been possible for much longer due to coronavirus, the decision to end recruitment was informed by less new information emerging from

later interviews. I found that most findings which emerged in later interviews were consistent with findings from earlier interviews, as such, I decided that it was not appropriate to continue with interviews, as it was unlikely that new themes would emerge.

Concurrent interviews and analysis allowed for revisions to ModRUM and the topic guide to be made iteratively and in response to interview findings, with revisions tested in subsequent interviews. For difficulties that were revealed during the think-aloud exercise, follow-up questions allowed a greater understanding of these issues and in some cases, how they could be improved or rectified. Verbal probing complemented the think-aloud task as it allowed me to gain valuable feedback on the design, formatting and length of ModRUM, from which decisions about refinements were made. The three-month recall period allowed identification of issues at all stages of the cognitive process of answering questions, whereas a longer period may have limited the identification of issues in the judgement and response processes as issues may have been more likely to occur in the retrieval process. If a retrieval issue occurs, judgement and response processes are not scored.

Interviewer-imposed bias was minimised as the think-aloud activity required minimal input during questionnaire completion. However, as two participants reported feeling under pressure while completing ModRUM while thinking-aloud in front of an interviewer, some interviewer-imposed bias was evident. Some participants also said they would have referred to their diaries/medical notes had an interviewer not been present. The artificial aspect of completing ModRUM during a think-aloud exercise does not reflect how it would be implemented in RCTs. As a result, it is unlikely that issues, which were likely to be a result of the artificial aspect of this study (i.e. thinking-aloud), would occur in usual administration of ModRUM, whether that be a self-complete or interviewer-administered version. The issues and struggles that occurred due to confusion over whether to include family/dependant resource-use may not have been an issue in an RCT setting, as participants are more likely to be aware that they are required to report their own resource use only. However, in an RCT, RUMs are often nested within a large booklet of outcome measures which may mean participants become fatigued and are more likely to rush through the RUM,

leading to missing or incorrect responses, suggesting the instructions are still likely to be useful.

When expressing interest in taking part, patients were asked to provide details on the frequency of primary and secondary care use in the last three months. The aim of this was to recruit patients who were active healthcare users. Responses to ModRUM indicated that some resources were well utilised by participants, yet others, including inpatient and day case stays, home visits and care from the ambulance service, were used by few or no participants. More issues were scored for resources that were used by more participants, such as outpatient appointments and prescribed medications. The four stages of the cognitive process of answering questions were less easy to differentiate when participants had not used a resource; responses were often immediate, as they generally did not require much thought. In these instances, participants appeared to comprehend the questions and no issues were observed. For questions where low or no usage was observed in this participant group, issues that may be experienced by patients who use these resources may not have been revealed.

To date, evidence of item identification, validation and piloting of existing RUMs is limited (9). As a result, researchers should consider potential measurement error when collecting and interpreting data captured via RUMs developed without thorough testing. Consideration should also be given to whether measurement error is likely to be systematic (leading to decreased precision of cost estimates in both arms) or differential (leading to biased estimates of incremental costs between treatment arms). For example, while most types of issues patients experienced could have been hypothesised based on existing literature, such as telescoping and leaving responses blank when the answer was zero, inclusion of family/dependant resource-use was an issue that was not anticipated. To our knowledge, this issue has not been addressed in other RUMs. Assuming this issue was not a feature of the artificial (non-RCT) setting of this study, if existing RUM introductions are not explicit about excluding family healthcare use, they could introduce systematic measurement error from an overestimation of costs which would bias and decrease the precision of cost estimates in both arms of the trial.

Independent raters were able to score transcripts and scores between raters were mostly consistent, with almost perfect agreement. Where disagreements were found, raters were able to agree on a final score. There were some areas in scoring that were more ambiguous than others, for example, raters agreed to score issues where P14 included ‘advanced nurse practitioner’ under ‘GP’ instead of ‘nurse’. While this may be considered an issue, it may also have been the correct place to include it from a costing perspective as the participant reported that the advanced nurse practitioner was providing consultations more similar to a GP than a practice nurse. After consideration of this issue, the research team agreed that the wording should remain the same, with no elaboration on which types of nurse to include. This was decided as adding type (e.g. advanced nurse practitioner) adds more jargon, with the potential for added comprehension issues.

In scoring transcripts, independent raters were reliant upon the participant revealing where they had made an issue either within the think-aloud exercise or during probing; however, the true figure may not have been recalled, as both methods require the participant to recall information that reveals the true figure. For outcome measurement of subjective constructs, revealing the true value is limited to think-aloud with probing. However, for resource-use this study could have been extended to allow for triangulation between patient-report (RUM), patient-report (interview transcript) and administrative data (e.g. healthcare records), which would require the assumption that such records are accurate.

As each transcript was scored by two or three raters independently, the opportunity for bias in scoring issues and struggles was minimised. I performed the qualitative analysis independently and formulated an analytic account which was shared with the research team for consideration prior to making changes to ModRUM. To minimise the potential for bias in the qualitative analysis, another member of the research team could have independently coded the transcripts to check that the interpretation and categorisation of data was consistent.

6.5.3 Comparison with the existing literature

ModRUM differs from existing RUMs as it is a concise, generic RUM designed to collect resource-use data in a standardised and consistent manner, with a modular design allowing flexibility to ensure it is relevant for a wide range of RCTs. While validation is rarely undertaken for self-report RUMs used in RCTs (9), the development process of ModRUM has been extensive and is ongoing. The CSRI is the most commonly used RUM (74); however, it differs from ModRUM as it was designed for interviewer-administration to capture resources related to mental health conditions and has subsequently been adapted many times inhibiting standardisation in implementation (49).

The systematic review described in **Chapter 3** highlighted that where validation is undertaken, criterion validity is the most commonly assessed psychometric property for RUMs, with most authors estimating the agreement between RUM data and data reported in electronic records. Despite RUMs being developed or adapted for almost every new trial, there is limited evidence of patient interviewing in their development. Where there is evidence, patients have informed item identification (130, 151) and item formulation (147), and similar to this study in interviews patients have also identified problems encountered during RUM completion (121, 157).

Ruof et al (2004) performed in-depth interviews with patients to identify the appropriate level of aggregation (detail captured) for each item included in their RUM for capturing resource use from patients with rheumatoid arthritis (147). Each patient was shown various levels of aggregation for each item and feedback indicated that there was a general preference for higher levels of aggregation (less detail), particularly for medications and diagnostics, from which a preliminary version of the RUM was developed (147).

Chernyak et al (2012) developed a RUM for capturing resource-use data from patients with diabetes mellitus (121). The RUM was tested in cognitive interviews with 43 patients, of which 19 tested a self-administered version and 24 tested an interviewer-administered version (121). Like this study, the researchers undertook behaviour coding of interview transcripts to identify problems experienced by participants in answering questions; however, coding was not based upon the four

processes of answering questions used in this study (121). The authors identified comprehension issues, including distinguishing between healthcare professionals, and information retrieval problems, such as inability to explain how they arrived at an answer (121). To enhance comprehension of the terminology, the wording was revised, additional instructions added and questions were restructured (e.g. splitting questions) (121).

'Think-aloud' interviews with concurrent and retrospective verbal probing were also used in a study to refine an adapted version of the CSRI, designed for proxy-completion by bereaved relatives of cancer patients (157). Nine interviews were conducted and revealed comprehension issues due to difficulty in deciding what to include under each group of healthcare services and retrieval issues related to uncertainty around the number of contacts, and information on hospital wards and medications (157). Subsequent refinements included asking for less detail and providing examples (157).

6.5.4 Implications for research practice

While self-reported RUMs have frequently been used as the primary source of resource-use data within RCTs, the advent of routinely collected electronic data has provided an alternative method (49). However, as accessing routine data can be costly and time consuming, at present self-report RUMs continues to be a commonly used method for collecting resource-use data (49).

In line with existing literature, I found that the ability to recall resource-use data was impacted by many factors including the recall period, resource type and frequency of use (56). A three-month recall period was used in this study as it is a commonly used recall period in RCTs (115). However, as economic data collection points are often determined by outcome measurement time points in RCTs, researchers will be able to adapt the ModRUM recall period as appropriate for their study. For alternative recall periods, researchers should consider implications to the acceptability for respondents and the validity of responses as the accuracy of results may diminish as the recall period increases (125). Similarly, as other adaptations are permitted, researchers should consider the impact on acceptability and validity of these

adaptations and whether further testing should be performed prior to implementation of ModRUM.

6.5.5 Implications for this research

Combining the findings from this chapter, interviews with health economists (**Chapter 5**) and the Delphi consensus survey (71) has allowed me to be confident that ModRUM captures what it is intended to measure and that no important items have been omitted, i.e. the content and face validity of ModRUM have been established. The remaining measurement properties, including feasibility, acceptability (completion and response rates), construct validity and criterion validity, could be tested in a larger quantitative patient pilot study (11). Piloting with patients allowed ModRUM to be tested in a setting that was more akin to how it would be administered in RCTs (i.e. without an interviewer present). Details on the patient pilot study are provided in **Chapter 8**.

6.5.6 Unanswered questions and future research

As I was unable to recruit any patients from non-white ethnic groups, future qualitative research focusing on harder to reach groups would allow statements to be made about the acceptability and validity of ModRUM for these groups. An alternative patient identification and recruitment strategy may be needed to reach these patients. Further research could also consider how issues may vary according to health conditions. As ModRUM is a generic RUM, patients were recruited without consideration to the conditions they may have; however, one participant revealed that their memory was impaired due to a stroke. Patient groups to prioritise for future research could include patients with impaired cognition (e.g. dementia), where development of a simplified or proxy-version may be more suitable; patients with high resource-usage (e.g. end of life care), where there may be increased burden of completing the questions; or patients who are likely to have used resources that less well-utilised in this study (e.g., ambulance or day case visit). While patients were recruited via primary care in this study, an alternative approach may be required to identify patients from these groups, such as through care homes or hospitals. In addition, ModRUM was developed for use in one healthcare system (UK NHS); for international use of ModRUM, translation and content validation would be required prior to implementation.

Many of the struggles identified related to uncertainty surrounding what to include within each question. Within a paper-based RUM the ability to provide an exhaustive list of the examples is limited without significantly increasing the length and potentially hindering the comprehensibility of the questionnaire. An online version of ModRUM could enhance comprehension by including drop down lists of examples and/or definitions provided when respondents hover over terms and improve usability by allowing an unlimited number of responses for commonly used healthcare (e.g. medications). An online version could also increase uptake of ModRUM in RCTs, as data collection methods move away from paper-based to online formats.

6.6 Conclusion

Cognitive interviewing with patients has been identified as an important step in the questionnaire development process, as it allows the assessment of patient comprehension and comprehensiveness of a new questionnaire. To date, in the development of RUMs, cognitive interviewing with patients has very rarely been undertaken. Using think-aloud interviews with retrospective verbal probing, patients' issues in the comprehension, recall, judgement and response processes of answering questions and struggles were identified, and feedback on the acceptability of ModRUM was gained. An iterative process allowed issues to be identified and rectified before testing revisions in further rounds of interviews. This study has provided evidence for the content validity and acceptability of the content, length, and layout of ModRUM in patients recruited from a primary care setting. The next Chapter reports on a study where health economists piloted the adaptation process of ModRUM and provided feedback in an online survey.

Chapter 7 Piloting the adaptation process of ModRUM with health economists

7.1 Chapter overview

This chapter reports on a pilot study with health economists to assess the feasibility and suitability of ModRUM for collecting resource-use data in RCTs. A User Guide was developed for ModRUM and health economists were advised to use the guide to adapt ModRUM for a recently funded NIHR HTA or PHR grant they were working on. Following their adaptations, health economists completed an online survey where they provided feedback on the ease of adaptation, clarity of the User Guide, likely ease of applying unit cost data to resource-use data generated by ModRUM, and likelihood and potential barriers to using ModRUM for capturing resource-use data in future trials.

This chapter begins with background information explaining the rationale and objectives of this study. Next, the study design, sampling and recruitment of health economists, data collection procedures and data analysis plan are outlined in the methods section. The results section provides information on the identification and recruitment of respondents, respondent and trial characteristics, and a summary of respondent feedback. The discussion section includes a summary of the main findings, strengths and limitations of this study, how this study compares to existing literature, implications for research practice and this research, and unanswered questions and future research. A brief summary concludes the chapter.

7.2 Background

To maintain consistency in the implementation of ModRUM, which will allow comparison across trials, the questions in ModRUM are standardised. However, to encourage uptake and ensure ModRUM is relevant to each trial, some aspects of ModRUM are adaptable. To date, flexibility, within an NHS perspective, is provided by the depth questions, which allow end-users to capture more details for items where increased precision is needed, such as for items that are key cost drivers or are highly utilised resources. Flexibility is also permitted in the recall period, type of resource use (all-cause versus condition-specific), examples provided to illustrate

healthcare that is included within each resource item and number of responses allowed in tables.

Despite the availability of a generic RUM (the ACQP) for capturing resource-use data (76), which was reviewed in more detail in **Chapter 3**, it was not well-utilised and there remains a precedent for developing and adapting RUMs on a trial-by-trial basis (9). In health measurement scale development, expert opinion is typically sought to make a judgement on whether the items adequately reflect the construct of interest (content validity) (79). As ModRUM has adaptable elements, an approach which deviates from health measurement scales where adaptation is strongly discouraged, I thought it would also be valuable to test and gain opinions on the adaptation process of ModRUM with potential future users.

To increase the likelihood of ModRUM being well-received and well-utilised by health economists for capturing resource-use data, the aim of this study was to assess the feasibility and suitability of ModRUM for collecting resource-use data in RCTs by piloting the adaptation process with health economists. I added this study to my research plans once it became clear in March 2020 that the patient pilot study (**Chapter 8**) would be delayed due to Covid-19. This study allowed me to identify the extent to which health economists would adapt ModRUM, develop and obtain feedback on a ModRUM User Guide, and gain feedback on the adaption process and suitability of ModRUM.

7.3 Methods

7.3.1 Study design

In this study, health economists who were working on recently funded NIHR HTA or PHR grants, piloted the adaptation process of ModRUM and reported feedback on the process in an online survey. Ethical approval for this study was provided by the University of Bristol Faculty of Health Sciences Research Ethics Committee in April 2020 (reference number: 102602).

7.3.2 Health economist sampling and recruitment

Health economists who were co-applicants on NIHR HTA and PHR grants, which were waiting to start as of March 2020 or research in progress which started between April 2019 and March 2020, were purposefully selected using details on recently funded NIHR grants which are publicly available in the NIHR Journals Library (181). These grants were selected as I thought the invitation to participate in this study would coincide with the health economist developing a RUM for the identified grant. For each grant, the following information was extracted and recorded in a Microsoft Excel spreadsheet:

- the grant name,
- the grant stage (in progress/waiting to start),
- the funding programme (HTA/PHR),
- whether the grant was an RCT,
- whether there was a health economic component,
- the lead health economist's name and institution,
- the disease area/condition, and
- the setting.

Grants were excluded if they:

- were not an RCT,
- did not have a health economics component,
- the grant included solely child participants, and/or
- the lead health economist was a member of the research team or a colleague at the University of Bristol

Following exclusion of ineligible grants, health economists were invited to participate in the study by email (**A5.1**, **Appendix 5**). An information sheet was attached to the email and included background information and a description of what participation would involve (**A5.2**).

Invitations were sent in two rounds. The first round included health economists who were named as the lead health economist on one grant. The second round included health economists who were named on multiple grants. The characteristics of included grants from the first round were used to identify gaps in the characteristics

with respect to funding programme, disease area/condition and setting. Grants in the second round were purposely selected to prioritise identified gaps.

Health economists had the option to take part by themselves, in conjunction with a colleague or to forward the email on to a colleague also working on the grant. The latter two options were included as, to my knowledge, it is often the more junior health economists who take on the responsibility of developing a RUM, with oversight from a senior health economist. Health economists who did not respond to the initial email invitation were sent one follow-up email approximately two weeks following the first email.

Two follow-up emails were sent to health economists who responded to the first email expressing their intention to take part, but did not subsequently participate. To encourage participation, in later follow-up emails, the average time it had taken to adapt ModRUM among respondents was included. Online and telephone contacts were also offered to meet with potential respondents to go through the survey with them.

7.3.3 Data collection

Health economists who agreed to participate were sent copies of the ModRUM core module (**Figure 7.1**) and ModRUM core module with depth questions (**A5.3**). I drafted the first version of the ModRUM User Guide, with the aim of creating a clear and concise set of instructions, that were also comprehensive enough to ensure that users adapt ModRUM in the intended way. The research team commented on this draft and the revised draft was given to respondents, who were asked to use the guide to prepare ModRUM for their recently funded trial. The ModRUM User Guide was a six-page document that included:

- a quick start guide, which concisely covered the important aspects of all adaptation procedures,
- an introduction, which described the purpose and outlined the design of ModRUM,
- a section on implementing ModRUM, which described the content and purpose of the core and depth questions, the current scope of ModRUM (NHS

funded healthcare only), and adaptable aspects (examples, recall period, all-cause/condition-specific resource use),
- details on validation of ModRUM; including past, present and plans for future validation, and
- formatting tips for adaptation of ModRUM in Microsoft Word

Several details were brief, with the intention to add more detail to the User Guide in the future; including information such as how to analyse ModRUM data, registration and copyright of ModRUM, and how to cite use of ModRUM (**A5.4**).

Once respondents had adapted ModRUM, they were asked to return their version of ModRUM to me by email and complete a brief online survey. The survey was prepared using Online Surveys (109). The survey included questions on ease of adaptation, suitability of ModRUM for capturing resource use in RCTs, clarity of the ModRUM User Guide and future use of ModRUM. Respondents were asked closed questions, with responses captured as 'yes' or 'no', or on 5-point Likert scales (110). Rather than following up closed questions with further closed questions, open questions where respondents could provide a rationale for their responses were used (111). Completion of the online survey by respondents was deemed to represent informed consent to participate.

7.3.4 Analysis

Respondents' trial-specific versions of ModRUM were reviewed and I summarised the adaptations which had been made. The quantitative survey data were downloaded to Microsoft Excel and analysed with simple descriptive statistics. Survey responses were also uploaded to NVivo 12, where categories were coded from closed questions (e.g. 'ease of adaptation'), sub-categories were informed by responses to closed questions (e.g. 'fairly easy') and more specific sub-categories emerged from the feedback (e.g. the 'user guide' made adaptation fairly easy). Feedback between respondents was compared and contrasted, to identify whether any common themes arose (100). I convened and led a research team meeting, where I presented the main findings to the team and how I thought they should be used to inform the development of ModRUM and the User Guide. We discussed the findings and agreed on how feedback should be used to inform further development.

Figure 7.1 ModRUM core module - pre-health economist pilot

ModRUM

Modular Resource-Use Measure

Participant Identification Number:

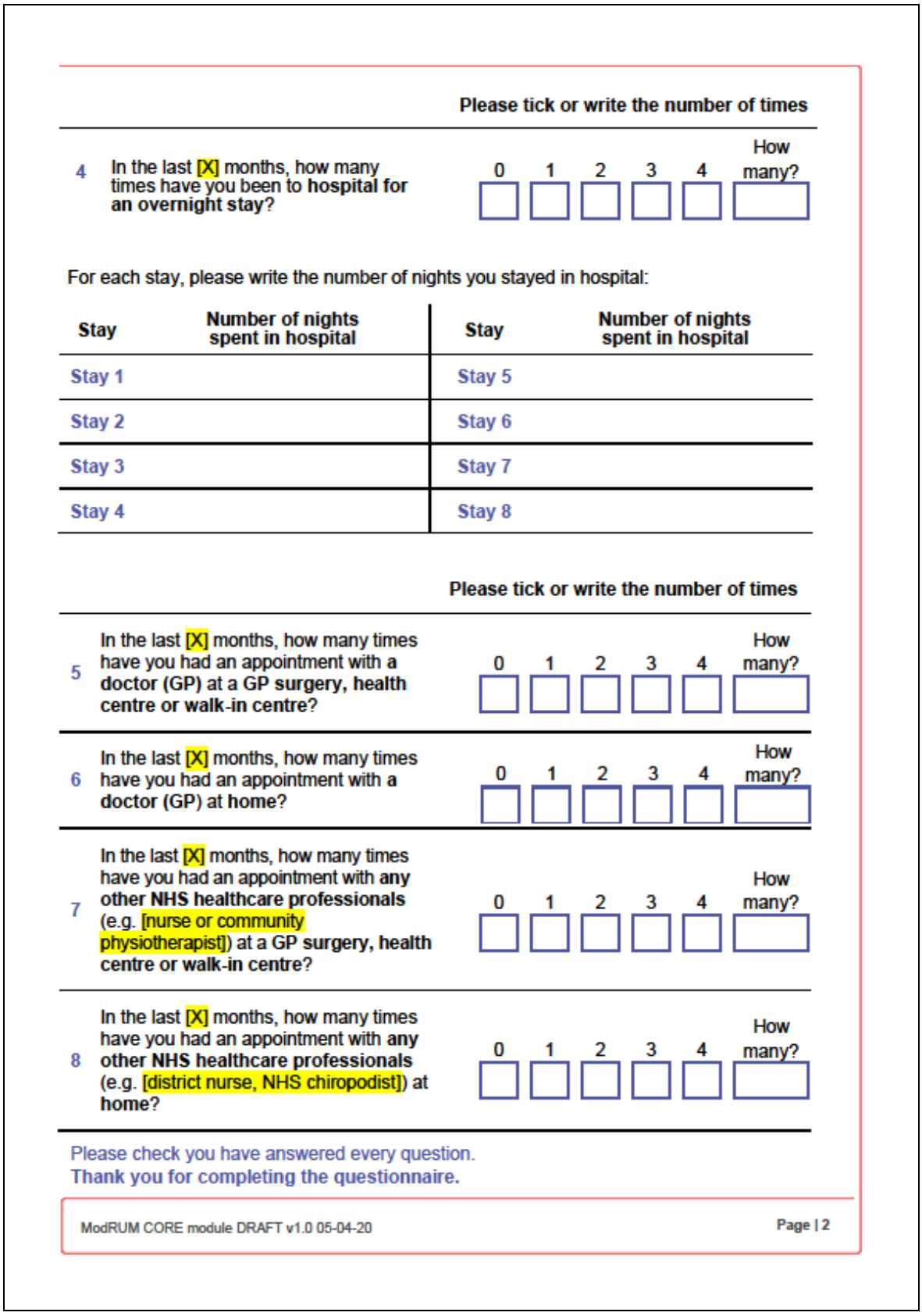
Your use of healthcare services

We would like you to answer some questions about the healthcare you have used [because of XXXXX/for any reason], **[in the last X months/since XXXXX]**. We only want you to **include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

		Please tick or write the number of times					
		0	1	2	3	4	How many?
1	In the last [X] months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	☐	☐	☐	☐	☐	☐
2	In the last [X] months, how many times have you been to **hospital for an outpatient appointment** (e.g. [to see a consultant or hospital physiotherapist, or to have an x-ray])?	☐	☐	☐	☐	☐	☐
3	In the last [X] months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)?	☐	☐	☐	☐	☐	☐

Figure 7.1 continued



Please tick or write the number of times

		0	1	2	3	4	How many?
4	In the last [X] months, how many times have you been to **hospital for an overnight stay?**	☐	☐	☐	☐	☐	☐

For each stay, please write the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital	Stay	Number of nights spent in hospital
Stay 1		Stay 5	
Stay 2		Stay 6	
Stay 3		Stay 7	
Stay 4		Stay 8	

Please tick or write the number of times

		0	1	2	3	4	How many?
5	In the last [X] months, how many times have you had an appointment with **a doctor (GP) at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
6	In the last [X] months, how many times have you had an appointment with **a doctor (GP) at home?**	☐	☐	☐	☐	☐	☐
7	In the last [X] months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. [nurse or community physiotherapist]) at **a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
8	In the last [X] months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. [district nurse, NHS chiropodist]) at **home?**	☐	☐	☐	☐	☐	☐

Please check you have answered every question.
Thank you for completing the questionnaire.

ModRUM CORE module DRAFT v1.0 05-04-20 Page | 2

7.4 Results

7.4.1 Respondent identification and recruitment

A search of the NIHR Journals Library identified 86 HTA and 19 PHR potentially eligible grants. Of these, 47 grants were excluded as they were not an RCT (n=13), had no health economics component (n=7), included child participants (n=21), or the lead health economist was a member of the research team or a University of Bristol colleague (n=6). The remaining 58 grants were led by 39 health economists, including nine health economists that were the lead health economist on multiple (up to four) grants.

Across two rounds, email invitations were sent to 39 health economists. Nine health economists did not respond and five declined to participate (**Table 7.1**). Reasons for declining included time constraints, limited capacity, poor timing, and trial delays due to Covid-19. Of 25 health economists who responded to the invitation expressing interest in participation, 15 did not subsequently participate. Follow-up emails were sent twice; however, many of these health economists responded to say that they were planning to take part, but time constraints and reasons related to Covid-19 had made it difficult to find the time. From the email invitations, 10 health economists participated. In addition to these, one member of the research team (JT) also took part in this study, as at the time she was already adapting ModRUM for a trial she was working on. Data were collected from respondents between July and October 2020.

Table 7.1 Health economist responses

	n	(%)
Did not respond	9	(23)
Declined to participate	5	(13)
Replied to the invitation but did not participate	15	(38)
Participated	10	(26)

7.4.2 Respondent characteristics

Respondents were based at eight different higher education institutions. Characteristics are presented in **Table 7.2**. Nine respondents had five or more years' experience of working in health economics and eight respondents had worked on five or more RCTs. Four respondents collaborated with a colleague to adapt ModRUM. Respondents reported that their collaborating colleagues had at least 11 years' experience of working in health economics and had worked on at least five trials.

7.4.3 Trial characteristics

Details of the 11 trials and their planned resource-use data collection methods are provided in **Table 7.3**. In brief, 10 trials were funded by the NIHR HTA programme. The trial setting was most often secondary care (65%), with other settings including primary care and the community. A broad range of conditions were covered and included cancer, frailty, workplace absence and intellectual disability.

Table 7.2 Characteristics of responders

	n	(%)
Number of years' experience working in health economics		
0 to 4 years	2	(18)
5 to 10 years	4	(36)
11 to 20 years	5	(45)
More than 20 years	0	(0)
Number of RCTs worked on		
0 to 4 trials	3	(27)
5 to 10 trials	4	(36)
More than 10 trials	4	(36)
Adapted ModRUM with a colleague	4	(36)
Number of years' experience colleague has worked in health economics		
0 to 4 years	0	(0)
5 to 10 years	0	(0)
11 to 20 years	2	(50)
More than 20 years	2	(50)
Number of RCTs colleague has worked on		
0 to 4 trials	0	(0)
5 to 10 trials	1	(25)
More than 10 trials	3	(75)

Table 7.3 Trial and resource-use data collection characteristics

	n	(%)
Programme		
HTA	10	91
PHR	1	9
Setting		
Secondary care	7	64
Primary care	2	18
Community	1	9
Community, social and educational	1	9
Area		
Cancer	2	18
Intellectual disability	1	9
Mental health	1	9
Mental health/intellectual disability	1	9
Orthopaedics	1	9
Frailty	1	9
Workplace absence	1	9
Cardiology	1	9
Infertility	1	9
Kidneys	1	9
Most suitable administration mode		
Self-complete	9	82
Interviewer-administered	2	18
Most suitable media administration mode		
Paper only	6	55
Telephone only	1	9
Online (computer and mobile)	2	18
Paper and online (computer)	1	9
Paper, telephone and online (computer and mobile)	1	9
Sectors beyond NHS-funded healthcare to be captured		
Informal care	7	64
Productivity	9	82
Social services	5	45
Criminal justice	1	9
Private*	1	9
Voluntary/third sector	1	9
Out of pocket costs	1	9

*unclear whether this referred to private healthcare, or something else

7.4.4 Summary of adapted ModRUM

On average, respondents reported spending 108 minutes (SD 63 minutes) adapting ModRUM for their trial. A summary of how respondents adapted ModRUM is provided in **Table 7.4**. Seven respondents returned their adapted version of ModRUM. All seven respondents used the core module and added depth questions. One respondent also returned their version of the core module only. The recall period used ranged from three to 12 months, which was a permitted adaptation. Six respondents adapted ModRUM to ask for resource use related to the condition of interest (e.g. 'because of your back pain'). Depth questions were most commonly asked for inpatient stays (n=5) and non-hospital-based other healthcare professionals (n=5). For the latter, four respondents added pre-specified healthcare professionals beyond 'GP' (see questions 10-12 in the core module with depth questions – **A5.3**) and four respondents added questions to capture remote (telephone/online) contacts.

Table 7.4 Summary of adapted versions of ModRUM

	n	(%)
Adapted ModRUM returned	7	(64)
ModRUM version adapted		
Core module with depth questions	6	(86)
Core module and core module with depth questions	1	(14)
Recall period used		
3 months	2	(29)
4 months	1	(14)
6 months	3	(43)
12 months	1	(14)
Resource-use type specified		
All-cause	0	(0)
Condition-specific	6	(86)
Not selected	1	(14)
Items where questions were included with added depth		
Ambulance service	2	(29)
Outpatient appointment	4	(57)
Day case visit	3	(43)
Inpatient stay	5	(71)
Non-hospital-based other healthcare professionals (including additional healthcare professionals e.g. nurse and modes e.g. telephone)	5	(71)
Prescribed medications	4	(57)

Four respondents made adaptations to ModRUM that were not explicitly stated as permitted in the user guide. These adaptations are summarised in **Table 7.5**.

Table 7.5 Adaptations beyond those specified as permitted in the ModRUM User Guide

Question	Adaptation	Respondent
A&E	Question not included	117
Ambulance service	Combined depth ambulance questions	111
Hospital outpatient	Core question duplicated to capture several pre-specified clinics and 'any other outpatient clinics'	111
Hospital outpatient	Healthcare professional seen' added to outpatient depth question table	125
Hospital inpatient	'Number of nights spent on intensive care ward' and 'number of nights spent on high dependency ward' added to inpatient depth question table	125
Hospital inpatient	'Name of hospital' and 'approximate date of first day in hospital' added, and 'department' and 'tests or surgical procedures' removed from inpatient depth question table	126
Hospital inpatient and day case	Questions not included	111
GP telephone or online	Adapted to 'a doctor (GP) or hospital doctor over the telephone or via video calling (e.g. Skype or similar)'	111
Non-hospital-based questions	Locations changed e.g. 'GP surgery, health centre or walk-in centre' changed to 'GP surgery'	111
Other healthcare professionals	'Private nurse' as an example	127
Non-hospital-based questions	Walk-in centre question added separately	111
Non-hospital-based questions	NHS 111/NHS 24 question added separately	111
All questions except hospital inpatient	'Because of your back pain' added to the end of each question	126

7.4.5 Summary of feedback

In this section and **Table 7.6**, responses to the survey are summarised.

7.1.1.1. Ease of making adaptations to ModRUM

Ten respondents reported that adapting ModRUM was ‘fairly easy’ or ‘easy’. Health economists primarily reported that the resource coverage and format made adaptation easy.

> R124: *All of the questions I needed were there and I could replace some of the core questions with some of the more in-depth questions so as to capture additional detail*
>
> R122: *The questions were very clear, well worded and laid out, with useful examples provided. The depth questions offered the opportunity to add extra detail which was very useful.*

Respondents who scored adaptation as ‘fairly easy’ were also asked to report on what they found difficult. Five respondents reported difficulty in editing ModRUM and described the process as *“fiddly”* (R101) and *“tricky to adapt”* (R126).

> R124: *Copying and pasting the in-depth questions into Word - questions would overlap, spacing would alter. It took more time to rectify this than to amend the actual questions.*

Although they scored the adaptation process as ‘fairly easy’, some respondents experienced other difficulties including the identification of which core and depth questions were equivalent, as for some resource items the question number differed in the core and depth questions. Some respondents reported that a level of detail in between the core and depth would be more appropriate for their trial. Some respondents also needed to capture resource use not currently included in ModRUM (e.g. private care and productivity losses).

R117 described the adaptation process as ‘difficult’. They reported several difficulties. First, they reported that some of the questions, such as ambulance services, were not relevant for their trial, which was set in secondary care and involved orthopaedic surgery; however, it was clear to them from the User Guide that items in the longer version of ModRUM are optional. They also said there was a lack of clarity, for example, in how response boxes should be used, and that the question

order not being adaptable was a difficulty as it does not allow for the most important questions to be included first.

Most suggested improvements related to improving the editing process. Respondents suggested that we could *"make it easier & (clearer) to 'drag and drop' items from long version into short version"* (R101), generate a question bank from which questions can be cut and pasted into ModRUM and potentially utilising alternative software.

Adaptations were reported or suggested that were beyond substituting core for depth questions. R124 reported amendments to *"the in-depth questions to record who a patient had seen at an outpatient clinic, or how many inpatient nights had been spent on different wards"*. R101 suggested that guidance could be added for patients utilising services from multidisciplinary teams.

> R101: *patients in this trial get their healthcare though multidisciplinary community mental health teams - rather than distinct healthcare professionals.*

Several of the suggested improvements could potentially be implemented by specifying them as options in the ModRUM User Guide. Suggestions included *"having the option of predefined fields for resource utilisation rather than examples"* (R103) and allowing an intermediate option between core and depth which could be used when *"…space is tight/participant burden is a key consideration, but still aiming to get a bit of extra detail to help with the costing"* (R122). R122 also suggested there could be more flexibility in the sizing of rows to allow them to ***"****…fit more into the questionnaire without using up as many extra pages - which can be a consideration for postage/printing costs etc."*. Respondents suggested that increased clarity could also be provided in the User Guide indicating which core and depth questions are interchangeable for each resource item. This would be helpful as the question number sometimes differs for the same resource item in the core and depth questions.

Table 7.6 Descriptive summary of survey responses

	Mean (SD) (minimum, maximum)
Adaptation time (minutes)	108 (63) (45, 240)
	n (%)
Ease of adaptation	
Very difficult	0 (0)
Difficult	1 (9)
Fairly easy	9 (82)
Easy	1 (9)
Very easy	0 (0)
Number of respondents who used the ModRUM User Guide	9 (82)
Clarity of the ModRUM User Guide	
Very unclear	0 (0)
Unclear	0 (0)
Fairly clear	1 (11)
Clear	6 (67)
Very clear	2 (22)
Ease of applying unit cost data to ModRUM responses	
Very difficult	0 (0)
Difficult	1 (9)
Fairly easy	3 (27)
Easy	4 (36)
Very easy	3 (27)
Feasible to develop a compendium of costs for ModRUM responses	
Yes	7 (64)
No	4 (36)
Likelihood of using ModRUM in the future	
Very unlikely	0 (0)
Unlikely	1 (9)
Fairly likely	1 (9)
Likely	5 (45)
Very likely	4 (36)
Can foresee barriers to using ModRUM for collecting resource-use data in future trials	
Yes	6 (55)
No	5 (45)

7.4.5.1 Clarity of the ModRUM User Guide

Nine respondents used the ModRUM User Guide to adapt ModRUM. Of the two respondents who did not use the guide, R103 reported "time" as the reason they did not read the guide and the other respondent was a member of the research team who had read the guide previously. Respondents reported that the guide was 'fairly clear' (n=1), 'clear' (n=6) or 'very clear' (n=2). Most feedback provided on the User Guide was favourable, with R111 describing it as *"easy to follow"* and R122 stating that the guide included *"useful formatting tips"*.

> R108: *The guide clearly explained how amendments could be made to the template.*

Specifically, respondents indicated that they "liked" (R117, R124) and found the 'quick start guide' at the beginning of the User Guide "useful" (R122).

R124 indicated that more clarity was needed about which depth and core questions are equivalent because the equivalent questions have different numbers in each version which "could lead to issues with double counting". R124 also reported an omission from the guide: *"the quick guide mentions section instructions and adding the extra space box, which I had expected to find more detail on in section 7 - I actually forgot to do this because it wasn't mentioned again"*.

7.4.5.2 Costing resource-use data from ModRUM

Respondents were asked to report how easy they thought it would be to apply unit costs to resource-use data captured in ModRUM. Seven respondents said that they thought application of unit costs would be 'easy or 'very easy', while three respondents thought it would be 'fairly easy and R117 thought it would be 'difficult'. R117 said that lots of (the depth) questions require free-text responses which would cause *"patients to write all sorts of different things making it very hard/sometimes unfeasible to properly cost the questionnaire"*. R108 also highlighted this issue. Other potential issues were highlighted by R103 who said there may be difficulty in *"assigning a HRG for hospital resource use", and R109* who said "*we don't know what type of healthcare professionals carry out outpatient visits, makes applying unit cost data more difficult… ideally we would like to know the healthcare professional (consultant vs nurse for example)".* The remainder of feedback on applying costs was positive with six respondents indicating that resource-use data could be easily

linked to “published costs” from “established sources”, and two respondents also indicating there was sufficient granularity.

> R111: *I think most of the options included are standard items you would expect to cost as part of an econ eval and they would be available from routine sources (e.g. NHS ref costs, pssru)*
>
> R122: *The resource use in the questions have published costs available, and the depth questions help to provide additional detail for more accurate costing.*

Respondents were also asked whether they thought it would be feasible to develop a compendium of costs for ModRUM responses (i.e. a document containing unit costs corresponding to each question). Seven respondents thought it would be feasible, and four respondents thought it would be unfeasible. One reason for the latter was that unit costs are updated on a regular basis, so a compendium would need to be continually updated. The second reason was that creating a compendium, particularly for depth questions, would be a cumbersome and potentially infeasible task, as the questions would capture a substantial list of resources.

> R124: *For some contacts e.g. outpatient clinics, the possibility exists for there to be very many different clinic types, across different specialties with different healthcare professionals seen. This is also the case for hospital inpatient admissions and procedures but would not be so much a problem for primary and community care contacts. Also such a compendium would need to be continually updated and maintained.*

R111 said that *“having specific unit costs might not be needed as for e.g. if you are working on a specific disease area such as oncology it would be more appropriate to cost an outpatient visit to this specific department than use the generic cost that would be published with the ModRUM”*. They said, *“an alternative could be to provide the sources of where the unit costs are available and what unit costs you would recommend for each item”*.

7.4.5.3 Future use of ModRUM

Ten respondents said that it is very likely, likely or fairly likely that they will use ModRUM in future. While R103 said it is fairly unlikely that they will use ModRUM in future, when asked for the reason, they reported that they *"already have a predefined instrument for our current trial"*, which suggested they misinterpreted the question. Respondents made several suggested improvements including: developing a *"supporting patient diary so that the patient can recall all the granular detail"* (R103), *"including specific categories rather than just a few examples"* so that resources that are likely to be highly utilised are prespecified in tables (R108), and putting GP and other healthcare professional questions (questions 7-12 of the core module with depth questions) into a table so they *are "asked more concisely… and could cover more than 2 types of healthcare professionals"* (R108).

Respondents were also asked if they could foresee any barriers to using ModRUM for collecting resource-use data in future trials. Six respondents reported that they could foresee barriers. Several issues had already been raised earlier in the survey, including lack of clarity about which core and depth questions are equivalent, potentially requiring a level of granularity between the core and depth questions, and difficulty adapting ModRUM in Microsoft Word. R101 said the lack of modules to capture a societal perspective, and in particular "for public/mental health, social care", could be a barrier. R117 thought the complexity and length of ModRUM may not be well accepted by trial teams and R108 thought that it may not be well-accepted by trial participants. However, it is unclear whether R108 understood that depth questions were not compulsory, suggesting that further guidance on this is required in the User Guide.

> R108: *ModRUM is very comprehensive, but the structure is quite long and requires a lot of free text from the survey respondent. This could potentially be off-putting to some respondents, and may cause some difficulty in interpreting the responses if they are not consistent.*

7.4.6 Summary of changes

Using respondents survey responses and corresponding codes, I created a summary document detailing the feedback obtained, which I presented at a research team meeting. Feedback was summarised under several categories including the

formatting of ModRUM, the user guide and content of ModRUM. For each category, I described the findings and informed the team how I thought the feedback should influence ongoing development of ModRUM. As a team, we then came to an agreement on what changes should and should not be made. This information is summarised in **Table 7.7**. In brief, we agreed that within the scope of this PhD, revisions could be made to improve clarity of the User Guide (see **A5.5** for the revised ModRUM User Guide). It was agreed that no further changes should be made to the content of ModRUM, as suggestions were either beyond the scope of ModRUM at present or deviated from the design principles for ModRUM outlined in **Table 1.1** (**Chapter 1**). Feedback provided on enhancing the adaptation process, by utilising alternative software and developing a complementary resource-use log, were beyond the scope of this thesis, but the research team agreed they should be considered for future research.

Table 7.7 Summary of respondent feedback and agreed changes

Category	Feedback summary	Decision	Rationale for the decision
Formatting of ModRUM	Utilise alternative software to improve ease of adaptation (e.g. functionality to drag and drop questions from depth into core)	Agreed, post-PhD	This study has highlighted that current formatting constraints may be inhibit uptake of ModRUM and future work should prioritise work to improve this.
	Provide a question bank	Potentially, post-PhD	
ModRUM User Guide	Clarification on which core and depth questions are equivalent, needed to avoid double-counting	Update guide	Additional detail was added to the ModRUM User Guide to increase clarity.
	Clarification on permitted changes		
	Instructions for the extra space box are only included in the 'quick start guide'		
	Add guidance on how to capture multidisciplinary teams	No change	This is potentially trial-specific and beyond the scope of the User Guide, which we ideally want to keep concise.
ModRUM content	Option to include healthcare professional for outpatients (e.g. consultant versus nurse)	No change	This moves away from the ModRUM design principles concise and comprehensible.
	Option for capturing inpatient stays on different wards	No change	This moves away from the ModRUM design principles concise and comprehensible.
	Option for an intermediate level, as core questions may be too brief and depth questions may be too comprehensive	No change	The User Guide already states that other items can be added if necessary. Specifically stating that an intermediate level is an option would encourage use of such a level, which would reduce consistency.
	Capture out-of-pocket expenses that may be incurred for prescribed medications	No change	This is beyond the scope of these modules of NHS-funded healthcare only.

Table 7.7 continued

Category	Feedback summary	Decision	Rationale for the decision
Question design	Allow the order of questions to be changed so the most important questions can come first.	No change	This suggestion was also raised by health economists in **Chapter 5**; however, no change was made as we want to retain consistency in how ModRUM is implemented.
	Allow tables to include pre-specified resources to make completion quicker	Potentially, to be considered further	This will be considered further as there are advantages to this approach, particularly if patients are likely to have multiple identical healthcare contacts (e.g. chemotherapy visits).
	Questions 7-12 of the depth module could be asked more concisely by using a table	No change	Tables were considered at the prototype development stage. Standalone questions were chosen for the core module to keep ModRUM uncluttered and have been used in the depth questions to maintain consistency in presentation of ModRUM.
Unit costs compendium	Provide sources of unit costs instead of an exhaustive list of unit costs	Agreed	Reporting sources of unit costs would be preferrable and easier to maintain. Sources were identified in the patient pilot (**Chapter 8**) and included in the revised User Guide. The User Guide could be extended post-PhD to describe how sources could be linked to each ModRUM question.
Other	Develop a complementary resource-use log to act as a memory aid	Potentially, post-PhD	A resource-use log may be developed in future, as they have been shown to reduce missing data of some resources (55).

7.5 Discussion

7.5.1 Main findings

In this study, health economists adapted ModRUM for a recently funded trial and provided feedback in an online survey. Most respondents found that it was feasible to adapt ModRUM and said they were likely to use ModRUM in the future, indicating that it is suitable for capturing resource-use data in trials. While the ModRUM User Guide was well-received, many respondents suggested that improvements could be made to simplify the adaptation process. Of those who returned their adapted version of ModRUM, all respondents added some depth questions to the core module, but the depth questions included differed dependent on the study. Several health economists made adaptations that were not specified as permitted in the User Guide.

7.5.2 Strengths and limitations of this study

Eleven health economists provided feedback on their experience of adapting ModRUM for a recently funded trial. This study provided a different perspective to the qualitative interviews with health economists (**Chapter 5**) as I attempted to replicate how health economists would prepare ModRUM for their trials. Respondents returned their adapted versions of ModRUM, which meant I could identify which areas they had changed and whether they had followed the User Guide. Responses to closed and follow-up open questions provided feedback which has informed where refinements could be made.

Using an online survey was a cost-effective and resourceful method for inviting a large number of health economists to participate. Alternative methods, such as an interviewer administered survey, may have introduced bias, whereby respondents modify their responses as they are providing feedback directly. However, as the questions were self-administered, the inability to probe respondents meant that in comparison with the interviews reported in **Chapter 5**, I did not obtain the same level of feedback from respondents and was unable to clarify feedback further (111). In addition, respondents in this study provided varying degrees of detail in their feedback.

Although the sample size was small, I was able to recruit health economists with a range of experience of working in health economics and working on RCTs, from across eight higher education institutions. Respondents adapted ModRUM for trials being conducted in a range of settings and areas. While no formal sample size was estimated as there was no precedent to follow. By identifying and approaching 39 health economists, I had hoped to obtain feedback from more than the 11 health economists who ultimately participated. The sampling strategy meant that only HTA and PHR funded trials were included. Of the 39 grants that were selected for this study, only six were PHR grants. Of the six health economists approached who were working on PHR grants, only one participated which limits the generalisability of this study to trials conducted in a non-HTA setting.

Fifteen health economists expressed some form of interest or intended to participate when replying to the initial contact email but did not participate. An increased number of respondents may have provided a greater breadth of feedback and provided more evidence for the feedback obtained. Many health economists cited time pressures or trial delays that had been exacerbated by the Covid-19 pandemic as the reason they had been unable to participate (n=11). After recognising that time pressures were a key constraint to participation, in later follow-up invitation emails I included the average time it had taken respondents who had already participated to adapt ModRUM. I also offered to meet (either online or via telephone) with potential respondents to go through the survey with them; however, there was no uptake of this method. Health economists may have also not responded due to a belief that ModRUM was unsuitable for their trial, which they may not have expressed in their correspondence with me. Three health economists did however allude to a lack of suitability, with two reporting that they were planning for the majority of data to be extracted from administrative data rather than from self-report, and one reporting that they required very specific and detailed resource-use information to enable micro-costing. ModRUM is not designed to enable micro-costing, it is intended as a way of capturing other healthcare use during trial follow up, which may complement methods to micro-cost interventions (e.g. case report forms).

7.5.3 Comparison with the existing literature

RUMs are generally developed on a trial-by-trial basis (9) and therefore it is unlikely that their developers consider how future users may implement their RUM in another trial. One RUM which does provide a guide for future users is the ACQP, which is described in detail in **Chapter 3** (76). In their discussion paper on the ACQP, Thompson and Wordsworth provide a bank of questions, categorised under resource type groups (e.g. *"single visits to health care facilities"*), from which users can formulate a RUM (76) (pg.2). This differs from ModRUM where users are provided with a ready-formulated RUM and a user guide which provides information on how the RUM can be adapted.

In many ways, the ACQP is less prescriptive than ModRUM (76). For example, ModRUM has a core module with questions that should be captured in all trials, whereas for the ACQP, the user should decide which items to include based upon their patient group. Another difference is the respondent instructions/preamble. In ModRUM, it is predefined with several parts which can be adapted to make ModRUM trial specific. For the ACQP, it is not prespecified, as the authors state that the preamble *"will differ widely between the contexts in which the questionnaire is used. Instead we have chosen simply to list the issues we believe should be addressed"* (76) (pg.9). The instructions for respondents at the beginning of ModRUM cover what to include (e.g. resources used as an NHS patient) and how to answer the questions (e.g. record an answer even if it is zero), whereas the ACQP goes beyond this and includes things such as a confidentiality statement and investigator contact details (76). As ModRUM is likely to be nested in a large battery of questionnaires, these have not been included as they are likely to be covered in a participant information sheet and/or cover letter. An area of commonality is the recall period; users are able to select this in both RUMs.

In design, ModRUM may be considered more like the CSRI than the ACQP, with respect to users making adaptations to a fully formed measure, rather than constructing the measure from a library of questions (74, 76). Beecham and Knapp have reported that while the CSRI was made bespoke for a specific context, one of the advantages of it is that it is easily adaptable (74). They also describe how these adaptations have been implemented, such as using alternative modes of

administration to that intended and extracting key questions from the CSRI to include in another measure (74). While uptake of the CSRI has been considerable in comparison to other RUMs, with it being included in over 100 studies, there does not appear to be clear guidelines on what adaptations can and cannot be made (75). As a result, the CSRI is susceptible to ad-hoc changes, with often little clarity on what changes have been made and how the changes impact validity. Uptake of ModRUM, akin to the CSRI is desirable; however, while offering more transparency, the more stringent adaptation rules may mean that ModRUM is less appealing to users.

With respect to development of a user guide and testing the guide with potential users (health economists), I am not aware that this has been conducted before. In the systematic review (**Chapter 3**), I found almost no evidence to suggest that independent health economists were involved in the piloting of existing RUMs. Where health economist involvement has been reported, it was in consensus meetings with clinicians to judge the completeness and consistency of the RUM and it was unclear whether the health economists were independent or part of the development team (142).

7.5.4 Implications for research practice

In this study, health economists provided mostly positive feedback on ModRUM and the accompanying User Guide, suggesting that ModRUM will be a useful tool for health economists capturing patient-reported resource use in future economic evaluations alongside RCTs. Suggested modifications to the User Guide were implemented to enhance clarity. Future work, which is beyond the scope of this thesis, such as utilising alternative software, could improve the ease of adaptation of ModRUM. By employing these changes, the user experience will be enhanced, which should increase uptake of ModRUM. Assuming ModRUM will be well-utilised in future trials, researchers and decision makers will benefit from increased comparability between trials of resource-use data.

7.5.5 Implications for this research

As there is a precedent for developing RUMs for each new trial, adaptations that were not specified as permitted in the User Guide were made. Based on respondent feedback, the large proportion of respondents who made unpermitted adaptations may have also been exacerbated by the hypothetical nature of this study. Some respondents already had clear ideas about the data they planned to collect, or they had already developed a RUM for use in their trial; it is likely that deviating from these ideas to align with ModRUM would have been more difficult than if ModRUM was an established RUM that was selected at the outset of the trial. To minimise unpermitted adaptations which may mean that questions are no longer valid and inhibit comparability across trials, more emphasis on this was added to the User Guide. Consideration will also be given to whether users of ModRUM should register and share their version of ModRUM with the research team prior to it being accepted for use in a trial. This could be an onerous task for the research team and therefore consideration should be given to whether there should be a license fee for commercial use to cover costs with, for example, a registration process similar to the EQ-5D, where a fee is charged for commercial use, but may not be charged for non-commercial research could be used (182).

While the aim of this research is to develop a new generic RUM that can be used in a wide range of trials and to perform initial validation, the ModRUM User Guide could also be expanded further to include guidance on adapting ModRUM, how modifications may impact validity, and when and how further validation should be performed. Understanding when modifications impact validity and when further validation is required is unclear; however, additional validation could be considered when changes are made to the questionnaire design, population and administration mode. For questionnaire design, assuming the population and administration mode remain the same, some changes to ModRUM may not require further validation if they have already been validated (e.g. a three-month recall period has been tested in this thesis). Other changes to the design may require further validation. For example, if ModRUM is adapted to capture resource use associated with a particular condition, validation may be required to test whether participants can accurately determine related and unrelated resources to that condition. For some populations, further validation may be required, for example if members of the population group

are cognitively impaired. Finally, the aim of this thesis is to develop and perform validation of a paper-based version of ModRUM for self-completion by patients themselves. Alternative administration modes, such as computer-based, interviewer-administration or proxy completion are likely to benefit from further validation.

7.5.6 Unanswered questions and future research

The feedback provided by health economists during this study suggests that many are likely to use ModRUM and that ModRUM is suitable for use in a broad range of future trials. While health economists are an important part of the target population who would utilise ModRUM, future research may capture the opinions of a wider group of trialists, as one respondent indicated that ModRUM may not be well-accepted by trial teams due to the length and complexity. However, this comment was from a respondent who chose the core module with depth questions version of ModRUM. In comparison to existing RUMs, ModRUM, in particular the core module, may be preferred by wider research teams as it is comparatively shorter and less complex than most RUMs. Trial chief investigators or trial managers may offer an alternative perspective on the acceptability of ModRUM. They often have a large number of patient-reported measures to include in a trial, so would be well-placed to judge the suitability of ModRUM, particularly with reference to length and participant-burden, when it is to be collected alongside other measures. In the context of a trial, with a battery of outcome measures alongside ModRUM, members of patient and public involvement groups could also comment on the acceptability of ModRUM. While gaining feedback from individuals with a broader range of perspectives would be valuable, it is also likely to yield diverging opinions on the optimal level of detail. The development of core questions, that are suitable for most trials, and depth questions, that capture more details that may be relevant to some trials, means that ModRUM can cater for different requirements on the level of depth.

In this study, one respondent suggested that the length and quantity of free text required may be off-putting for some respondents. This feedback emulates comments made by two participants, who completed the core module with depth questions version of ModRUM, in patient interviews (**Chapter 6**). They suggested that ModRUM may not have been acceptable had they used more resources. Future research could explore the acceptability of ModRUM (in particular, the core module

with depth questions) with patients who use a substantial amount of healthcare and in the context of a trial, where ModRUM is completed alongside many other measures.

Beyond this theis, future work should focus on the development of breadth modules which would provide flexibility to capture resources beyond an NHS perspective. In this study, all but one respondent reported collecting resource-use data from beyond an NHS perspective, indicating that there would be value in developing such modules. Informal care, social care and productivity losses were most commonly reported in addition to the NHS perspective. As instruments are already available and well-used for measuring productivity, including Institute for Medical Technology Assessment Productivity Cost Questionnaire (183) and the Work Productivity and Activity Impairment Questionnaire (184), the development of social care and informal care modules has already been prioritised for future research (see **Chapter 9**, Section **9.4.1**). The addition of the social care module will mean that ModRUM can capture resources from an NHS and PSS perspective which is recommended in the NICE reference case.

7.6 Conclusion

In this pilot study, most health economists reported that it is feasible to adapt ModRUM and that ModRUM is suitable for capturing resource-use data in their trials. Favourable opinions were generally provided on the ModRUM User Guide. Respondents made several suggestions on improvements that could be made to simplify the adaptation process.

Chapter 8 Piloting with patients

8.1 Chapter overview

This chapter reports on a study where ModRUM was piloted with patients recruited from a primary care setting. Patients were invited via postal invitation to complete ModRUM along with a characteristics form and a quality-of-life questionnaire, and to provide consent for their responses to ModRUM to be compared with their GP medical records. Responses were analysed using descriptive statistics and statistical analyses to assess the acceptability, feasibility, construct validity and criterion validity of ModRUM.

This chapter begins with a background and objectives section, which outlines the justification for, and objectives of, this study. A methods section follows where the study design and data analysis plan are outlined. The results section is split into several sub-sections, including a section reporting on participant characteristics, quality of life and resource use, and sections reporting on acceptability, feasibility, construct validity and criterion validity. The discussion includes a summary of the main findings, strengths and limitations of this study, a comparison to existing literature, implications for research practice, and unanswered questions and future research. A brief conclusion is provided at the end of the chapter.

8.2 Background and objectives

Once the content and face validity of a new instrument have been established, the remaining measurement properties should be tested (11). Measurement properties including feasibility, acceptability, construct validity and criterion validity can be tested in a larger quantitative study (11). The feasibility of a new instrument requires the instrument to be viable for respondents to complete and for researchers to administer and analyse (79), while the acceptability assesses whether the instrument is tolerable to respondents (81). Construct validity can be established through hypothesis testing to assess whether the association between scores from the instrument correlate as expected to another instrument measuring the same or a related construct, or to a respondent characteristic which is hypothesised to be associated with the construct of interest (79). To test the criterion validity of a new

instrument, the scores from the new instrument are compared with the scores of another measure, which is ideally the 'gold-standard' measure (79).

Evidence from existing publications on pilot testing and the assessment of construct and criterion validity for existing RUMs is presented in **Chapter 3**. The review identified that piloting of existing RUMs has involved a variety of methods and has been conducted for a variety of purposes. The main purpose of piloting for existing RUMs was to identify issues and refine the RUM, with the aim of improving acceptability to respondents and increasing data quality (75, 121, 123, 130, 131, 135, 136, 146, 149, 150). The aims of the study reported in this chapter were to: (1) assess the feasibility and acceptability of completing ModRUM, (2) assess the feasibility of cleaning data, and identifying and applying unit cost data to ModRUM responses, and (3) test the construct and criterion validity of ModRUM.

8.3 Methods

8.3.1 Study design

In this pilot study, patients at participating GP practices were invited, via post, to complete ModRUM, a characteristics form and a quality-of-life questionnaire. Patients were also asked to provide consent for their primary care medical records, on consultations and prescribed medications, to be shared with the research team. Ethical approval for this study was provided by South Central - Berkshire B Research Ethics Committee (REC reference 19/SC/0244).

8.3.2 Site identification and recruitment

GP practices were recruited to take part in both the cognitive interviews study (**Chapter 6**) and this patient pilot study. GP practice identification and recruitment is described in section **6.3.2**.

8.3.3 Patient identification and recruitment

Sample size

The sample size for this study was determined pragmatically. Prior to the Covid-19 pandemic, I had planned to recruit patients from GP practice waiting rooms. Using this approach, I was aiming to recruit 400 patients, which would have exceeded the sample size of many studies assessing agreement between self-report and

administrative data, which were included in a systematic review (50). However, due to Covid-19, face-to-face recruitment and data collection were not permitted. Recruitment via post meant that the financial cost of conducting the study significantly increased, to cover support from GP practices and postal expenses. Budgetary constraints meant that up to 800 patients could be invited to participate in this study. I was uncertain what the response rate would be in this study, as I do not believe there has been an equivalent study that provides indicative response rates. However, based on response rates to postal surveys in other areas, it was likely that the response rate would be below 30 percent. For example, the GP Patient Survey 2020, which included multiple reminders to non-responders, achieved a response rate of 32 percent. While a study exploring the effect on response rate of personalisation, questionnaire length and type of reminder for a travel behaviour, physical activity and the environment achieved an overall response rate of 17 percent, with the response rate prior to reminders below 10 percent (185, 186).

Identification and eligibility assessment

I liaised with my contact at each GP practice to identify patients to take part in this study. First, a member of the GP practice administrative team extracted a list of adult patients who had had an appointment (face-to-face or remote) with a member of the clinical team (e.g. GP, nurse) within the last four weeks. This list was then shared with a GP at the practice, who screened patients for eligibility. GPs were asked to screen patients subject to the following eligibility criteria:

- aged 18 or over,
- capable of understanding and completing a questionnaire in English and
- capable of giving informed consent.

The number of invites each practice was requested to send is presented in **Table 8.1**. To account for potentially lower response rates, practices in more deprived areas were asked to send more invitations. Due to time constraints and to simplify logistics for GP practices, each practice was assigned to send either the core module or the core module with depth questions to their patients. Three practices were assigned to sending the core module with depth questions, so that overall more invitations for the core module with depth questions version would be sent, to account for potentially lower response rates to the longer questionnaire.

Table 8.1 Number of invites to be sent, by practice

Practice number	Deprivation score (1=most deprived; 10=least deprived)	ModRUM version	Number of patients the practice was asked to invite
1	2	Core with depth	160
2	5	Core with depth	160
3	4	Core	200
4	10	Core	150
5	10	Core with depth	130

Recruitment

I prepared mailout packs in Docmail, which is a hybrid mailing service. As I was unable to access patient details, a member of the GP practice administrative team uploaded patient contact details to Docmail. Docmail then printed out and sent the study documentation directly to patients. To potentially increase the response rate, invitation letters to patients were addressed to the patient from a GP at their GP practice (**A6.1**, **Appendix 6**) (187). Patients who wished to participate were asked to complete the documents and return them in an enclosed pre-paid return envelope.

8.3.4 Data collection

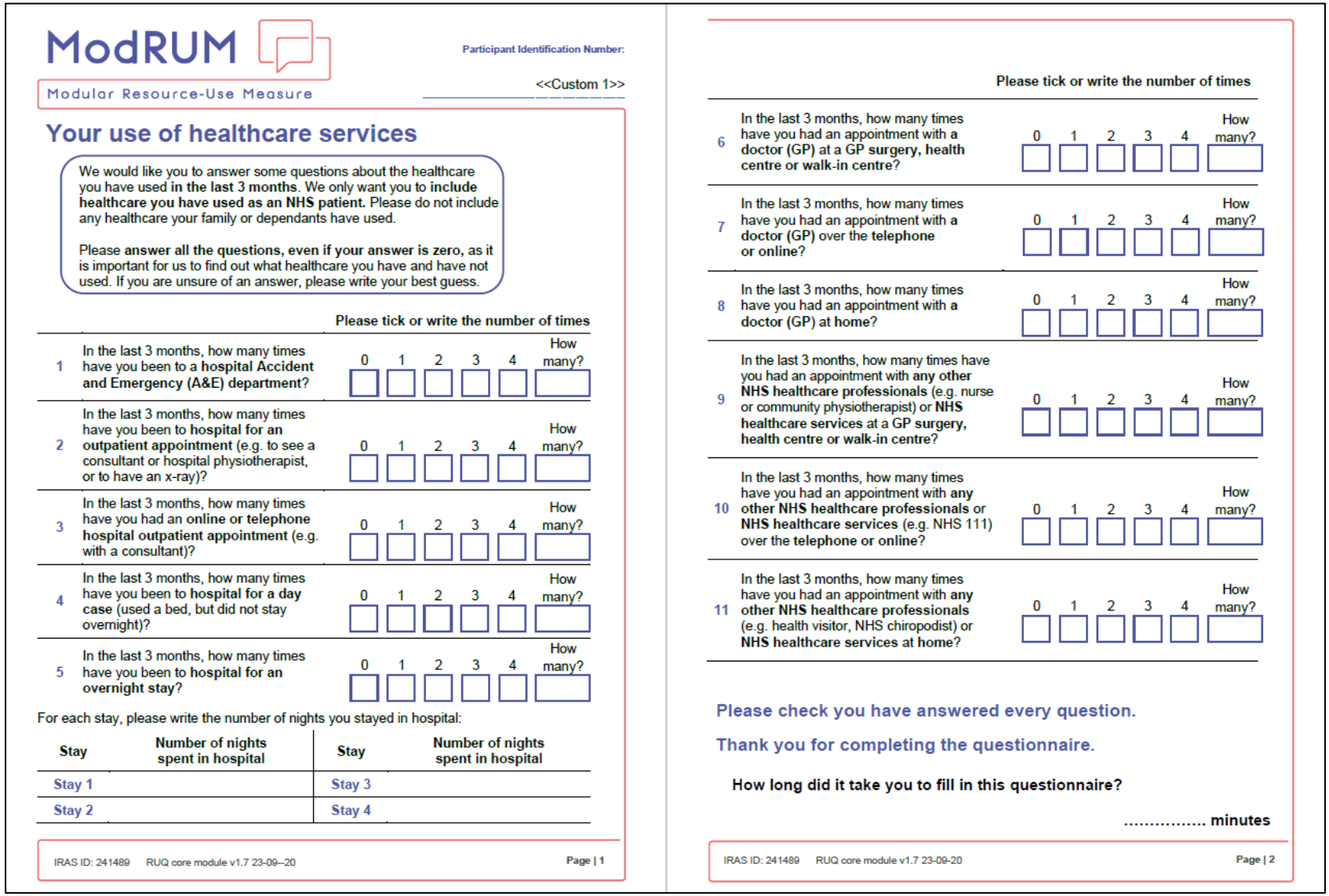
Mailout packs included ModRUM core module (labelled ModRUM-C hereinafter) (**Figure 8.1**) or ModRUM core module with depth questions (labelled ModRUM-CD hereinafter) (**A6.2**). All questions referred to a three-month recall period, which represents a commonly used recall period in trials (115). Given the rapid increase in remote consultations due to the Covid-19 pandemic, questions on remote consultations were added to ModRUM for this pilot study. Mailout packs also included a patient information sheet (**A6.3**), two consent forms (**A6.4**), the EQ-5D-5L, a patient characteristics form (**A6.5**) and a contact details form. To increase personalisation of the information sheet, I included a picture of myself (187). Participants were asked to complete the contact details form if they wanted to receive a summary of results at the end of the study and/or they wanted to enter a competition to win a £100 Love2Shop voucher.

The EQ-5D-5L is a generic preference-based measure that is often used in RCTs to estimate health-related quality of life (30). Approval to use the EQ-5D-5L was obtained from the EuroQol Research Foundation in January 2019. The main reason for including the EQ-5D-5L was to assess the construct validity of ModRUM; however, the inclusion of the EQ-5D-5L and the patient characteristics form in the mailout pack also helped to increase the external validity of the study, as it meant that completing ModRUM was more similar to completing a RUM in a trial setting, where the completion of multiple questionnaires is the norm. Participants were required to complete the EQ-VAS, as part of the EQ-5D-5L, this involves indicating their health today by marking a point on a VAS scale where 100 is 'the best health you can imagine' and zero is 'the worst health you can imagine' (30). To overcome potential imprecision in marking a point, participants are also asked to write the number, from zero to 100 in a box alongside the scale (188).

To inform acceptability, at the end of ModRUM, a question was added where participants were asked to report how long it took them to complete ModRUM.

At least eight weeks following completion of ModRUM, I provided my contact at each GP practice with a list of participant identification numbers, for participants who consented for their medical records to be shared with the research team. A data extraction script for EMIS Web was developed by contacts at two GP practices. Each practice used the script to extract data on participant consultations and prescribed medications. Identifiable data was replaced with participant identification numbers, by GP practice contacts, and anonymised data was securely sent to me.

Figure 8.1 ModRUM core module - patient pilot version



ModRUM

Modular Resource-Use Measure

Participant Identification Number:

<<Custom 1>>

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used **in the last 3 months**. We only want you to **include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

		Please tick or write the number of times					
		0	1	2	3	4	How many?
1	In the last 3 months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	☐	☐	☐	☐	☐	☐
2	In the last 3 months, how many times have you been to **hospital for an outpatient appointment** (e.g. to see a consultant or hospital physiotherapist, or to have an x-ray)?	☐	☐	☐	☐	☐	☐
3	In the last 3 months, how many times have you had an **online or telephone hospital outpatient appointment** (e.g. with a consultant)?	☐	☐	☐	☐	☐	☐
4	In the last 3 months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)?	☐	☐	☐	☐	☐	☐
5	In the last 3 months, how many times have you been to **hospital for an overnight stay?**	☐	☐	☐	☐	☐	☐

For each stay, please write the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital	Stay	Number of nights spent in hospital
Stay 1		Stay 3	
Stay 2		Stay 4	

		Please tick or write the number of times					
		0	1	2	3	4	How many?
6	In the last 3 months, how many times have you had an appointment with a **doctor (GP) at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
7	In the last 3 months, how many times have you had an appointment with a **doctor (GP) over the telephone or online?**	☐	☐	☐	☐	☐	☐
8	In the last 3 months, how many times have you had an appointment with a **doctor (GP) at home?**	☐	☐	☐	☐	☐	☐
9	In the last 3 months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. nurse or community physiotherapist) or **NHS healthcare services at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
10	In the last 3 months, how many times have you had an appointment with **any other NHS healthcare professionals** or **NHS healthcare services** (e.g. NHS 111) over the **telephone or online?**	☐	☐	☐	☐	☐	☐
11	In the last 3 months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. health visitor, NHS chiropodist) or **NHS healthcare services at home?**	☐	☐	☐	☐	☐	☐

Please check you have answered every question.

Thank you for completing the questionnaire.

How long did it take you to fill in this questionnaire?

..... **minutes**

8.3.5 Data analysis

I developed a Microsoft Access database where I entered data from ModRUM, the EQ-5D-5L and the patient characteristics form. Validation rules were included to restrict entered data to permitted formats and ranges. To check for data entry errors, once data had been entered for all participants, I compared all entered data to paper responses. Database data and primary care medical record extracts which were recorded on Microsoft Excel spreadsheets, were imported to Stata 17, where data cleaning and analyses were conducted. In Stata, utility values were estimated from EQ-5D-5L scores using the validated mapping function from the EQ-5D-3L (33, 34).

8.3.5.1 Methods for assessing acceptability and feasibility

Participant acceptability was assessed using questionnaire response rates and participant-reported completion time, by ModRUM version. The former is presented as the percentage of patients who consented to participate and completed ModRUM, and the latter using descriptive statistics. The impact of GP practice deprivation level and ModRUM version on the response rate was also considered using logistic regression. Participant feasibility was assessed using question completion rates and by reviewing issues participants experienced in answering ModRUM questions. Question completion rates are presented as the percentage of participants answering each question. Issues are presented descriptively, alongside the number of participants experiencing each issue.

The feasibility of using ModRUM data for costing purposes (researcher feasibility) was assessed based on data cleaning requirements and the ability to identify relevant unit costs for each resource. Rather than reporting information on cleaning and costing in this methods section, information is reported in section **8.4.3** under ‘researcher feasibility’, as assessment of the feasibility of cleaning and costing ModRUM data is an outcome of this chapter.

8.3.5.2 Methods for assessing construct validity

Construct validity was assessed via hypothesis testing including known-group analyses. In hypothesis testing, assuming that hypotheses are valid, hypotheses can be tested by assessing the extent to which scores obtained from an instrument are consistent with predicted hypotheses about the relationship between the instrument

and other variables or instruments (69). Known group validity asserts that an instrument is able to discriminate between groups where a difference in outcome is anticipated (69, 79). To assess known-group validity, Wilcoxon rank-sum (Mann-Whitney U) and Kruskal-Wallis H tests, described further below, were conducted to determine whether there were differences between groups (113). These tests are non-parametric methods, meaning they do not make the assumption that the outcome variable is normally distributed, which is the case for non-normally distributed total healthcare costs (113). The tests involve ranking participants by the outcome variable (total healthcare costs in this study) and examining whether medians of groups are different (113). The Wilcoxon rank-sum test was used when there were two groups (i.e. sex) and the Kruskal-Wallis H test was used when there were more than two groups (i.e. age group, number of long term conditions and age on leaving full time education). For the purposes of consistency in this known group validity assessment only, total cost estimates were based on the core questions, which were asked of all respondents, as ModRUM-C is nested within ModRUM-CD.

A generalised linear model (GLM) was employed to assess a hypothesised relationship, outlined further below, between quality-of-life scores and total healthcare costs. In this model, total costs were estimated using ModRUM-C or ModRUM-CD dependent on which version each participant completed as this could be adjusted for within the model. A number of other explanatory variables were also included in the model, including sex, age and GP practice deprivation score. Multiple model specifications were considered and compared using the Stata command *linktest*, histograms, percentile plots of deviance residuals and Akaike's information criterion. To assess the correlation between explanatory variables, the variance inflation factor (VIF) was estimated for explanatory variables and a correlation matrix was formed. A GLM model was implemented with an identity link function and gamma distribution to account for the positively skewed distribution of costs/units of resource use. A clustered sandwich estimator (vce(cluster) option in Stata) was used to obtain robust variance estimates that adjust for potential similarity of participants within GP practices. The model was specified as:

$$total\_cost = \beta_0 + \beta_1 mod\_version_i + \beta_2 sex_i + \beta_3 age\_over\_65_i + \beta_4 LTC_i + \beta_5 education_i + \beta_6 ethnic\_group_i + \beta_7 utility_i + \beta_8 deprivation_i$$

where *total_cost* was the total cost estimated using core or depth unit costs, dependant on the version of ModRUM the participant completed. Variable $mod\_version_i$ represents whether participant i completed ModRUM-C or ModRUM-CD. Categorical variables included sex (sex_i: female/male), age group ($age\_over\_65_i$: aged 65 and under, aged over 65), number of long-term conditions (LTC_i: none, one, more than one), age on leaving full time education ($education_i$: 16 or under, 17 or 18, 19 or over) and ethnic group ($ethnic\_group_i$: white/non-white). Continuous variables included EQ-5D-5L utility score ($utility_i$) and GP practice deprivation score ($deprivation_i$). Evidence of construct validity is provided if associations appear as hypothesised.

Prior to analysing data, I made several hypotheses about the relationship between total healthcare costs and patient characteristics and quality of life. I presented the hypotheses to the research team and the agreed hypotheses were as follows:

- Hypothesis 1: Older participants will have higher total healthcare costs than younger participants
    - Rationale: Previous research has demonstrated that increased age is associated with more healthcare use (189, 190).
- Hypothesis 2: Participants with more long-term conditions will have higher total healthcare costs than participants with no or one long-term condition.
    - Rationale: In England, 30% of the population who have long-term conditions account for 70% of spending (191). As the number of long-term conditions increases, so does healthcare and medication use (189). In a systematic review of 55 studies on the characteristics of high cost patients, the authors found that prevalence of chronic conditions and multimorbidity was common among high cost patients (192).
- Hypothesis 3: Participants with lower self-reported quality-of-life, as estimated using the EQ-5D-5L, will have higher total healthcare costs, than those with higher self-reported quality-of-life.
    - Rationale: Self-reported quality of life is indicative of an individual's health status. People who report lower quality of life are likely to be in

poorer health. People in poorer health are more likely to use healthcare and have higher healthcare costs.

The hypotheses above were framed on anticipated associations for members of the general population. It should be noted that in the interest of testing ModRUM with active healthcare users, patients were invited to the study if they had had a recent appointment at the GP practice. This criterion meant that the sample is a subset of the general population, who are active healthcare users. Furthermore, the observation of associations is dependent upon the sample size, with only strong associations likely to be observed in a small sample.

Potential associations were also explored for sex, age leaving full time education and GP practice deprivation level. For sex, while gender differences in health-seeking behaviours have been observed, with men less likely to seek care for physical and mental health conditions (193), the sample in this study is patients who have already sought healthcare, so they are not representative of all males in the general population. Furthermore, in another study while gender differences in primary consultation rates have been observed, which were only partially accounted for by reproductive reasons, once receipt of medication for similar underlying morbidities was controlled for, differences were almost eradicated (194). As education is an indicator of social disadvantage, and people who are socially disadvantaged are more likely to suffer ill health, resulting in increased healthcare consumption and cost, increased healthcare costs estimated using ModRUM, may be seen for participants who left full time education at a younger age (184). Participants registered at GP practices in more deprived areas may also be more socially disadvantaged and tend to be sicker suggesting they may need more healthcare (184). However, need does not necessarily indicate that they will consume more healthcare. Based on existing literature, I did not expect to see a difference based on ethnic groups. Morris et al. found that primary care use was higher for ethnic minorities while secondary care utilisation was lower; however, only some results were significant (195). As the sample included over 50,000 individuals, it is unlikely that in this study a significant association will be observed (195).

8.3.5.3 Methods for assessing criterion validity

Criterion validity assesses the extent to which the scores of an instrument agree with an accepted 'gold standard' that is considered to measure the construct of interest accurately (69, 79, 81). To test criterion validity, I estimated the level of agreement between healthcare resources in ModRUM that had corresponding data in available in GP medical records.

GP record data, which had initially been captured by the practices on EMIS Web, was provided in Microsoft Excel spreadsheets. This data was uploaded to Stata 17, where data cleaning was conducted. For consultations, data were provided with multiple rows for each appointment. It was possible to distinguish appointments on the same day if they were conducted by healthcare professionals with different roles (e.g. salaried general practitioner versus general medical practitioner) and/or by different modes of appointment (e.g. GP surgery versus telephone consultation). Due to the format of the data, if there were multiple rows with the same healthcare professional, via the same mode and on the same day, one consultation was assumed to have occurred. To minimise the work involved for GP practice staff, they were asked to extract data from 3 months prior to ModRUM completion date for the date the first participant answered ModRUM at their practice, until the ModRUM completion date of the last participant who completed ModRUM at their practice. Once I received the data, I kept only data for 3 months prior to completion date for each participant. Two participants did not provide the date they completed ModRUM, for these I used the median completion date of participants recruited at the same practice. Once I had formatted the medical record data, appropriate unit costs were sourced and applied.

First, I estimated the sensitivity and specificity to compare binary (yes/no) reporting for each resource. For resource use, sensitivity is the proportion of participants that have use of a resource recorded in their medical records, who also report using that resource in ModRUM (113). Specificity is the proportion of participants who have no use of a resource recorded in their medical records, that are correctly identified as not using the resource in ModRUM (113). Next, I estimated Lin's concordance correlation coefficient (CCC) for resource use and costs (114). Lin's CCC was selected over Pearson's correlation as it can be used to compare continuous, non-

normally distributed data. It incorporates measures of precision (Pearson's correlation) and accuracy (114). Lin's CCC is scaled between -1 and 1, where -1 indicates perfect reversed agreement and 1 indicates perfect agreement (114). The *concord* command was used in Stata to estimate Lin's CCC (196). Following previous studies assessing agreement between self-report and medical record data, Lin's CCC (p_c) was interpreted according to the following categories: poor (less than 0.40), fair (0.40 to 0.59), good (0.60 to 0.74) and excellent (0.75 to 1.00) (52, 146, 197).

To visualise differences, I produced Bland-Altman plots (198). For each participant, the differences between ModRUM and GP records for visits and costs were plotted on the y-axis against the mean of each pair (ModRUM and GP records) on the x-axis (198). The plot includes lines showing the mean difference and 95% limits of agreement. The mean difference shows whether there is systematic bias between methods, while the limits of agreement depict where 95% of differences between measurement methods are expected to lie, assuming normality of differences (198).

8.4 Results

In total, 717 patients were invited to participate in the study and 100 (14%) patients participated. Six further patients responded; however, they could not be included in the study because they did not return a consent form (n=2), did not provide consent for their data (medical notes and data collected during the study) to be used (n=2), did not sign the consent form (n=1) or they replied after the study had closed (n=1). More detailed information on response rates, such as by ModRUM version, is presented below in section **8.4.2**, which describes the acceptability of ModRUM.

8.4.1 Participant characteristics, quality of life and resource use

Participant-reported data was collected between November 2020 and March 2021, and GP medical record data was obtained between May and June 2021. Participant characteristics are presented in **Table 8.2**. Of the 100 participants recruited, 61% were female and 55% were aged 66 or over. Only one participant was recruited who was aged 30 or under. Most participants were of white ethnicity (95%). Over half of participants had at least one long term condition (58%). 44% of participants left full

time education aged 16 or under and 35% of participants left full time education aged 19 or over.

Quality of life scores, as estimated using the EQ-5D-5L, are presented in **Table 8.3**. The mean EQ-5D-5L utility score for all participants was 0.750 (SD: 0.249). On average, the score was slightly higher for participants who completed ModRUM-C, than for participants who completed ModRUM-CD (0.772 [SD: 0.212] versus 0.726 [SD: 0.285]).

Mean healthcare utilisation and costs are presented in **Table 8.4**, by ModRUM version. Unit cost sources for ModRUM are reported in **Table 8.9** and **Table 8.10**, as part of the researcher feasibility section, and for GP medical record data in **A6.6**. In both versions, remote consultations with a GP were the most commonly used resource (ModRUM-C: 1.90 contacts, ModRUM-CD: 1.84 contacts). Most resources had a similar number of mean contacts across ModRUM versions, with the exception of GP surgery contacts which were more common for ModRUM-CD (1.18 versus 0.60). Other healthcare professional contacts were higher for ModRUM-C; however, once other healthcare professional and nurse contacts were added for ModRUM-CD, the number of contacts were similar. The mean total cost was higher for ModRUM-CD (£537 (SD: £1045) versus £462 (SD: £802)). The large standard deviation for both versions reflects that a minority of patients had costly inpatient stays.

Table 8.2 Participant characteristics

	n	(%)
Sex		
Female	61	(61)
Male	39	(39)
Age group		
18-30	1	(1)
31-45	14	(14)
46-55	15	(15)
56-65	15	(15)
66-75	32	(32)
76 or over	23	(23)
Ethnic group		
Asian/Asian British	2	(2)
Mixed/Multiple ethnic groups	1	(1)
Other	1	(1)
Prefer not to say	1	(1)
White	95	(95)
Long term conditions		
More than one	34	(34)
One	24	(24)
None	39	(39)
Missing	3	(3)
Age on leaving full-time education		
16 or under	44	(44)
17 or 18	18	(18)
19 or over	35	(35)
Prefer not to say	1	(1)
Missing	2	(2)

Table 8.3 EQ-5D-5L scores

	n	(%)	Mean	(SD)	[Min, Max]
All (N=100)					
Utility score	99	(99)	0.750	(0.249)	[-0.227, 1]
VAS scale score	83	(83)	77	(20)	[0, 100]
VAS box score	96	(96)	76	(20)	[0, 100]
Core (N=53)					
Utility score	52	(98)	0.772	(0.212)	[-0.200, 1]
VAS scale score	45	(85)	77	(21)	[0, 100]
VAS box score	51	(96)	77	(20)	[0, 100]
Core plus depth (N=47)					
Utility score	47	(100)	0.726	(0.285)	[-0.227, 1]
VAS scale score	38	(81)	77	(19)	[0, 97]
VAS box score	45	(96)	75	(21)	[0, 100]

VAS: Visual analogue scale

Table 8.4 Healthcare utilisation and costs, by ModRUM version

	Core module (N=53)						Core module with depth questions (N=47)					
			Resource use		Cost (£)				Resource use		Cost (£)	
	n	(%)	Mean	[Min, Max]	Mean	(SD)	n	(%)	Mean	[Min, Max]	Mean	(SD)
A&E	53	(100)	0.15	[0, 1]	25	(60)	47	(100)	0.17	[0, 2]	28	(72)
Ambulance (convey)	-	-	-	-	-	-	47	(100)	0.06	[0, 1]	13	(52)
Ambulance (treat)	-	-	-	-	-	-	47	(100)	0.04	[0, 1]	11	(53)
Outpatient (f2f)	53	(100)	0.58	[0, 6]	72	(132)	46	(98)	0.61	[0, 5]	78	(158)
Outpatient (remote)	53	(100)	0.36	[0, 3]	28	(52)	46	(98)	0.43	[0, 4]	35	(71)
Day case	53	(100)	0.17	[0, 4]	128	(486)	46	(98)	0.07	[0, 1]	72	(322)
Inpatient stays	53	(100)	0.08	[0, 2]	101	(493)	47	(100)	0.06	[0, 1]	281	(1162)
GP (surgery)	52	(98)	0.60	[0, 4]	20	(29)	45	(96)	1.18	[0, 7]	39	(41)
GP (remote)	52	(98)	1.90	[0, 10]	49	(42)	44	(94)	1.84	[0, 12]	47	(48)
GP (home)	52	(98)	0.08	[0, 4]	6	(47)	44	(94)	0.00	[0, 0]	0	(0)
Nurse (surgery)	-	-	-	-	-	-	45	(96)	0.84	[0, 5]	8	(10)
Nurse (remote)	-	-	-	-	-	-	44	(94)	0.20	[0, 2]	1	(2)
Nurse (home)	-	-	-	-	-	-	44	(94)	0.41	[0, 14]	16	(84)
Other HCP (surgery)	52	(98)	1.00	[0, 7]	16	(27)	47	(100)	0.28	[0, 4]	14	(51)
Other HCP (remote)	52	(98)	0.33	[0, 2]	6	(11)	43	(91)	0.35	[0, 5]	8	(28)
Other HCP (home)	52	(98)	0.15	[0, 7]	6	(39)	46	(98)	0.09	[0, 2]	7	(32)
Medications	-	-	-	-	-	-	45	(96)	-	-	67	(105)
Mean total cost	52	(98)	-	-	462	(802)	38	(81)	-	-	537	(1045)

*HCP: healthcare professional

8.4.2 Acceptability

Overall the response rate was 14%. The response rate was higher for patients invited to complete ModRUM-C. For ModRUM-C, 268 patients were invited, and 53 (20%) patients participated. For ModRUM-CD, 449 patients were invited, and 47 (10%) patients participated. Using logistic regression, after controlling for practice deprivation score, for patients who received ModRUM-C, the odds of taking part were 1.74 times as large as for patients who received ModRUM-CD (95% confidence interval (CI): 1.12 to 2.72, p=0.014). **Table 8.5** presents response rates by GP practice. In addition to varying by the version of ModRUM completed, response rates also varied by practice deprivation score, with lower response rates from patients registered at practices in more deprived areas. After controlling for ModRUM version, a one-unit improvement in the deprivation level of the GP practice, meant that the odds of patients participating increased by a factor of 1.11 (95% CI: 1.04 to 1.19, p=0.003).

The mean and median participant-reported ModRUM completion times were similar for both versions (**Table 8.6)**. The maximum reported completion time of 25 minutes was reported for ModRUM-C, which possibly indicates that some participants included time spent completing all documents in the mail pack. All other times reported were 12 minutes or less. Once the 25-minute outlier was omitted, the mean completion time for ModRUM-C reduced to 4.9 minutes, compared with 5.7 minutes for ModRUM-CD; however, this did not alter the median time, which was 5 minutes for both versions of ModRUM.

Table 8.5 Response rates, by GP practice

Practice	Practice deprivation score	ModRUM version	Number of patients invited	Number of patients participating (%)
1	2	Core with depth	160	10 (6)
2	5	Core with depth	159	17 (11)
3	4	Core	119	18 (15)
4	10	Core	149	35 (23)
5	10	Core with depth	130	20 (15)
Total			717	100 (14)

Table 8.6 Participant-reported time to complete ModRUM

	n	Mean minutes (SD)	Median minutes (IQR)	[Min, Max]
Core	42	5.4 (4.3)	5.0 (3.0-6.0)	[1, 25]
Core with depth	40	5.7 (3.8)	5.0 (2.5-8.0)	[1, 12]

8.4.3 Feasibility

8.4.3.1 Participant feasibility

Prior to cleaning the data, question completion rates for ModRUM items ranged from 96 to 100 percent for ModRUM-C and 91 to 100 percent for ModRUM-CD. Question completion rates are reported in **Table 8.7** and relate to top level questions, as opposed to details reported in tables within ModRUM.

In **Table 8.8**, I have summarised the issues that occurred when participants completed the questions. Despite not being an issue in terms of analysis, as most responses were consistent, 19% of participants completed both the tick box and 'how many?' box for at least one question. One participant missed an entire page of questions when completing ModRUM-C, while seven participants who completed ModRUM-CD missed at least one page. Of these seven, two participants reported missed questions (GP/nurse contacts) under the other healthcare professional question, suggesting they may not have seen the questions, as opposed to missing them intentionally. One participant, who completed ModRUM-C, recorded only positive responses to the questions, with responses to other questions left missing.

For ModRUM-CD, five participants did not complete the tick box question, which was the top-level question asking how many times a resource was used, but they did complete the tables below, where more detail on the resource use was required. This meant that top-level answers could often be inferred from answers in the tables. Two participants recorded remote outpatient appointments under the face-to-face question. Five participants either missed the number of times a medication was prescribed, or reported an answer in a different metric to what was asked for (e.g. 'every 2 months', instead of one or two times during the three month recall period).

Several participants who completed ModRUM-CD reported issues with the pre-paid return envelope. The size of the envelope provided was the only option provided by the mailing company; however, given the high-quality paper and additional pages of ModRUM-CD, the study documents only just fitted in the provided pre-paid envelope. Several participants returned ModRUM-CD in their own envelope. The envelope being an insufficient size may have negatively impacted the response rate to ModRUM-CD.

Table 8.7 Question completion rates, by ModRUM version

Core module (N=53)	Pre-cleaning		Post-cleaning	
	n	(%)	n	(%)
A&E	52	(98)	53	(100)
Outpatient (f2f)	52	(98)	53	(100)
Outpatient (remote)	52	(98)	53	(100)
Day case	52	(98)	53	(100)
Inpatient stays	53	(100)	53	(100)
GP (surgery)	51	(96)	52	(98)
GP (remote)	52	(98)	52	(98)
GP (home)	51	(96)	52	(98)
Other HCP (surgery)	52	(98)	52	(98)
Other HCP (remote)	52	(98)	52	(98)
Other HCP (home)	51	(96)	52	(98)
Core module with depth questions (N=47)	**Pre-cleaning**		**Post-cleaning**	
	n	**(%)**	**n**	**(%)**
A&E	47	(100)	47	(100)
Ambulance (convey)	47	(100)	47	(100)
Ambulance (treat)	47	(100)	47	(100)
Outpatient (f2f)	46	(98)	46	(98)
Outpatient (remote)	45	(96)	46	(98)
Day case	46	(98)	46	(98)
Inpatient stays	47	(100)	47	(100)
GP (surgery)	43	(91)	45	(96)
GP (remote)	44	(94)	44	(94)
GP (home)	44	(94)	44	(94)
Nurse (surgery)	44	(94)	45	(96)
Nurse (remote)	44	(94)	44	(94)
Nurse (home)	44	(94)	44	(94)
Other HCP (surgery)	45	(96)	47	(100)
Other HCP (remote)	43	(91)	43	(91)
Other HCP (home)	46	(98)	46	(98)
Medications	43	(91)	45	(96)

*HCP: healthcare professional

Table 8.8 Question completion issues, by ModRUM version

	Number of participants (%)
Core module only (N=53)	
Used both tick box and 'how many' box for at least one question	
Answers consistent	12 (23)
Answers inconsistent	1 (2)
Full page(s) without responses	1 (2)
Only provided answers when resources used	1 (2)
Core module with depth questions only (N=47)	
Used both tick box and 'how many' box for at least one question	
Answers consistent	7 (15)
Answers inconsistent	0 (0)
Full page(s) without responses	7 (15)
Missed questions included under other questions e.g. GP included in other HCP	2 (4)
Tick box answer missing	
Answer can be inferred from table response	4 (9)
Answer cannot be inferred from table response (other HCP)	1 (2)
Outpatient appointments	
Remote included under face-to-face, but crossed out and rewritten	1 (2)
Remote included under face-to-face, note to refer to previous question added	1 (2)
Day case reported under outpatient and repeated under day case	1 (2)
Other healthcare professional contacts	
GP included in the other HCP table	1 (2)
Used one row for each appointment to the same HCP	1 (2)
Number of times missed in table, can be inferred from tick box	1 (2)
Tick box answer inconsistent with table answers	1 (2)
Prescribed medications	
Number of times missed in the table for some/all medications	2 (4)
Every [X] months/when needed/ongoing/constant repeat prescriptions reported instead of number of times	3 (6)

*HCP: healthcare professional

8.4.3.2 Researcher feasibility

Data cleaning and analysis were performed in Stata 17. To minimise data missingness, several assumptions were made. Question completion rates post-cleaning are presented in **Table 8.7**. Minimal cleaning was required for ModRUM-C. Where one participant only reported positive answers, the unanswered questions were assumed to be zero. More cleaning was required for ModRUM-CD. This included inputting values for tick boxes where answers were only reported in the tables and moving answers that were in the incorrect position to the relevant question (e.g. moving GP contacts from other healthcare professional to GP, when the GP question was missed and the appointment listed was clearly a GP appointment). Multiple imputation of missing data was not performed, as such, the analysis only includes participants who provided complete data or for whom simple imputation, using the assumptions above, could be performed.

Appropriate national unit costs were located for all participant-reported resources in both versions of ModRUM. Unit costs used to value ModRUM-C resources are reported in **Table 8.9**. All unit costs were for the year 2019, where unit costs were not available for 2019, past costs were inflated to 2019 prices using the NHS cost inflation index (66). For the ModRUM-C, locating and applying unit costs took considerably less time, as the same unit cost was used for each resource across all participants. More detailed information was provided for questions on outpatient appointments, day case and inpatient stays, other healthcare professional consultations and prescribed medications in ModRUM-CD. Unit costs used to value resources in ModRUM-CD are reported in **Table 8.10**. Some items in ModRUM-CD did not differ from items in ModRUM-C (A&E, GP), as such, costs reported in **Table 8.9** were used for these resources.

The design of ModRUM means that questions that appear in ModRUM-C are embedded in ModRUM-CD, where for most items, the core question is the top-level question in ModRUM-CD, with a table below to record further details. The design meant that for participants who completed ModRUM-CD, I could compare how estimated costs differed dependent on what level of detail was used for costing (**Table 8.11**). Costs were higher for all but one resource when unit costs sourced for costing ModRUM-CD data were used, and in total they were 41 percent higher. The

largest contributors to this difference were hospital inpatient and day case admissions, for which this sample included three participants who had inpatient admissions and three participants who had day case admissions. The results here suggest that the unit costs used for these resources when costing using core level data were underestimates, as more detailed data led to the use of higher unit costs.

Table 8.9 Unit costs for the core module

Healthcare resource	Unit cost (£)	Source detail
A&E	166.05	Index tab: Accident and Emergency (48).
Hospital outpatient appointment		
Face-to-face	124.81	Weighted average of CL and NCL tabs, excluding non-face-to-face and paediatric/child only items (48).
Non-face-to-face	77.89	Weighted average of CL and NCL tabs, excluding face-to-face and paediatric/child only items (48).
Hospital day case	755.59	Weighted average of DC tab, excluding paediatric (48)
Hospital inpatient stay (per night)	535.41	Weighted average EL, NEL and NES tabs, excluding paediatric/child only items (48). Duration based on HES data: mean episode durations, including aged 20 and over (5.85 nights) (199).
General practitioner		
GP surgery	33.19	9.22 minute consultation (including direct care costs, excluding qualification costs) (66).
Home	84.24	11.4 minute home visit, 12 minutes travel time, £3.60 per minute of patient contact (66, 200).
Online/telephone	25.56	7.1 minute telephone call, £3.60 per minute of patient contact (66, 200).
Other healthcare professional		
GP surgery	16.28	Assume Band 5 nurse: £63 per hour patient related work, 15.5 minutes (66, 201).
Home	39.68	Assume a district nurse: CHS tab, district nurse, adult, face-to-face (48).
Online/telephone	18.28	Assume NHS 111 call or community health services: Average of NHS 111 (price reported, inflated using the NHS cost inflation index (66, 202)) and CHS tab, weighted average of all adult, non-face-to-face currency description (48).

Table 8.10 Unit costs for core module with depth questions responses

Healthcare resource	Unit cost (£)	Source detail
Ambulance		
See and treat or refer	209.38	AMB tab: See and treat or refer (48).
See and treat and convey	257.34	AMB tab: See and treat and convey (48).
Hospital outpatient appointment		
Face-to-face	Varies	Costed by test/procedure, clinic type and reason (48).
Non-face-to-face	Varies	Costed by clinic type and reason (48).
Hospital day case	Varies	Costed by test/procedure, department and reason (48).
Hospital inpatient stay	Varies	Costed by test/procedure, department and reason, with number of nights used to indicate long or short stay (48).
Nurse		
GP surgery	16.28	Assume Band 5 nurse: 63 per hour patient related work, 15.5 minutes (66, 201).
Home	39.68	Assume a district nurse: CHS tab, district nurse, adult, face-to-face (48).
Online/telephone	6.30	Assume Band 5 nurse: 63 per hour patient related work, same duration as specialist nurse (66, 201).
Prescribed medications	Varies	Costed using name and dose (where provided) using cost per item (203).

Table 8.10 continued

Healthcare resource	Unit cost (£)	Source detail
Other healthcare professional		
Surgery, health or walk-in centre		
Occupational therapist	83.17	AHP tab: Occupational therapist, Adult, One to One (48).
Physiotherapist	62.90	AHP tab: Physiotherapist, Adult, One to One (48).
Midwife	62.84	HVM tab: Community midwife, weighted average of ante and post-natal visits (48).
Mental health services	57.42	IAPTMHCC and IAPTMHCCIA tabs: weighted average (48).
Flu jab	9.82	Assumed practice nurse: (as above).
X-Ray/scan at walk-in centre	72.54	AE tab: Weighted average of Type 03 and 04 non admitted investigation (48).
Home		
Occupational therapist	83.17	AHP tab: Occupational therapist, Adult, One to One (48).
Physiotherapist	62.90	AHP tab: Physiotherapist, Adult, One to One (48).
Speech therapist	106.51	AHP tab: Speech and language therapist, Adult, One to One (48).
NHS 111 call	9.00	NHS 111 call: inflated using the NHS cost inflation index (66, 202).
Online/telephone		
Occupational therapist	83.17	AHP tab: Occupational therapist, Adult, One to One (48).
Physiotherapist	62.90	AHP tab: Physiotherapist, Adult, One to One (48).
Pharmacist	8.90	Assume Band 6: 89 per hour patient related work, assume same duration as specialist nurse, 6 minutes (66, 201).

Table 8.11 Costing comparison for participants who completed ModRUM-CD

	n	Mean resource use	ModRUM-CD costs (£) Mean	(SD)	ModRUM-C costs (£) Mean	(SD)
Outpatient (f2f)	46	0.61	78	(158)	75	(139)
Outpatient (remote)	46	0.43	35	(71)	33	(66)
Day case	46	0.07	72	(322)	49	(189)
Inpatient stays	47	0.06	281	(1162)	228	(1132)
Other HCP (surgery)[12]	45	1.04	17	(36)	10	(13)
Other HCP (remote)[12]	41	0.59	9	(28)	11	(20)
Other HCP (home)[12]	43	0.51	24	(94)	20	(88)
Total	39		347	(956)	246	(550)

1: HCP=healthcare professional; 2: Depth=nurse plus other healthcare professional

8.4.4 Construct validity

The results from known-group analyses are presented in **Table 8.12**. There was no evidence of a difference in total healthcare costs between sex, age and age on leaving full time education groups. However, there was good evidence against the null hypothesis that median total healthcare costs are the same irrespective of number of long-term conditions, which suggests total costs differ dependent on number of long-term conditions ($p<0.05$).

Hypothesis testing provided some support for the construct validity of ModRUM (**Table 8.13**). Results from assessments of model fit were as follows: the link function was tested and results suggested that the model was correctly specified, with no relevant variables omitted and the correct link function included. The variance inflation factor was estimated to test for multicollinearity. All variance inflation factors were less than 1.64, suggesting low correlation between explanatory variables. Estimating correlation coefficients, in a correlation matrix also confirmed this result. A lower AIC, in comparison with other model specifications, indicated a better model fit. Total healthcare costs as estimated using ModRUM, were negatively associated with health-related quality of life ($p<0.001$); in other words, participants with higher self-reported healthcare costs, reported lower EQ-5D-5L scores. Total healthcare costs

were positively associated with GP practice deprivation score (p<0.001), with increased healthcare costs observed for participants registered at GP practices in less deprived areas. In contrast to the hypothesised relationship between age and total healthcare costs, when other variables were controlled for, the results from the model suggest that on average participants aged over 65 have lower total healthcare costs that those aged 65 and under (p=0.002).

Table 8.12 Known-group analyses: rank test results

Hypothesis	Known groups	n	Rank sum[1]	p-value
1	Sex			
	Female	54	2379.0	0.521
	Male	36	1716.0	
2	Age group			
	18-30	1	29.5	
	31-45	14	745.5	
	46-55	12	596.5	0.538
	56-65	15	624.5	
	66-75	26	1251.5	
	76 or over	22	847.5	
3	Number of long-term conditions			
	None	36	1299.0	
	One	20	821.0	0.013
	More than one	30	1621.0	
4	Age on leaving full time education			
	16 or under	37	1597.5	
	17 or 18	16	620.0	0.618
	19 and over	33	1523.5	

1: higher rank equates to higher total resource-use costs

Table 8.13 Hypothesis testing: generalised linear regression results

	N[1]	Adjusted cost (£) Marginal mean (95% CI)		Mean difference (95% CI)		p-value
ModRUM version						
Core module	47	639	(454 to 823)			
Core module with depth questions	35	364	(187 to 541)	-275	(-410 to -140)	<0.001
Sex						
Male	34	580	(421 to 739)			
Female	48	480	(293 to 667)	-100	(-180 to 31)	0.049
Age group						
65 and under	40	632	(467 to 797)			
Over 65	42	416	(220 to 612)	-216	(-350 to -82)	0.002
Ethnic group						
Non-white	3	479	(244 to 715)			
White	79	523	(354 to 692)	44	(-133 to 220)	0.629
Number of long-term conditions						
None	36	287	(243 to 331)			
One	19	348	(227 to 468)	61	(-53 to 175)	0.298
More than one	27	956	(409 to 1503)	670	(87 to 1,252)	0.024
Age on leaving full time education						
16 or under	33	477	(337 to 617)			
17 or 18	16	415	(212 to 619)	-62	(-135 to 12)	0.102
19 or over	33	617	(424 to 811)	140	(40 to 241)	0.006
EQ-5D-5L score[2]	82			-47	(-57 to -36)	<0.001
GP practice deprivation score[3]	82			22	(15 to 29)	<0.001

1: Including only participants for whom total cost could be estimated (imputation was not performed), 2: Rescaled to increments of 0.1, 3: On a scale of 1 to 10, where 1 is most deprived and 10 is least deprived

8.4.5 Criterion validity

GP medical records were obtained for 99 participants. Records were unavailable for one participant as they had left the GP practice they were registered at when invitations were sent out.

8.4.5.1 Sensitivity and specificity

Sensitivity and specificity were estimated for GP and other healthcare professional contacts, and prescribed medications (**Table 8.14**). Sensitivity was high across all resources (>0.83), indicating that participants were likely to report use of a resource, when it was indicated in the medical records that they had used the resource. Specificity estimates were more variable, with the lowest value for GP contacts, which was also the most commonly utilised resource amongst participants (i.e. few patients had zero contacts, which was expected given the inclusion criteria of a clinical appointment at the GP practice in the last four weeks). Lower values of specificity could indicate that ModRUM is picking up resource use not captured in the medical records, or it could be a result of incorrectly reporting resource usage when it had not occurred within the recall period (telescoping). When compared with healthcare professional contacts, specificity for prescribed medications was relatively high at 0.88 (95% CI: 0.47 to 1.00).

Table 8.14 Sensitivity and specificity, ModRUM versus medical record data

				Sensitivity (95% CI)	Specificity (95% CI)
General practitioner contacts					
		ModRUM			
		Yes	No		
Medical record	Yes	80	2	0.98 (0.92 to 1.00)	0.33 (0.10 to 0.65)
	No	8	4		
Other healthcare professional contacts					
		ModRUM			
		Yes	No		
Medical record	Yes	41	8	0.84 (0.70 to 0.93)	0.55 (0.39 to 0.70)
	No	19	23		
Prescribed medications[1]					
		ModRUM			
		Yes	No		
Medical record	Yes	35	1	0.97 (0.86 to 1.00)	0.88 (0.47 to 1.00)
	No	1	7		

1: prescribed medication question only included in ModRUM-CD

8.4.5.2 Agreement between ModRUM and medical records

In **Table 8.15,** I have presented mean contacts and costs as estimated using ModRUM and GP record data. Mean resource use and costs were higher in ModRUM than the medical records for GP and other healthcare professional contacts. For GP contacts, the mean difference in contacts was 0.39 (95% CI: 0.06 to 0.73), while the mean difference in costs was £16.05 (95% CI: £4.74 to £27.35). For other healthcare professional contacts, the mean difference in contacts was 0.62 (95% CI: 0.13 to 1.10), while the mean difference in costs was £23.70 (95% CI: £7.76 to £39.65). The estimated mean cost of prescribed medications was 42% higher for GP medical record data than ModRUM data.

Based on Lin's CCC, there was a good level of agreement between ModRUM and medical records for GP contacts and costs, and prescribed medication costs (**Table 8.15**). There was a poor level of agreement for other healthcare professional contacts and costs. Bland-Altman plots are presented for GP and other healthcare professional contacts and costs in **Figure 8.2** and for prescribed medication costs in **Figure 8.3**.The line y=0 represents perfect agreement between data sources, while the difference between this line and the observed average agreement represents bias. In **Figure 8.2**, for all plots, the observed average agreement is above zero, suggesting ModRUM estimates are on average larger than estimates based on GP record data. For other healthcare professionals contacts and costs, where the mean is above three, it is evident that larger values in ModRUM are driving the bias, potentially suggesting ModRUM is capturing healthcare not captured in GP records. For prescribed medication cost, the observed average agreement is below zero, suggesting ModRUM estimates are on average smaller than estimates based on GP record data. No trend is observed based on mean cost. The 95% limits of agreement indicate the ranges of differences between data sources to be expected for each individual. A smaller range was observed for GP than other healthcare professional contacts, while for costs, the largest range was for prescribed medications.

Table 8.15 Agreement between ModRUM and medical record healthcare contacts and costs, by item

	n	ModRUM		Medical records		Mean difference		p_c*	95% limits of agreement
		Mean	(SD)	Mean	(SD)	(95% CI)			
General practitioner									
Contacts	94	2.79	(2.15)	2.39	(2.09)	0.39	(0.06 to 0.73)	0.693	(-2.80 to 3.58)
Cost (£)	94	80.31	(69.09)	64.27	(57.14)	16.05	(4.74 to 27.35)	0.602	(-92.12 to 124.21)
Other HCP									
Contacts	91	1.67	(2.36)	1.05	(1.29)	0.62	(0.13 to 1.10)	0.224	(-3.98 to 5.21)
Cost (£)	91	36.40	(74.65)	12.69	(20.57)	23.70	(7.76 to 39.65)	0.021	(-126.35 to 173.76)
Prescribed medications									
Cost (£)	44	68.54	(105.74)	97.38	(166.31)	-28.84	(-60.71 to 3.02)	0.702	(-234.26 to 176.58)

* p_c: Lin's concordance correlation coefficient

Figure 8.2 Bland-Altman plots (primary and community care contacts and costs)

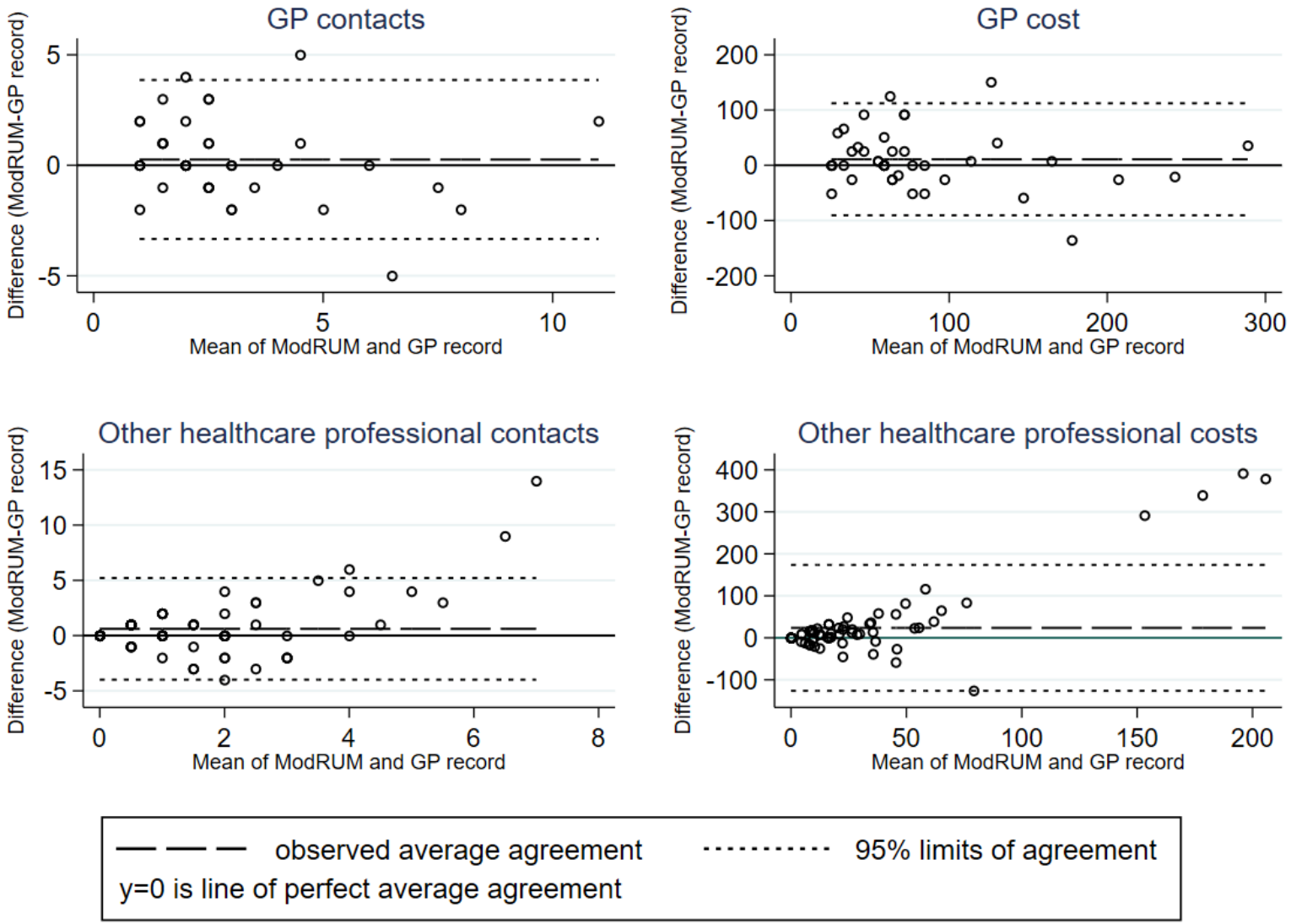
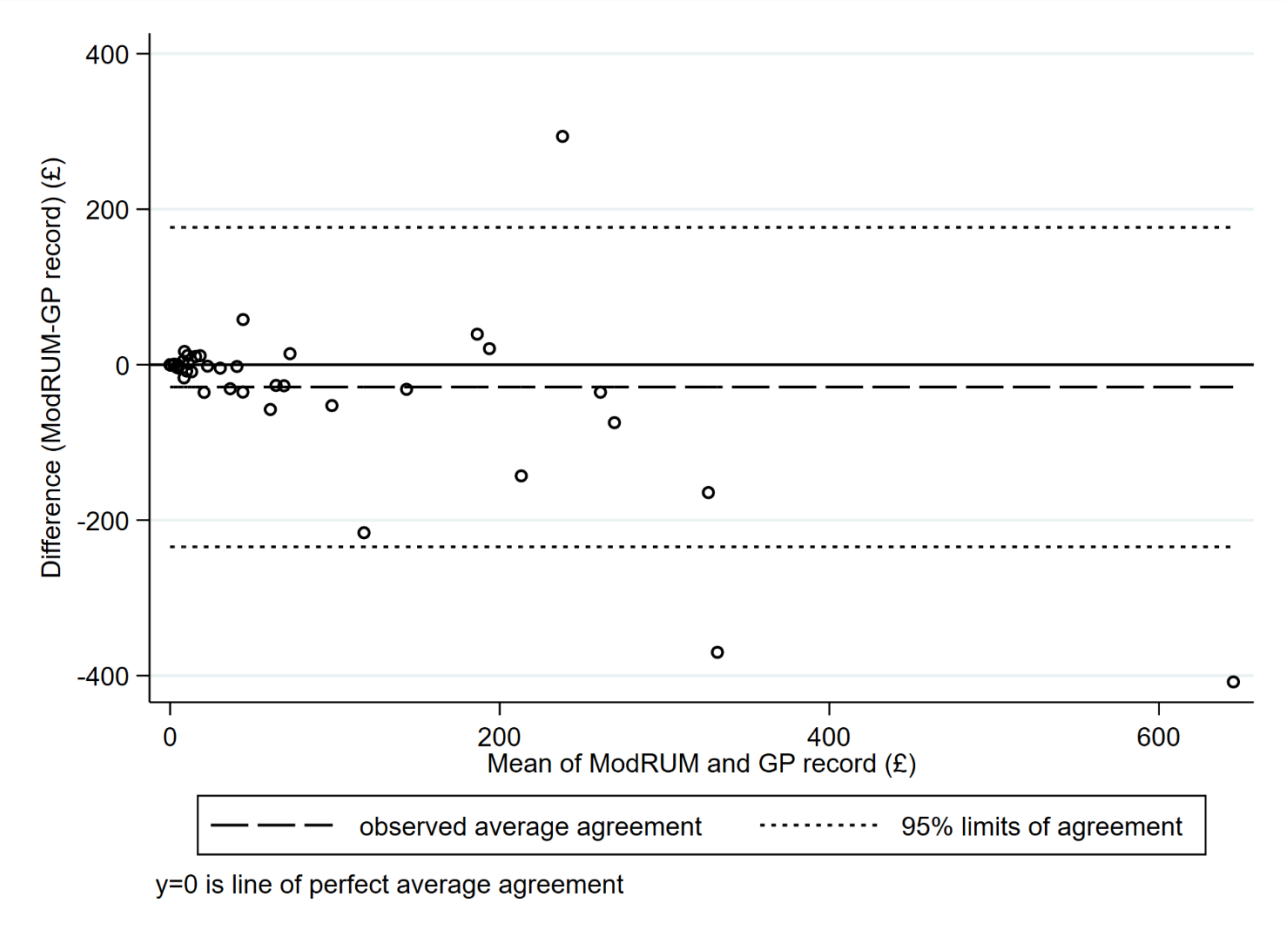


Figure 8.3 Bland-Altman plot (prescribed medication cost)



8.5 Discussion

8.5.1 Main findings

In this study, ModRUM was piloted with 100 patients. Despite completion times being similar, based on response rates, ModRUM-C appeared to be more acceptable to patients than ModRUM-CD. The results of this study provide some evidence for the validity of ModRUM for collecting resource-use data from patients recruited in a primary care setting. Validity was demonstrated through hypothesis testing, including known-group analyses, and comparison of ModRUM results to GP medical records.

8.5.2 Strengths and weaknesses

Despite delays and making amendments to this study due to the Covid-19 pandemic, I successfully piloted ModRUM with 100 patients recruited from a primary care setting. I was able to assess the acceptability of ModRUM through response rates and self-reported completion time. Overall, the response rate was 14%, which was higher for patients who were invited to complete ModRUM-C than for patients who were invited to complete ModRUM-CD (20% versus 10%). These rates were consistent with previous research on response rates to postal surveys (186). The response rate does not reflect the likely response rate to ModRUM in a trial-based economic evaluation, as the recruitment method was not consistent with an RCT, where participants are likely to be considerably more engaged. The response rate may have also been negatively impacted by the Covid-19 pandemic. Invitations were sent during winter 2020/21, which included periods when England was in national lockdown, meaning people may have been less able or more reluctant to participate. For example, patients may have been willing to participate, but due to guidance for vulnerable people to shield at home, were not able to post the study documents. Question completion rates were at least 96% for ModRUM-C and 91% for ModRUM-CD. The sample included a sufficient number of participants to test and provide preliminary evidence for the feasibility of costing resource-use data reported in ModRUM in this group.

Most participants completed the patient characteristics form and the EQ-5D-5L, which meant that I could assess construct validity. Construct validity was assessed using established statistical techniques. Results were compared with pre-defined

hypotheses and support for the validity of ModRUM was obtained for hypotheses made regarding health-related quality of life and number of long-term conditions. Participants with lower EQ-5D-5L scores had higher total healthcare costs ($p<0.001$). Participants with long-term conditions had higher total healthcare costs ($p=0.013$). While total healthcare costs were higher for participants who completed ModRUM-CD based on raw costs, in the GLM model, where participant characteristics and GP practice deprivation score were controlled for, the opposite was observed, with ModRUM-C costs higher ($P<0.001$). While I hypothesised that older patients would have higher total healthcare costs, in the regression model, I found that younger participants had higher total healthcare costs ($p=0.002$). It is likely that this result was impacted by the sampling strategy, where to be recruited to the study patients required a recent appointment at their GP practice, and also by other characteristics being controlled for in the model, including long term conditions and quality of life. In contrast to the relationship anticipated, in the regression model I found that on average participants registered at practices in less deprived areas had higher healthcare costs ($p<0.001$).

With the exception of one participant who had left their GP practice, I was able to obtain medical record data for all participants and used this to assess criterion validity. Criterion validity could only be assessed for a subset of questions included in ModRUM, where corresponding data were available in the medical records. This included other primary and community-based healthcare professionals; however, these are unlikely to be comprehensively covered in the medical records, with services such as NHS 111 captured in ModRUM, but not in primary care medical records. With good agreement observed for GP contacts and costs and prescribed medication costs, and results on average higher for GP and other healthcare professional contacts and costs in ModRUM, this study provides support for using ModRUM as an alternative to medical records for resource-use data.

While the sample was not as large as originally desired, this study was successfully adapted and conducted during the Covid-19 pandemic. I was fortunate that the GP practices I had initially recruited had the capacity to increase the time they spent on this study to assist with the identification and recruitment of patients. I was able to adopt several strategies to maximise recruitment, such as invitations being sent via

patients' GP practices and increasing personalisation of the information sheet by including a picture of myself (187). Other strategies to increase the response rate, such as reminders, may have increased the response rate further; however, as I did not have access to patient details, this would have needed to have been led by the GP practice, adding burden on the practices, and requiring additional funds.

As I was unable to compare the characteristics of responders to non-responders, it is unclear whether the sample is representative of the population, which with a 14% response rate, may impede the external validity of the findings. Comparing the sample population with 2020 UK population estimates, males (39% versus 48%) and people aged under 66 (45% versus 18%) were underrepresented in this study (204). As anticipated, practice deprivation level was associated with whether a patient participated. Increasing the number of invites sent from practices in more deprived practices meant that I had representation from these practices. The sample included slightly more participants (55%) from the GP practices in the least deprived areas and more female participants (61%). Participants were mostly balanced across categories for number of long-term conditions and age on leaving full time education. Equal representation was not achieved across age groups, with only one participant recruited from the 18 to 30 age group and 55% of participants aged 66 or over. However, this was expected based on the identification process, where patients were required to have had a clinical appointment at their GP practice in the last four weeks. Also, having a larger proportion of older participants is likely to be more representative of the ages of people participating in many trials. Representation from patients from non-white ethnic groups was low, but again this was expected based on ethnicity estimates across the participating practices (**Table 6.2**) (179).

Questions on remote consultations were added after qualitative testing; as such, the content and face validity of these questions were not assessed. The new questions were kept consistent with questions already included, with respect to wording and format. Despite this, I identified that two participants included remote outpatient visits under the face-to-face outpatient question. The participants marked or rectified where this had occurred; however, to minimise participant burden and double counting, additional explanation could be included to explain that there are separate in person and remote outpatient questions.

As I was unable to recruit patients in person, I used a hybrid mail service. On reflection, response rates and question completion rates, particularly for ModRUM-CD, may have improved if I had ordered samples, prior to sending copies to patients. First, due to the high-quality paper and the number of pages to return, the pre-paid envelope was not large enough for participants who received ModRUM-CD. A number of patients added notes to say the envelope was too small or they used their own envelope to return their documentation, which may have resulted in patients who had completed the questionnaire not responding. Second, the documentation was also sent as individual sheets. A booklet may have improved question completion rates, as it was evident that where participants had missed entire pages of ModRUM-CD, it was likely that the participant had not seen the question as opposed to missing it due to being unable to recall the data and respond.

As mentioned above, only some items could be validated against GP medical records. In my initial NHS ethics application, I had planned to perform a manual search to identify other resource use, such as letters to the practice informing of secondary care utilisation; however, due to stricter research governance rules on accessing patient data this was not permitted. While this task could have been performed by a member of the GP practice team, I did not have the funds for this, and my contacts expressed uncertainty on the completeness of this data. An alternative option for validation would be to access this data elsewhere, such as HES; however, again cost and time constraints meant that this was not a viable option.

Constraints associated with amending the study so that it was conducted remotely meant that I dropped plans to assess reliability and to compare recall lengths. I had planned to conduct a test-retest reliability study to assess reliability by posting a second ModRUM to participants, to be completed a week following initial ModRUM completion. Assuming recruitment was running as planned, I was also planning to administer a three-month recall period version of ModRUM to the first 200 patients recruited, and a six-month recall period version to the second 200 patients recruited. These recall periods are commonly used periods in existing economic evaluations.

Having both would have allowed me to compare psychometric properties between them.

8.5.3 Comparison to existing literature

In **Chapter 3**, I summarised literature describing the piloting of RUMs. A range of methods were used, including cognitive interviews (121), which were similar to the study I conducted and reported in **Chapter 6**; testing of an interview-administered RUM with follow-up feedback questions [42]; and piloting of a postal patient-completed RUM, where follow-up debriefing questions were subsequently sent in the post (130). This study followed other pilot studies where the aim was to assess patient burden via completion time (136) and acceptability via data completeness, response rates and/or missingness (123, 130, 135, 146).

Construct validity, including known-group and convergent, were assessed by Ness et al. for the Multiple Sclerosis Health Resource Utilization Survey (142). Known-group validity was assessed by patients categorised into four quarters of the Expanded Disability Status Score, and the authors found significant differences in healthcare costs between groups (142). Convergent validity was assessed by comparing health costs to patient-reported disability, lost ability to partake in daily routine and activities and health-related quality of life (142). All results were significant, including health-related quality of life, which was negatively associated with total costs, which is consistent with the result found in this study (142).

Noben et al. reviewed studies reporting comparisons of self-report and administrative data and rated the methodological reporting quality of the 16 included studies (50). They concluded that *"the exchangeability of self-reported and administrative health care resource use measurements can only be cautiously supported based on the presented validation evidence"*, and that the evidence did not support one method over the other (50) (pg.104). I believe that the Noben et al findings suggest that administrative data is not a 'gold standard' for self-report; as such, the assessment of criterion validity may have been more appropriately labelled an assessment of construct validity.

Of the six included studies that were considered to have sufficient quality, Noben et al. concluded that patients generally reported lower estimates for resource use, in particular physician and physiotherapy contacts, when compared with administrative data, which contrasts with the results of this study where GP and other healthcare

professional contacts were higher by patient report (50). Similar in design to this study, Patel et al. compared patient-report data with GP medical record data for a random sample of primary care patients (87). Patient-report was collected by post using an adapted version of the CSRI, with a six month recall period (74, 87). The authors found no significant difference between data sources for number of contacts, with agreement, as estimated using Lin's CCC, high (p_c=0.756) (87). More granular information was captured from participants for costing, which allowed self-report GP contacts to be costed by duration of consultation (87). For this, costs were higher when estimated using medical records, with agreement considered fair (p_c=0.609) (87). These results were consistent with this study and Byford et al, who also used the CSRI and observed relatively high agreement for GP contacts (p_c=0.631) (52). A low level of agreement for other healthcare professionals in this study, was consistent with the findings of Byford et al., who found low levels of agreement for practice nurse, community psychologist and community psychiatric nurse contacts (all p_c<0.350) (52). Despite a difference in average cost, the good level of agreement of prescribed medication costs observed in this study (p_c=0.702), was in contrast to existing research, where poor agreement has been observed (146).

8.5.4 Implications for research practice

Based on coverage of resource-use items, it is debatable whether GP medical record data is a 'gold standard' to compare with ModRUM. Some only consider a 'gold standard' to be a longer version of a questionnaire (81), so, for example ModRUM-CD would be a 'gold standard' for ModRUM-C. The results in this study, where for GP and other healthcare professional contacts resource use was higher by self-report, suggests that for these resources, ModRUM may be more comprehensive. As the locations where patients receive treatments and providers become more diverse, for economic evaluation, a validated patient-report measure, which has increased breadth, may be preferable to collecting medical record data from a range of sources covering healthcare utilisation data on primary, secondary and community healthcare providers. For administrative data to remain a feasible option for economic evaluations, increased diversity of providers means that data would ideally be obtained from a higher level (i.e. at integrated care system level, as opposed to individual providers).

For prescribed medications, high levels of sensitivity and specificity indicated that participant report was generally consistent with medical records for binary responses on whether they had used medications. While the agreement between data sources was considered good, the total cost of prescribed medications was 42% higher when costed using medical record data. This could be the result of less accurate recall by participant report, assuming that it is unlikely that medications that were not prescribed would be included in medical records. The estimated cost difference could also have been impacted by the alternative costing approaches, with more detailed medical record data, allowing for increased precision in cost estimates. However, these differences may also have been impacted by the level of detail captured in each method, with the different detail requiring different cost estimates.

When deciding between patient report and administrative data, researchers must make trade-offs, and this should include ease of access to the data. In this study, accessing GP medical record data was not straightforward. I originally planned to personally extract resource-use data from GP records, which has been feasible in previous studies; however, due to stricter rules on accessing patient data, since the implementation of General Data Protection Regulation in 2018, this was not permitted. As I was reliant on GP practice staff, at a time where they were under significant time pressures, there were some delays in developing a search for EMIS Web and obtaining medical record data. I began requesting medical record data from participating practices on February 1st, 2021, but it was not until June 17th, 2021 when all medical record data was obtained, which was three months following completion of participant data collection.

Researchers should carefully consider the amount of detail they ask participants to provide. Both the response rate and question completion rates were higher for the shorter version of ModRUM. Where researchers are uncertain what level of detail is appropriate to collect, depth questions could be included in the feasibility or internal pilot phase of an RCT. Costing this information using both top-level core questions and detailed information from tables, could indicate whether questions can be made more concise. If the researcher chooses to use core module questions in the main trial, the detail provided during the internal pilot or feasibility study could also be used

to inform the most appropriate unit costs to use in the final analysis. For example, if a large proportion of outpatient appointments are performed in Orthopaedics, the researcher may choose an orthopaedics unit cost to cost all appointments, as opposed to a generic outpatient unit cost.

8.5.5 Unanswered questions and future research

In this study, I recruited patients from primary care. The sample size was not as large as originally planned. Representation from certain groups was low (non-white ethnic groups and patients aged 30 and under). For this reason, results from this study cannot provide evidence for the acceptability, feasibility and validity of ModRUM in these groups. All psychometric assessment of ModRUM in this thesis has been conducted with patients recruited from primary care. Although this provides evidence of the validity of ModRUM in this patient group, further validation in other patient groups is required.

The Covid-19 pandemic caused rapid changes to the way patients access healthcare. For this pilot, additional questions were added on remote consultations to account for an increased proportion of remote healthcare. This amendment followed the ModRUM User Guide, where flexibility is permitted to include pertinent resources. However, these added questions did not undergo qualitative testing. As increased utilisation of remote services is likely to continue, validation of these questions would be beneficial.

Covid-19 may result in other changes that impact how this research develops, and what is a priority for future development of ModRUM. For example, use of remote data collection in RCTs may have accelerated due to Covid-19; as such, a priority for future research is developing and validating an online version of ModRUM.

8.6 Conclusion

In this chapter, I have reported on a pilot study of ModRUM with a range of patients recruited from a primary care setting. This study provides preliminary evidence for the feasibility, acceptability, construct validity and criterion validity of ModRUM in this population. In the next and final chapter of this thesis, I have provided a summary of this research.

Chapter 9 Discussion and conclusion

9.1 Research summary

This thesis has contributed to what is considered a relatively under-researched area of trial-based research: resource-use measurement (10-12). To date, there has been a trend for designing or adapting self-report RUMs for each new RCT (9). While this has advantages, as RUMs can be made specific for each trial and patient burden can be minimised if some resource-use data can be obtained from alternate sources, it is also problematic as within the scope of a trial, validation and piloting are rarely performed (10). Bespoke RUMs also lack standardisation, and this contributes to heterogeneity across economic evaluations, which makes it more difficult to compare results across trials and synthesise results (10, 205). The primary aim of this thesis was to develop and perform the initial validation of a new self-report modular RUM (ModRUM). The aim for ModRUM was that it should be standardised, generic, flexible, precise, concise, comprehensible, and transparent. The purpose of each of these principles is described below.

1. Standardised – to increase consistency in resource-use measurement across RCTs, to improve comparability of results,
2. Generic – for use in RCTs assessing a wide range of conditions,
3. Flexible – to allow for pre-defined adaptations, such as adding depth questions, so that ModRUM can be made more relevant to each trial, which should increase uptake,
4. Precise – to allow more detail to be captured for key cost drivers or highly utilised resources, which will increase the accuracy of cost estimates,
5. Concise – to minimise participant burden, with the aim of reducing missing data,
6. Comprehensible – to patients, to avoid confusion and improve accuracy of data,
7. Transparent – to allow researchers implementing ModRUM to clearly report what resource-use data have been captured.

When developing ModRUM, I considered how to balance these principles. For example, ModRUM is standardised in multiple ways including that the wording of questions cannot be altered; however, there is flexibility in areas such as level of

depth which will allow it to be made relevant to different trials. This approach aligns with that reported by Thorn et al. whereby *"standardisation should be based on broad principles rather than being an attempt to create identical RUMs for all occasions"* (10) (pg.157).

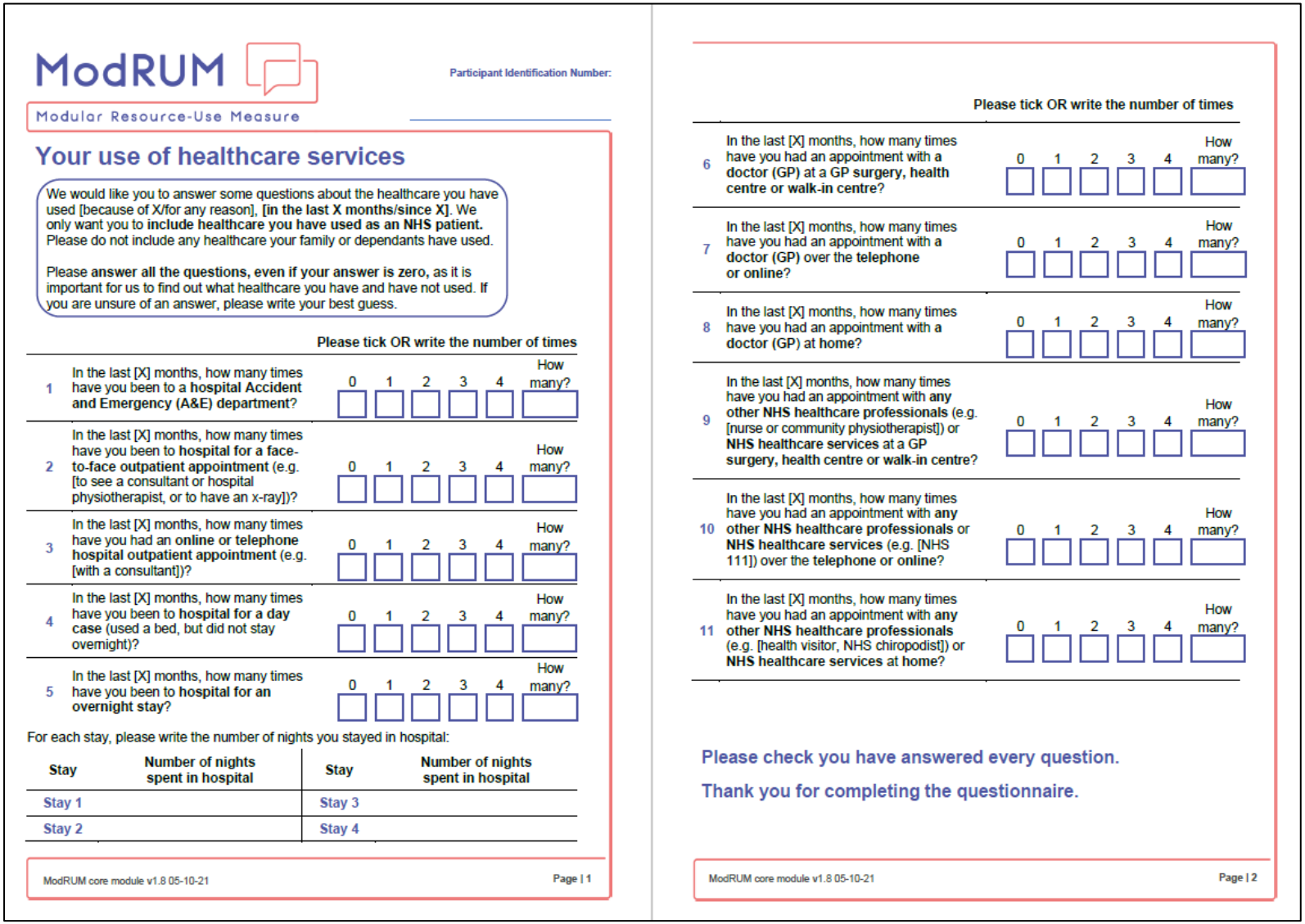
In order to meet the aims of this thesis, I employed several methods, including qualitative and quantitative research methods, in a number of studies, each of which had specific objectives:

- In **Chapter 3**, I identified only 34 RUMs where literature on their development, including information on how items were identified and piloted, was available. The review highlighted that for these RUMs, a range of methods have been used in the development, but limited psychometric assessment was reported. While details on development and psychometric assessment were sparse, the information that was reported was used to inform the development of ModRUM.
- In **Chapter 4**, I described how core healthcare items identified in a Delphi consensus study, by my PhD supervisors (JT, SN, WH) and colleagues, were formulated into a prototype of ModRUM (71). I utilised RUMs stored within DIRUM to inform the wording, formatting and layout of ModRUM (62). The content and design of the prototype of ModRUM benefited from feedback from PROM developers, while input at a later stage in the research from a professional design company enhanced the aesthetics of ModRUM.
- In **Chapter 5**, qualitative interviews with 10 health economists led to changes to ModRUM, including revisions to the content of core and depth questions. Changes were made to ensure that ModRUM had face and content validity and was adequate for capturing self-reported healthcare resource-use data.
- In **Chapter 6**, 20 patients recruited from primary care completed ModRUM during qualitative 'think aloud' interviews with retrospective probing. Issues identified during completion and in probing led to revisions of ModRUM which minimised ambiguity and enhanced comprehensibility. This study was essential as it demonstrated that participant responses were consistent with what questions intended to measure (content validity). In this study, most participants also reported that the content, length, and layout of ModRUM were acceptable.

- In **Chapter 7**, ModRUM was piloted with 11 health economists, who were asked to hypothetically adapt ModRUM for use in one of their recently funded trials, and to provide feedback in an online survey. Most respondents found that it was feasible to adapt ModRUM and indicated they were likely to use ModRUM in the future. Health economists highlighted that the adaptation process could be improved, so that it is less time consuming.
- In **Chapter 8**, ModRUM was piloted with 100 patients recruited from primary care. I found that it was feasible to clean and cost data generated from ModRUM. Preliminary evidence for the feasibility for respondents, acceptability, and construct and criterion validity was also found in this population.

The final version of ModRUM core module is presented in **Figure 9.1** and the final version of ModRUM core module with depth questions is presented in **A7.1** (**Appendix 7**).

Figure 9.1 ModRUM core module (final version)



ModRUM

Modular Resource-Use Measure

Participant Identification Number:

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used [because of X/for any reason], **[in the last X months/since X]**. We only want you to **include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

Please tick OR write the number of times

		0	1	2	3	4	How many?
1	In the last [X] months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	☐	☐	☐	☐	☐	☐
2	In the last [X] months, how many times have you been to **hospital for a face-to-face outpatient appointment** (e.g. [to see a consultant or hospital physiotherapist, or to have an x-ray])?	☐	☐	☐	☐	☐	☐
3	In the last [X] months, how many times have you had an **online or telephone hospital outpatient appointment** (e.g. [with a consultant])?	☐	☐	☐	☐	☐	☐
4	In the last [X] months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)?	☐	☐	☐	☐	☐	☐
5	In the last [X] months, how many times have you been to **hospital for an overnight stay?**	☐	☐	☐	☐	☐	☐

For each stay, please write the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital	Stay	Number of nights spent in hospital
Stay 1		Stay 3	
Stay 2		Stay 4	

ModRUM core module v1.8 05-10-21 Page | 1

Please tick OR write the number of times

		0	1	2	3	4	How many?
6	In the last [X] months, how many times have you had an appointment with a **doctor (GP) at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
7	In the last [X] months, how many times have you had an appointment with a **doctor (GP) over the telephone or online?**	☐	☐	☐	☐	☐	☐
8	In the last [X] months, how many times have you had an appointment with a **doctor (GP) at home?**	☐	☐	☐	☐	☐	☐
9	In the last [X] months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. [nurse or community physiotherapist]) or **NHS healthcare services** at a GP **surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
10	In the last [X] months, how many times have you had an appointment with **any other NHS healthcare professionals** or **NHS healthcare services** (e.g. [NHS 111]) over the **telephone or online?**	☐	☐	☐	☐	☐	☐
11	In the last [X] months, how many times have you had an appointment with **any other NHS healthcare professionals** (e.g. [health visitor, NHS chiropodist]) or **NHS healthcare services at home?**	☐	☐	☐	☐	☐	☐

Please check you have answered every question.

Thank you for completing the questionnaire.

ModRUM core module v1.8 05-10-21 Page | 2

9.2 Strengths and limitations of this research

More detailed strengths and limitations are provided for each chapter in their respective discussion sections. In this section, I provide a summary of the strengths and limitations of this thesis.

9.2.1 Strengths

Exploratory work to generate items for a new RUM was conducted prior to this PhD research, by my PhD supervisors (JT, SN, WH) and colleagues (71). This PhD research has taken these items and successfully formulated them into a RUM and undertaken preliminary validation of the measure with a range of stakeholders. Assessment undertaken in this thesis has shown that ModRUM has content and face validity, and is suitable for use in RCTs, for collecting self-report healthcare utilisation data that can be costed. ModRUM was found to be acceptable to patients recruited from a primary care setting, and preliminary evidence was found on the feasibility, and construct and criterion validity of ModRUM.

At the time of conducting this research, although several publications provided some information on how to develop and/or validate a RUM, the information provided was brief in comparison to vast literature on PROM development (9, 10, 69, 75). As such, this thesis benefited from drawing upon literature from related disciplines, including, health measurement scale, PROM and COS development. The development of ModRUM has benefited from an iterative approach with multiple rounds of testing, as advocated in health measurement scale development literature (79). Once the prototype was developed, in line with PROM development literature, the assessment of content validity was prioritised in qualitative studies with health economists and patients (85). From the assessment of content validity, I found that questions captured the information they were intended to capture (85) and that the ModRUM is relevant, comprehensive and comprehensible to patients (73).

The development of ModRUM involved both qualitative and quantitative methods. For the studies with patients, quantitative testing complemented the qualitative interviews. For example, interviews provided me with a deeper understanding, with the ability to probe when issues occurred to reveal the source of the issue, while the quantitative pilot study afforded me the opportunity to test ModRUM in a more

realistic setting, with a larger number of patients and without interviewer-imposed bias. The pilot study also allowed me to test measurement properties, which required statistical analysis, that could not be assessed in a qualitative study.

While the Covid-19 pandemic caused delays to the patient pilot, the delays allowed me to incorporate additional testing with health economists (**Chapter 7**), which included the development and assessment of a user guide for ModRUM. The ACQP is the only existing generic RUM that I identified in my literature review; however, there has not been significant uptake among health economists within RCTs (76). In the context of low uptake of a generic RUM to date, this additional study allowed me to explore potential barriers to implementing ModRUM in trials. Most health economists indicated that they would use ModRUM in the future and, to increase uptake, suggestions were provided on how the adaptation process could be improved.

9.2.2 Limitations

The review of existing RUMs was performed to identify what methods have been used in the development and validation of existing RUMs. The review had some limitations. First, while I used multiple sources to identify RUMs, I also kept the inclusion criteria strict by only including RUMs that had details on item selection or piloting available. As such, RUMs were excluded if they had undergone validation, but for which details on item selection and piloting were not available. This was an intentional step as I believe that many of these publications would be reporting criterion validity. Information extracted on criterion validation from included RUMs was sufficient for informing methods of assessing criterion validity in this study. A second limitation of the review was that I solely performed the screening, eligibility assessment and extraction of data. Ideally a second reviewer would have conducted these activities, to minimise the risk of omitting a relevant article, or not extracting relevant information.

Both studies testing ModRUM with patients were susceptible to selection bias, as patients were recruited from the BNSSG region of England only, whereas ModRUM will ideally be used in RCTs conducted with patients recruited from areas around the

UK. To counteract this limitation with respect to regional variations in terminology, in the studies with health economists, participants were recruited from around the UK, and in interviews, they were specifically asked whether they thought any terminology would be problematic for patients.

While clinicians have been involved in the development of other RUMs (see Chapter 3 for further detail), I did not involve clinicians in the development of ModRUM. This may be deemed a limitation. However, where clinicians have been involved in the development of other RUMs, the RUMs are designed to be used in a specific population. In this instance, clinicians are likely to be most knowledgeable on the range of healthcare resources patients in a specific population may use. As ModRUM is a generic RUM, health economists were selected to inform the development as they were likely to have extensive experience of collecting resource-use data from patients in a wide range of trials.

Qualitative research involves a subjective approach, with findings from qualitative research influenced by the interviewer and participants (84). Prior to and while undertaking this thesis, I was working as a health economist conducting economic evaluations alongside RCTs. In this role, I worked with patient-reported data collected in RUMs which were designed specifically for each trial. Though there was value in my understanding of the topic as the interviewer, I was also careful not to allow my experience to influence the findings of the qualitative research (84). Using established qualitative methods was beneficial in minimising this risk. For example, the use of constant comparison allowed me to identify when common themes emerged, as opposed to paying undue attention to areas that I am more familiar with (98, 103). To minimise the risk of bias further, it would have been beneficial for the transcripts to have been coded by a second researcher.

I believe the Covid-19 pandemic negatively impacted this thesis in several areas. Firstly, although recruitment for patient ‘think-aloud’ interviews was completed prior to the first lockdown, nearing the end of the recruitment period, in March 2020, I experienced some reluctance from patients, who had initially expressed an interest in taking part. Secondly, the patient pilot study was substantially impacted (**Chapter 8**). Significant amendments were made to the patient pilot study design, so that it

was feasible to deliver without recruiting from GP practice waiting rooms, where direct contact with potential participants would have been required. These changes meant that the desired sample size was not achieved, as I was limited financially in terms of the number of postal invitations I could send. Had I been able to recruit patients face-to-face, as originally planned, I believe the personal approach and lack of restrictions on the number of patients that could be approached would have led to increased participation. Furthermore, the patient pilot took place over winter 2020/21, during which time national lockdowns were in place, which may have negatively impacted the response rate, as patients were less able or willing to take part. A larger sample size would have increased certainty in estimates. The sample size achieved means the results are preliminary and further testing is needed. Finally, recruitment to the health economist pilot was inhibited (**Chapter 7**). Of 25 health economists who expressed an interest in taking part, only 10 participated, with reasons for not participating including delays to trial start dates, meaning health economists were not in a position to adapt ModRUM for their recently funded trial, and due to time constraints of the health economist, which some mentioned were exacerbated by the pandemic.

9.3 Contribution and implications for applied research

9.3.1 ModRUM in comparison to existing methods

ModRUM is a new, generic modular RUM that is designed for self-completion by trial participants. Several adaptations are permitted to make it relevant for a wide range of economic evaluations conducted alongside UK-based RCTs of healthcare interventions. One such adaptation is the option to supplement core questions with depth questions. While the core has several advantages, such as less burden on participants and analysts, the depth questions provide an option for increased precision. In a trial-based economic evaluation, the use of depth questions for certain items would be warranted if differences in the use of such items are anticipated by trial arm.

Alternative methods for collecting resource-use data throughout a trial include bespoke RUMs, which are often developed or adapted for each new trial (9), and administrative data, which can include records from individual healthcare providers, such as GP practices or hospitals, and large administrative datasets, such as the

Clinical Practice Research Datalink (CPRD) for primary care data and HES for secondary care data (49, 54, 206). In **Table 9.1**, I have tabulated differences between ModRUM and the current alternative methods of resource-use data collection for use in UK-based economic evaluations with respect to several areas, including access, scope, detail and participant burden. A comprehensive list of aspects to consider when selecting between self-report and routine (administrative) data, has previously been presented by Franklin and Thorn (49) (pgs.10-11).

Of the three methods described, access issues, in terms of registering to use a RUM or administrative dataset, are least relevant to developing a bespoke RUM. However, designing a bespoke RUM and adequately validating it is likely to be the most time-consuming approach. It would be less time consuming if validation was omitted, which has been a common albeit sub-optimal approach. One aim of creating a generic measure was to reduce duplication of research efforts in designing bespoke RUMs. With less research time spent on designing bespoke RUMs, time could be redirected to performing psychometric assessment of ModRUM, such as content validation, if it is being used in a population that it has not previously been validated in. Access would be most difficult for administrative data, potentially most significantly for large administrative datasets, where there is likely to be a time lag between an event and when the data are available for research. Considerable input may also be required if data are collected directly from individual healthcare providers, such as GP practices, where dependent on trial design, there could be hundreds of sites to collect data from. Furthermore, administrative data may not be available for all participants if they have opted out of their confidential patient information data being used for research purposes.

As using administrative data takes the burden off participants, improvements to access, cost and the format of administrative data, would make it a more appealing method. Administrative data may be preferrable to self-report in studies where participants are likely to have higher healthcare utilisation, as this may indicate poorer health, and a lengthy questionnaire may not be appropriate or acceptable to these participants, which could result in missing or poorer quality data. ModRUM core module, may also be appropriate for participants when participant burden is more of a concern, as it is considerably more concise than many existing RUMs.

In addition to the methods reported above, a new RUM has recently been launched that has been internationally standardised, with English, Dutch and German versions available (207). The PECUNIA RUM is the output of a large international European Union funded grant which aimed to establish standardised costing and outcome assessment measures (207). The development of the PECUNIA RUM included a review of resource-use measurement recommendations (46) and qualitative working including a focus group of health economists and interviews with former mental healthcare users and carers. Full details on the development, beyond the initial review are yet to be published. Consistent with ModRUM, the PECUNIA RUM is a self-report RUM that is designed for use in an adult population (207). While the PECUNIA RUM fosters increased standardisation internationally, ModRUM is designed for UK-based studies. The breadth of the PECUNIA RUM means it can be used for costing in a societal perspective. It is an extensive 37-page questionnaire that covers health, social and informal care, productivity losses, education and justice (207). This is advantageous when compared with ModRUM, which currently only includes a healthcare module; however, it differs in design to ModRUM, as a distinguishing factor of ModRUM is the succinct design.

Table 9.1 Comparing ModRUM to alternatives for data collection

	ModRUM	Bespoke RUM	Administrative data
Access	- Readily available via registration. - Free for non-commercial use. - Minor adaptations may be required prior to implementation. - Further testing recommended prior to implementation if no evidence of validity in the population under study. - Will need to be included in ethics applications.	- Made bespoke, so no delays in accessing; however, if done appropriately, designing and validating the RUM will be time-consuming. - Will need to be included in ethics applications.	- Potential delays in accessing the data, which may be more problematic with large administrative datasets. - Stringent and time-consuming information governance procedures, which are subject to change without warning. - Comparatively costly, for example, NHS administrative staff may need to be reimbursed for extracting data, and national data providers (e.g. NHS Digital and CPRD) have high charges (54, 206). - The national data opt-out service will mean some data for some people is not available for research purposes.
Scope	- At present, limited to NHS-funded healthcare, but the user/researcher has the option to add other items relevant to their RCT. - Future modules to be developed will capture social care, informal care and personal expenses.	- All sectors (e.g. healthcare, education) considered relevant can be included.	- Theoretically, all databases with participant-identifiable data could be used; however, access issues may limit this (49). - Some sectors are not covered by administrative data (e.g. informal care).
Detail	- Brief core questions and detailed depth questions are available. - The level of detail is limited by what participants are aware of and are able to recall (e.g. whether an appointment is consultant led or not, may not be known by the participant).	- Any level of detail that trial participants are able to recall can be covered.	- Potentially more detailed clinical data than patients can recall.

Table 9.1 continued

	ModRUM	Bespoke RUM	Administrative data
Participant burden	- Completion likely required at several points during the trial. - Acceptability to primary care patients shown in this thesis. - The core module is brief in comparison to many bespoke RUMs.	- Completion likely required at several points during the trial. - May not be user-friendly if testing is not conducted prior to implementation in the trial.	- No participant completion burden. - Participants must consent for researchers to access their data.
Researcher burden	- It is feasible to clean and cost data produced by ModRUM. - Burden can be minimised by only including depth questions when necessary, as free text responses vastly increase data processing and analysis time. - Generic code and/or coding rules could save time when implementing ModRUM in multiple RCTs.	- Data cleaning and coding may take longer, or data may not be in the optimal format if time was not spent considering this during RUM development.	- It can be time-consuming to clean administrative data that has not been designed for research (costing in economic evaluation) purposes. - Data may need to be combined from multiple sources.
Validity	- Evidence for the face and content validity. - Preliminary evidence of the acceptability, feasibility, and construct and criterion validity with primary care patients. - Relies on patient recall, with the potential for recall bias - More likely to have missing data than some administrative data.	- Rarely validated prior to implementation. - Relies on patient recall, with the potential for recall bias - More likely to have missing data than some administrative data.	- Validation by the data generators is likely, but validity for use in economic evaluations unlikely (49). - Some research has been conducted in this area, for example, for HES outpatient data, which was found to be reasonably valid (208).

9.3.2 Approach to developing and testing a new RUM

In **Chapter 3**, I found that reporting of the development of RUMs is rare and even when it is conducted the detail provided is often limited. While several publications provide brief details on steps to take in developing a RUM (9, 10, 46, 75), the variety of methods and lack of reporting found in my review of RUMs highlighted that there is no established precedent for RUM development and reporting of such development. While the aim of this research was not to provide a guide for developing RUMs, this thesis provides a detailed set of steps for developing a new RUM. These steps will be used to inform methods for developing and assessing new ModRUM modules covering other sectors.

9.3.3 Launching ModRUM

To foster uptake of ModRUM, several strategies have been or will be employed to make health economists aware that ModRUM is available as an alternative to current methods for collecting resource-use data. Some awareness was achieved as part of this research, as 21 health economists from around the UK participated in interviews or piloting of ModRUM. Over the course of this PhD, I have disseminated details on the development of ModRUM via a peer-reviewed journal publication and reporting results at conferences, where UK-based health economists and trialists were in attendance; including two Health Economists Study Group meetings, the International Clinical Trials Methodology Conference, and the European Health Economics Association PhD Conference (175, 209, 210). To increase awareness further, I plan to disseminate further via peer-reviewed journal publications. ModRUM will also be publicised via social media, have a dedicated page on the Health Economics Bristol website and be listed within DIRUM (62). Users of ModRUM will also be required to cite use of ModRUM, which may increase uptake.

9.3.4 Accessing ModRUM

In addition to considering how to publicise ModRUM, it was important to consider how potential users will access and use it. To increase comparability across future trials, it is important to maximise uptake amongst the target audience, by minimising any barriers to access. To do that, ModRUM will be made available at no cost for non-commercial use. While no registration process would be the most simplistic

approach, it was agreed with the research team (PhD supervisors) that researchers should register for a license to use ModRUM. Using a short registration process will be valuable for several reasons: [1] the pilot with health economists highlighted that users may make unpermitted adaptations, so registration means that adaptations can be monitored, [2] a record of which studies ModRUM has been used in will be created, including details of further validation, which could be useful to other potential users who want to use ModRUM in a similar population, and [3] a dedicated registration page will ensure users access the most recent version of ModRUM and the corresponding User Guide (211). Without registration, users of ModRUM may go on to make unpermitted adaptations, or use ModRUM in a population it was not designed for, which may mean that data generated from an adapted version are not valid (212). For example, translating ModRUM for use in a different language or country, requires more than simply translating ModRUM as terminology may have different meanings in different languages; further psychometric assessment ought to be conducted. A registration process will differentiate ModRUM from existing RUMs, and increase consistency with outcome measures, such as the EQ-5D, where registration and charging for commercial use of a measure is commonplace (182).

9.4 Future development

Ideas for future development have been described in the discussion sections of multiple chapters throughout this thesis. In this section, I summarise two keys areas that could be prioritised for future development which could increase the uptake and longevity of ModRUM.

9.4.1 Breadth bolt-on modules

The aspects of ModRUM developed in this thesis only offer flexibility in the level of detail to capture; future work should prioritise the development of breadth bolt-on modules, which would provide flexibility to capture resources beyond the healthcare sector in a consistent manner to healthcare resources. This would increase the relevance of ModRUM to trials capturing resource-use beyond an NHS perspective. The development of ModRUM was performed in the context of the UK healthcare system, where guidelines for conducting economic evaluations developed by NICE assert that in the primary analysis of an economic evaluation of healthcare

interventions, an NHS and PSS perspective should be taken (23). In the Delphi study, where items for ModRUM were identified, it was concluded that a core set of items for a new RUM should include NHS items alone (71). As PSS resources may not be relevant to capture in all trials, it was suggested that PSS items could form a separate bolt-on module (71). Within this thesis, health economists indicated that it would be desirable to have items on social care (**Chapter 5** and **Chapter 7**).

Utilising experience gained in the development of the healthcare module of ModRUM, research funding was secured via the NIHR Research for Patient Benefit programme to develop social care, informal care and personal expenses modules for ModRUM. The research will follow a similar process to that established in this research, including qualitative interviews with health economists and 'think-aloud' interviews and piloting with people who access social care. Prior to these activities, items for inclusion will be informed by a literature review and refined in focus groups with health economists and people who access social care services.

Development of concise modules for other sectors, such as social care may be more difficult as the range of resources is potentially more diverse, or more variable dependant on condition. For example, in a trial incorporating a social care perspective, home care is likely to be vital for trials with frail, elderly populations, whereas for a trial assessing mental health interventions, a social care perspective may be relevant, but home care is less likely to be used by participants or related to the intervention.

9.4.2 Mode of administration

Many RCTs now provide options for trial participants to complete study documentation in online or app-based questionnaires as an alternative to paper-based questionnaires. To keep up with these advances, it will be important to prioritise the development of alternative modes of administration. Without a readily available online version, researchers may choose not to use ModRUM, or they may develop an online or app-based version of ModRUM themselves, which would require further validation. Development of bespoke online versions could introduce inconsistencies in the way ModRUM is implemented. Online and app-based modes of administration would afford the opportunity to reduce participant burden via skip

logic (for example, where depth questions only appear if a positive response is provided to the top-level question), minimise missing data through reminders to missed responses and/or enforcing responses, minimise the ability for respondents to input answers outside the range of responses and provide more examples (for example, hovering over words could bring up more examples or definitions of potentially problematic terms).

9.5 Future testing

Areas where further testing is required have been discussed throughout the discussion sections of this thesis. This section summarises areas where future testing of ModRUM could be prioritised.

9.5.1 Testing ModRUM within randomised controlled trials

In this thesis, all testing of ModRUM has been conducted outside of the RCT setting it is ultimately designed to be used in. Although the pilot studies (**Chapter 7** and **Chapter 8**) aimed to reflect the way health economists would adapt and patients would complete ModRUM, there are likely to be differences which means that testing within trials is needed. For example, although the patient pilot study (**Chapter 8**) was designed to mimic the way trial participants would complete ModRUM (i.e. recruiting active healthcare users; postally administered, with an accompanying outcome measure), there are likely to be differences in a trial context. One of the main differences could be that, for some trials, participants will use significantly more resources than most patients used in the pilot study. As such, the ModRUM-completion burden would be higher for these participants, particularly if depth questions are included. As indicated in interviews with patients (**Chapter 6**), ModRUM core with depth questions, may not be acceptable to trial participants who have higher healthcare utilisation. One area, that may require further consideration, is for trials where participants have multiple appointments for the same reason (e.g. for delivery of radiotherapy fractions). While depth questions may be needed in such a trial, further design considerations may be needed so that participants do not have to complete multiple rows in a table for the same reason.

A further question is, when is additional validation required prior to administering in an RCT? While adaptations are permitted, it is unclear as to what extent of adaptation would require additional testing prior to administration. Validation is a time-consuming process, so ideally once ModRUM has been tested and used in a specific patient group, then that version of ModRUM would be valid in other studies with the same patient group. However, end users may want to make adaptations, and a judgement would need to be made as to whether the validation evidence holds true in the revised version. To overcome this, one option for further consideration could be to follow the design of COSs for specific areas of health or healthcare and have core resource sets for specific areas (70). This would maintain flexibility of ModRUM across areas, but could increase standardisation within specific areas of health or healthcare, where appropriate.

9.5.2 Testing ModRUM with under-represented groups

Although the sampling strategy employed in the cognitive interviews study (**Chapter 6**) prioritised recruitment of groups we believed would be under-represented, no patients from non-white ethnic groups and few younger patients returned reply forms. A similar result was found in the pilot study (**Chapter 8**). This result was not unexpected as the proportion of non-white patients at the participating practices was low, and invitation to the study required patients to visit their practice/have an appointment, which was less probable for younger patients. To test ModRUM with non-white ethnic groups, an alternative recruitment method may be required, which could involve recruiting practices with a higher proportion of non-white patients, or recruitment from an alternative setting.

9.5.3 Testing of other measurement properties of ModRUM

Prior to alterations made to the design of the patient pilot study due to the Covid-19 pandemic, I had intended to test reliability. As it was unclear what the response rate would be to the postal invitation for the pilot study, all available budget was spent on maximising the number of invitations. As a result, reliability testing was dropped from the pilot study. In the review of existing RUMs, I found that assessment of reliability had only been reported for three studies. When developing ModRUM, consideration was given to undertaking reliability testing of a RUM, to leave a gap in time sufficient enough for respondents not to recall their previous answer, yet not so long that the

response would have changed. A future research study could include the assessment of test-retest reliability which would involve administering ModRUM to the same individuals at different times to establish whether the answers they provide are reproducible and reliable (79).

9.6 Conclusion

In this thesis, I have developed and performed the initial validation of the healthcare module of a new modular resource-use measure: ModRUM. ModRUM is a standardised RUM, with a concise core healthcare module that includes 11 questions and optional depth questions, which can be included to capture additional detail for resources where increased precision may be required. ModRUM is a generic measure, and some permitted adaptability means that it will be relevant for capturing resource-use data in a wide range of RCTs, which will improve comparability of results. Using an iterative approach in the development and testing, with qualitative and quantitative methods, ModRUM was found to be: (1) valid, with respect to content, comprehension, and comprehensiveness (2) acceptable in terms of content, length, and layout, to patients recruited from primary care; (3) feasible to adapt for use in economic evaluations; and (4) suitable for costing purposes in economic evaluations alongside RCTs.

References

1. Morris S, Devlin N, Parkin D. Economic Analysis in Health Care. West Sussex, England: John Wiley & Sons, Ltd; 2007.
2. The World Bank. Current health expenditure per capita (current US$) 2021 [Available from: https://data.worldbank.org/indicator/SH.XPD.CHEX.PC.CD?end=2018&start=2000&view=chart].
3. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. London, UK: National Institute for Health and Care Excellence; 2013.
4. McGuire A. Theoretical concepts in the economic evaluation of health care. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001. p. 1-21.
5. Palmer S, Torgerson D. Definitions of efficiency. British Medical Journal. 1999;318:1136.
6. Patel A. Equity and Efficiency. In: Razzouk D, editor. Mental Health Economics. Cham, Switzerland: Springer Nature; 2017.
7. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. New York: Oxford University Press; 2005.
8. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-Effectiveness Analysis Alongside Clinical Trials II—An ISPOR Good Research Practices Task Force Report. Value in Health. 2015;18:161-72.
9. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK health technology assessment program. Value in Health. 2010;13(8):867-72.
10. Thorn J, Coast J, Cohen D, Hollingworth W, Knapp M, Noble S, et al. Resource-use measurement based on patient recall: issues and challenges for economic evaluation. Applied Health Economics and Health Policy. 2013;11(3):155-61.
11. de Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press; 2011.
12. Mokkink L, Terwee C, Knol D, Stratford P, Alonso J, Patrick D, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Medical Research Methodology. 2010;10(22).
13. Hurley J. An Overview of the Normative Economics of the Health Sector. In: Culyer A, Newhouse J, editors. Handbook of Health Economics. Amsterdam: Elsevier; 2000. p. 55-118.
14. Coast J, Smith R, Lorgelly P. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. Social Science & Medicine. 2008;67:1190-8.
15. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. Welfarism vs. extra-welfarism. Journal of Health Economics. 2008;27:325-38.
16. Coast J, Mitchell P, Goranitis I. Ethics and Values in Welfarism and Extra-Welfarism. In: Razzouk D, editor. Mental Health Economics. Cham, Switzerland: Springer Nature; 2017.
17. Coast J. Maximisation in extra-welfarism: A critique of the current position in health economics. Social Science & Medicine. 2009;69:786-92.

18. Hicks JR. The foundations of welfare economics. Economic Journal. 1939;49:694-712.
19. Clarke PE. Methods for evaluating health and health care: Underlying theory and implications for practical application. In: McIntosh E, Clarke P, Frew EJ, Louviere J, editors. Applied Methods of Cost-Benefit Analysis in Health Care. Oxford: Oxford University Press; 2010.
20. Culyer A. The normative economics of health care finance and provision. Oxford Review of Economic Policy. 1989;5(1):34-58.
21. Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press; 2017.
22. Buchanan J, Wordsworth S. Welfarism versus extra-welfarism: can the choice of economic evaluation approach impact on the adoption decisions recommended by economic evaluation studies? PharmacoEconomics. 2015;33:571-9.
23. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual 2018, Oct 31 ed. London: National Institute for Health and Care Excellence (NICE); 2014.
24. Gray A, Clarke P, Wolstenholme J, Wordsworth S. Applied Methods of Cost-effectiveness Analysis in Health Care. Oxford: Oxford University Press; 2011.
25. Razzouk D. Cost-Benefit Analysis. In: Razzouk D, editor. Mental Health Economics. Cham, Switzerland: Springer Nature; 2017.
26. Coast J. Is economic evaluation in touch with society's health values? British Medical Journal. 2004;329:1233-6.
27. Mauskopf J, Paul J, Grant D, Stergachis A. The role of cost-consequence analysis in healthcare decision-making. PharmacoEconomics. 1998;13(3):277-88.
28. Franklin M. Cost Utility Analysis. In: Razzouk D, editor. Mental Health Economics. Cham, Switzerland: Springer Nature; 2017.
29. EuroQol group. EuroQol - a new facility for the measurement of health-related quality-of-life. Health Policy. 1990;16(3):199-208.
30. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of Life Research. 2011;20(10):1727-36.
31. Dolan P. Modeling valuations for EuroQol health states. Medical Care. 1997;35(11):1095-108.
32. Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Economics. 2017;27(1):7-22.
33. National Institute for Health and Care Excellence. Position statement on use of the EQ-5D-5L value set for England (updated October 2019) 2019 [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l].
34. van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health. 2012;15(5):708-15.
35. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Economics. 2005;14:487-96.
36. National Institute for Health and Clinical Excellence. Social value judgements: Principles for the development of NICE guidance. 2nd ed2008.
37. National Institute for Health and Care Excellence. Interim process and methods of the highly specialised technologies programme: updated to reflect 2017

changes 2017 [Available from: https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-highly-specialised-technologies-guidance].
38. Briggs A, O'Brien BJ. The death of cost-minimization analysis? Health Economics. 2001;10(2):179-84.
39. Dakin H, Wordsworth S. Cost-minimisation analysis versus cost-effectiveness analysis, revisited. Health Economics. 2013;22:22-34.
40. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? Health Economics. 2006;15:677-87.
41. Petrou S, Gray A. Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting British Medical Journal. 2011;342.
42. Raftery J. Costing in economic evaluation. British Medical Journal. 2000;320:1597.
43. Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. 2nd ed. Oxford: Oxford University Press; 2015.
44. Hughes D, Charles J, Dawoud D, Tudor Edwards R, Holmes E, Jones C, et al. Conducting economic evaluations alongside randomised trials: current methodological issues and novel approaches. PharmacoEconomics. 2016;34(5):447-61.
45. Byford S, Raftery J. Perspectives in economic evaluation. British Medical Journal. 1998;316:1529-30.
46. Janssen LMM, Drost RMWA, Paulus ATG, Garfield K, Hollingworth W, Noble S, et al. Aspects and challenges of resource use measurement in health economics: towards a comprehensive measurement framework. PharmacoEconomics. 2021;39(9):983-93.
47. Potter S, Davies C, Davies G, Rice C, Hollingworth W. The use of micro-costing in economic analyses of surgical interventions: a systematic review. Health Economics Review. 2020;10(3).
48. NHS England and NHS Improvement. National Schedule of NHS Costs 2018/19. 2020 [Available from: https://www.england.nhs.uk/national-cost-collection/].
49. Franklin M, Thorn J. Self-reported and routinely collected electronic healthcare resource-use data for trial-based economic evaluations: the current state of play in England and considerations for the future. BMC Medical Research Methodology. 2019;19(8):1-13.
50. Noben CY, de Rijk A, Nijhuis F, Kottner J, Evers S. The exchangeability of self-reports and administrative health care resource use measurements: assessment of the methodological reporting quality. Journal of Clinical Epidemiology. 2016;74:93-106.
51. Leggett LE, Khadaroo RG, Holroyd-Leduc J, Lorenzetti DL, Hanson H, Wagg A, et al. Measuring resource utilization: a systematic review of validated self-reported questionnaires. Medicine. 2016;95(10):e2759.
52. Byford S, Leese M, Knapp M, Seivewright H, Cameron S, Jones V, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. Health Economics. 2007;16(5):531-6.
53. van Lier L, Bosmans J, van Hout H, Mokkink L, van den Hout W, de Wit G, et al. Consensus-based cross-European recommendations for the identification, measurement and valuation of costs in health economic evaluations: a European Delphi study. The European Journal of Health Economics. 2018;19:993-1008.

54. NHS Digital. Hospital Episode Statistics (HES). Available from: [https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics].
55. Marques E, Johnson E, Gooberman-Hill R, Blom A, Noble S. Using resource use logs to reduce the amount of missing data in economic evaluations alongside trials. Value in Health. 2013;16(1):195-201.
56. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. Medical Care Research and Review. 2006;63(2):217-35.
57. Kjellsson G, Clarkec P, Gerdtham U-G. Forgetting to remember or remembering to forget: A study of the recall period length in health care survey questions. Journal of Health Economics. 2014;35:34-46.
58. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: the current state of play. Health Economics. 2012;21(2):187-200.
59. Sabariego Tomás C. Assessment and determinants of health care utilization among patients with musculoskeletal conditions undergoing outpatient rehabilitation in Germany. Munich: Ludwig Maximilian University of Munich; 2011.
60. Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. Health Technology Assessment. 1999;3(6).
61. Kashner T, Stensland M, Lind L, Wicker A, Rush A, Golden R, et al. Measuring use and cost of care for patients with mood disorders: the utilization and cost inventory. Medical Care. 2009;47(2):184-90.
62. Ridyard CH, Hughes DA, Dirum Team. Development of a database of instruments for resource-use measurement: purpose, feasibility, and design. Value in Health. 2012;15(5):650-5.
63. McCrone P, Weich S. Mental health care costs: paucity of measurement Social Psychiatry and Psychiatric Epidemiology. 1996;31(2):70-7.
64. Ridyard C, Hughes D, DIRUM Team. Taxonomy of methods of resource use measurement. Health Economics. 2015;24:372-8.
65. McIntosh E. Shadow Pricing in Health Care Cost-Benefit Analyses. In: McIntosh E, Clarke P, Frew EJ, Louviere J, editors. Applied Methods of Cost-Benefit Analysis in Health Care. Oxford: Oxford University Press; 2010. p. 39-54.
66. Curtis LA, Burns A. Unit Costs of Health & Social Care 2020. University of Kent: PSSRU; 2020.
67. Joint Formulary Committee. British National Formulary (online) London: BMJ Group and Pharmaceutical Press.
68. van den Berg B, Bleichrodt H, Eeckhoudt L. The economic value of informal care: a study of informal caregivers' and patients' willingness to pay and willingness to accept for informal care. Health Economics. 2005;14:363-76.
69. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of Clinical Epidemiology. 2010;63(7):737-45.
70. Williamson P, Altman D, Bagley H, Barnes K, Blazeby J, Brookes S, et al. The COMET Handbook: version 1.0. Trials. 2017;18:1-50.
71. Thorn JC, Brookes ST, Ridyard C, Riley R, Hughes DA, Wordsworth S, et al. Core items for a standardized resource use measure (ISRUM): expert delphi consensus survey. Value in Health. 2018;21(6):640-9.

72. Ridyard C, Hughes D. Review of resource-use measures in UK economic evaluations, In: Curtis L, Burns A, editors. Unit Costs of Health and Social Care 2015. 2015.
73. Prinsen C, Mokkink L, Bouter L, Alonso J, Patrick D, de Vet H, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018;27:1147-57.
74. Beecham J, Knapp M. Measuring Mental Health Needs. In: Thornicroft G, editor. Costing psychiatric interventions. 2nd ed. London: Gaskell; 2001. p. 200-24.
75. Chisholm D, Knapp MRJ, Knudsen HC, Amaddeo F, Gaite L, van Wijngaarden B, et al. Client Socio-Demographic and Service Receipt Inventory – European Version: development of an instrument for international research: EPSILON Study 5. The British Journal Psychiatry. 2000;177(suppl 39):s28-s33.
76. Thompson S, Wordsworth S. An annotated cost questionnaire for completion by patients. HERU Discussion Paper 03/01. 2001.
77. Franklin M, Lomas J, Walker S, Young T. An educational review about using cost data for the purpose of cost-effectiveness analysis. PharmacoEconomics. 2019.
78. Knapp M, Beecham J. Reduced list costings: examination of an informed short cut in mental health research. Health Economics. 1993;2(4):313-22.
79. Streiner DL, Norman GR, Cairney J. Health measurement scales: A practical guide to their development and use. 5th ed. Oxford: Oxford University Press; 2015.
80. Patrick DL, Burke LB, Gwaltney CJ, Kline Leidy N, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. Value in Health. 2011;14(8):967-77.
81. Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating patient-based outcome measures for use in clinical trials: a review. Health Technology Assessment. 1998;2(14):1-74.
82. Hendriks MRC, Al MJ, Bleijlevens MHC, van Haastregt JCM, Crebolder HFJM, van Eijk JTM, et al. Continuous versus intermittent data collection of health care utilization. Medical Decision Making. 2013;33(8):998-1008.
83. Hammersley M. What's wrong with ethnography? Methodological explorations. London: Routledge; 1992.
84. Coast J, Jackson L. Theoretical and methodological positions and the choice to use qualitative methods. In: Coast J, editor. Qualitative Methods in Health Economics. London, UK: Rowman & Littlefield Ltd; 2017.
85. Patrick DL, Burke LB, Gwaltney CJ, Kline Leidy N, Martin ML, Molsen E, et al. Content validity - establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2 - assessing respondent understanding. Value in Health. 2011;14(8):978-88.
86. Coast J. Qualitative Methods in Health Economics. London, UK: Rowman & Littlefield Ltd; 2017.
87. Patel A, Rendu A, Moran P, Leese M, Mann A, Knapp M. A comparison of two methods of collecting economic data in primary care. Family Practice. 2005;22(3):323-7.
88. Willis GB. Cognitive Interviewing A Tool for Improving Questionnaire Design. California: Sage Publications, Inc; 2005.
89. Tourangeau R. Cognitive sciences and survey methods. In: Jabine T, Straf M, Tanur J, Tourangeau R, editors. Cognitive Aspects of Survey Methodology: Building

a Bridge Between Disciplines. Washington, DC: National Academy Press; 1984. p. 73-100.
90. Corbin J, Strauss A. Basics of Qualitative Research. 4th ed. California, US: Sage Publications; 2015.
91. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. Trials. 2012;13.
92. Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome set-standards for development: The COS-STAD recommendations. PLOS Medicine. 2017;14(11):e1002447.
93. Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PLOS ONE. 2019;14(12):e0225980.
94. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLOS Medicine. 2011;8(1):e1000393.
95. Coast J, Jackson L. Understanding primary data analysis. In: Coast J, editor. Qualitative methods in health economics. London, UK: Rowman & Littlefield Ltd; 2017. p. 92-118.
96. Bailey C, Kinghorn P, Orlando R, Coast J. Using 'think-aloud' and interview data to explore patient and proxy completion of health and capability measures at the end of life. In: Coast J, editor. Qualitative Methods in Health Economics. London, UK2017.
97. Glaser B, Strauss A. The discovery of grounded theory. Chicago: Aldine; 1967.
98. Glaser BG. The constant comparison method of qualitative analysis. Social Problems. 1965;12(4):436-45.
99. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLOS Medicine. 2009;6(7):e1000097.
100. Patton MQ. Qualitative Research and Evaluation Methods. 3rd ed. Thousand Oaks, California: Sage; 2002.
101. DeVellis RF. Scale development: theory and applications. Los Angeles: Sage; 2017.
102. Rubin HJ, Rubin IS. Qualitative Interviewing: The Art of Hearing Data. 3rd ed. Los Angeles, California: Sage Publications Inc.; 2012.
103. NVivo qualitative data analysis software, Version 12 Pro. QSR International Pty Ltd; 2018.
104. Flick U. An introduction to qualitative research. 5th ed. Los Angeles: Sage; 2014.
105. Willis GB. Cognitive Interviewing - A How To Guide: Research Triangle Institute; 1999.
106. Horwood J, Pollard B, Ayis S, McIlvenna T, Johnston M. Listening to patients: using verbal data in the validation of the Aberdeen measures of impairment, activity limitation and participation restriction (Ab-IAP) BMC Musculoskeletal Disorders. 2010;11(182):1-13.
107. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Gaithersburg, USA: Advanced Analytics, LLC; 2014.

108. Klein D. Implementing a general framework for assessing interrater agreement in Stata. The Stata Journal. 2018;18(4):871-901.
109. Jisc. Online surveys (formerly BOS) [Available from: www.onlinesurveys.ac.uk].
110. Likert R. A technique for the measurement of attitudes. Archives of Psychology. 1932;22:5-55.
111. Fowler F. Survey Research Methods. 4th ed. ed. Los Angeles, California: Sage Publications; 2014.
112. StataCorp. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC; 2021.
113. Kirkwood BR, Sterne JAC. Medical Statistics. 2nd ed. Malden, Massachusetts: Blackwell Science; 2003.
114. Lin L-K. A concordance correlation coefficient to evaluate reproducibility. Biometrics. 1989;45:255-68.
115. Thorn J, Ridyard C, Riley R, Brookes S, Hughes D, Wordsworth S, et al. Identification of items for a standardised resource-use measure: review of current instruments. Trials. 2015;16(Suppl 2):O26.
116. Faria R, Gomes M, Epstein D, White I. A guide to handling missing data in cost-effectiveness analysis conducted within randomised controlled trials. PharmacoEconomics. 2014;32:1157-70.
117. Leurent B, Gomes M, Carpenter JR. Missing data in trial-based cost-effectiveness analysis: an incomplete journey. Health Economics. 2018;27:1024-40.
118. Mayer S, Paulus A, Łaszewska A, Simon J, Drost R, Ruwaard D, et al. Resource-Use Measurement Instruments for Intersectoral Costs and Benefits in the Education and Criminal Justice Sectors. PharmacoEconomics 2017:1-14.
119. Choi BCK, Pak AWP. A catalog of biases in questionnaires. Preventing Chronic Disease. 2005;2(1):A13.
120. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.
121. Chernyak N, Ernsting C, Icks A. Pre-test of questions on health-related resource use and expenditure, using behaviour coding and cognitive interviewing techniques. BMC Health Services Research. 2012;12(303):1-8.
122. Beemster TT, van Velzen JM, van Bennekom COM, Reneman MF, Frings-Dresen MHW. Test–retest reliability, agreement and responsiveness of productivity loss (iPCQ-VR) and healthcare utilization (TiCP-VR) questionnaires for sick workers with chronic musculoskeletal pain. Journal of Occupational Rehabilitation. 2019;29:91-103.
123. Beresford B, Mann R, Parker G, Kanaan M, Faria R, Rabiee P, et al. Reablement services for people at risk of needing social care: the MoRe mixed-methods evaluation. Health Services Research and Delivery. 2019;7(16).
124. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan SS, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Services Research. 2013;13.
125. Icks A, Dittrich A, Brüne M, Kuss O, Hoyer A, Haastert B, et al. Agreement found between self-reported and health insurance data on physician visits comparing different recall lengths. Journal of Clinical Epidemiology. 2017;82:167-72.
126. Chernyak N. Self-reported health care utilization: measurement issues, data validity and implications for design of health surveys and economic evaluations: an

empirical investigation among patients with diabetes mellitus: University of Bremen; 2011.
127. Chisholm D, Conroy S, Glangeaud-Freudenthal N, Oates M, Asten P, Barry S, et al. Health services research into postnatal depression: results from a preliminary cross-cultural study. British Journal of Psychiatry. 2004;184.
128. Clarke CE, Patel S, Ives N, Rick CE, Woolley R, Wheatley K, et al. Clinical effectiveness and cost-effectiveness of physiotherapy and occupational therapy versus no therapy in mild to moderate Parkinson's disease: a large pragmatic randomised controlled trial (PD REHAB). Health Technology Assessment. 2016;20(63).
129. Cooke M, Marsh J, Clark M, Nakash R, Jarvis R, Hutton J, et al. Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. the cast trial. Health Technology Assessment. 2009;13(13):1-158.
130. Cooper NJ, Mugford M, Symmons DP, Barrett EM, Scott DG. Development of resource-use and expenditure questionnaires for use in rheumatology research. Journal of Rheumatology. 2003;30(11):2485-91.
131. Fox M, Voordouw J, Mugford M, Cornelisse J, Antonides G, Frewer L. Social and economic costs of food allergies in Europe: development of a questionnaire to measure costs and health utility. Health Services Research. 2009;44(5):1662-78.
132. Fox M, Cerecedo I, de la Hoz Caballer B, Zamora J, Rokicka E, Starosta P, et al. Social and economic costs of food allergies in Europe: translation and trans-cultural adaptation into different languages of a questionnaire to measure costs and health utility. HEG Working Paper. 2013;No. 13-02.
133. Gordon L, Patrao T, Hawkes A. Can colorectal cancer survivors recall their medications and doctor visits reliably? BMC Health Services Research. 2012;12.
134. Gray R, Rick C, Rowland N, Williams A, Gray A, Clarke C, et al. A large randomised assessment of the relative cost-effectiveness of different classes of drugs for Parkinson's disease (PD MED) 2010 [Available from: https://www.dirum.org/instruments/details/29#].
135. Griffin D, Wall P, Realpe A, Adams A, Parsons N, Hobson R, et al. UK FASHIoN: feasibility study of a randomised controlled trial of arthroscopic surgery for hip impingement compared with best conservative care. Health Technology Assessment. 2016;20(32).
136. Guzman J, Pelosi P, Bombardier C. Capturing health care utilization after occupational low-back pain: development of an interviewer-administered questionnaire. Journal of Clinical Epidemiology. 1999;52(5):419-27.
137. Hirst A, Dutton S, Wu O, Briggs A, Edwards C, Waldenmaier L, et al. A multi-centre retrospective cohort study comparing the efficacy, safety and cost-effectiveness of hysterectomy and uterine artery embolisation for the treatment of symptomatic uterine fibroids. The HOPEFUL study. Health Technology Assessment. 2008;12(5):1-248.
138. Houchen-Wolloff L, Gardiner N, Devi R, Robertson N, Jolly K, Marshall T, et al. Web-based cardiac rehabilitation alternative for those declining or dropping out of conventional rehabilitation: results of the WREN feasibility randomised controlled trial. Open Heart. 2018;5:1-8.
139. Mauch V, Woods N, Kirubi B, Kipruto H, Sitienei J, Klinkenberg E. Assessing access barriers to tuberculosis care with the tool to estimate patients' costs: pilot results from two districts in Kenya. BMC Public Health. 2011;11:1-9.

140. Mirandola M, Bisoffi G, Bonizzato P, Amaddeo F. Collecting psychiatric resources utilisation data to calculate costs of care: a comparison between a service receipt interview and a case register. Social Psychiatry and Psychiatric Epidemiology. 1999;34:541-7.
141. Murray D, MacLennan G, Breeman S, Dakin H, Johnston L, Campbell M, et al. A randomised controlled trial of the clinical effectiveness and cost-effectiveness of different knee prostheses: the Knee Arthroplasty Trial (KAT). Health Technology Assessment. 2014;18(19):1-268.
142. Ness N-H, Haase R, Kern R, Schriefer D, Ettle B, Cornelissen C, et al. The multiple sclerosis health resource utilization survey (MS-HRS): development and validation study. Journal of Medical Internet Research. 2020;22(3).
143. O'Dowd H, Gladwell P, Rogers C, Hollinghurst S, Gregory A. Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme. Health Technology Assessment. 2006;10(37):1-140.
144. McIntosh E, Gray A, Daniels J, Gill S, Ives N, Jenkinson C, et al. Cost-utility analysis of deep brain stimulation surgery plus best medical therapy versus best medical therapy in patients with Parkinson's: economic evaluation alongside the PD SURG trial. Movement Disorders. 2016;31(8):1173-82.
145. Peek G, Elbourne D, Mugford M, Tiruvoipati R, Wilson A, Allen E, et al. Randomised controlled trial and parallel economic evaluation of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR). Health Technology Assessment. 2010;14(35):1-90.
146. Pinto D, Robertson MC, Hansen P, Abbott JH. Good agreement between questionnaire and administrative databases for health care use and costs in patients with osteoarthritis. BMC Medical Research Methodology. 2011;11.
147. Ruof J, Huelsemann JL, Mittendorf T, Handelmann S, von der Schulenburg JM, Zeidler H, et al. Patient-reported health care utilization in rheumatoid arthritis: what level of detail is required? Arthritis & Rheumatism. 2004;51(5):774-81.
148. Hulsemann JL, Ruof J, Zeidler H, Mittendorf T. Costs in rheumatology: results and lessons learned from the 'Hannover Costing Study'. Rheumatology International. 2006;26(8):704-11.
149. Russell A, Gaunt D, Cooper K, Horwood J, Barton S, Ensum I, et al. Guided self-help for depression in autistic adults: the ADEPT feasibility RCT. Health Technology Assessment. 2019;23(68).
150. Sabes-Figuera R, McCrone P, Sharac J, Csipke E, Craig TKJ, Rose DS, et al. Developing a tool for collecting and costing activity data on psychiatric inpatient wards. Epidemiology and Psychiatric Sciences. 2012;21:393-9.
151. Schweikert B, Hahmann H, Leidl R. Development and first assessment of a questionnaire for health care utilization and costs for cardiac patients. BMC Health Services Research. 2008;8:1-11.
152. van den Brink M, van den Hout WB, Stiggelbout AM, van de Velde CJH, Kievit J. Cost measurement in economic evaluations of health care: whom to ask? Medical Care. 2004;42(8):740-6.
153. van den Brink M, van den Hout WB, Stiggelbout AM, Putter H, van de Velde CJH, Kievit J. Self-reports of health-care utilization: diary or questionnaire? International Journal of Technology Assessment in Health Care. 2005;21(298-304).
154. Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, et al. Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations. Health Technology Assessment. 2004;8(50):1-127.

155. Williams J, Alam M, Alrubaiy L, Clement C, Cohen D, Grey M, et al. Comparison Of iNfliximab and ciclosporin in STeroid Resistant Ulcerative Colitis: pragmatic randomised Trial and economic evaluation (CONSTRUCT). Health Technology Assessment. 2016;20(44).
156. Wyatt K, Henley W, Anderson L, Anderson R, Nikolaou V, Stein K, et al. The effectiveness and cost of enzyme replacement and substrate reduction therapies: a longitudinal cohort study of people with lysosomal storage disorders. Health Technology Assessment. 2012;16(39):1-543.
157. Gomes B, McCrone P, Hall S, Riley J, Koffman J, Higginson IJ. Cognitive interviewing of bereaved relatives to improve the measurement of health outcomes and care utilisation at the end of life in a mortality followback survey. Supportive Care in Cancer. 2013;21:2835-44.
158. Klingshirn H, Mittrach R, Braitmayer K, Strobl R, Bender A, Grill E, et al. RECAPDOC - a questionnaire for the documentation of rehabilitation care utilization in individuals with disorders of consciousness in long-term care in Germany: development and pretesting. BMC Health Services Research. 2018;18:1-9.
159. Mokkink L, de Vet H, Prinsen C, Patrick D, Alonso J, Bouter L, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018;27:1171–9.
160. Heslin M, Babalola O, Ibrahim F, Stringer D, Scott D, Patel A. A comparison of different approaches for costing medication use in an economic evaluation. Value in Health. 2018;21(2):185-92.
161. Bradburn N, Sudman S, Wansink B. Asking Questions: The Definitive Guide to Questionnaire Design - For Market Research, Political Polls, and Social and Health Questionnaires. San Francisco, CA: Jossey-Bass; 2004.
162. Mays N, Pope C. Qualitative research in health care: assessing quality in qualitative research. British Medical Journal. 2000;320:50-2.
163. Hahn EA, DeVellis RF, Bode RK, Garcia SF, Castel LD, Eisen SV, et al. Measuring social health in the patient-reported outcomes measurement information system (PROMIS): item bank development and testing. Quality of Life Research. 2010;19:1035-44.
164. NHS Improvement. National Schedule of Reference Costs 2017/18 2018 [Available from: https://webarchive.nationalarchives.gov.uk/20200501111106/https://improvement.nhs.uk/resources/reference-costs/].
165. Yeo A, Legard R, Keegan J, Ward K, McNaughton Nicholls C, Lewis J. In-depth interviews. In: Ritchie J, Lewis J, McNaughton Nicholls C, Ormston R, editors. Qualitative Research Practice: A guide for Social Science Students and Researchers. Second ed. London: SAGE Publications Ltd; 2014.
166. Brod M, Tesler L, Christensen T. Qualitative research and content validity: developing best practices based on science and experience. Quality of Life Research. 2009;18:1263-78.
167. Beecham J, Knapp M. CSRI 1992 [Available from: http://www.dirum.org/instruments/details/44, Date accessed: 20/06/2020].
168. McIntosh E, Gray A. PDMED resource utilisation questionnaire 2000 [Available from: http://www.dirum.org/instruments/details/29 Date accessed: 20/06/2020].
169. McPherson K. HOPEFUL Study Patient Questionnaire 2003 [Available from: http://www.dirum.org/instruments/details/34, Date accessed: 20/06/2020].

170. Mugford M. CESAR Patient Cost Questionnaire 2003 [Available from: http://www.dirum.org/instruments/details/31, Date accessed: 20/06/2020].
171. Wordsworth S. An annotated cost questionnaire for completion by patients 2001 [Available from: http://www.dirum.org/instruments/details/28, Date accessed: 20/06/2020].
172. Grant AM, Boachie C, Cotton SC, Faria R, Bojke L, Epstein DM, et al. Clinical and economic evaluation of laparoscopic surgery compared with medical management for gastro-oesophageal reflux disease: 5-year follow-up of multicentre randomised trial (the REFLUX trial). Health Technology Assessment. 2013;17(22).
173. White J, Withers K, Lencioni M, Carolan-Rees G, Wilkes A, Wood K, et al. Cardiff cardiac ablation patient-reported outcome measure (C-CAP): validation of a new questionnaire set for patients undergoing catheter ablation for cardiac arrhythmias in the UK. Quality of Life Research. 2016;25(6):1571-83.
174. NHS. The NHS Long Term Plan 2019 [Available from: https://www.longtermplan.nhs.uk/publication/nhs-long-term-plan/].
175. Garfield K, Husbands S, Thorn JC, Noble S, Hollingworth W. Development of a brief, generic, modular resource-use measure (ModRUM): cognitive interviews with patients. BMC Health Services Research. 2021;21.
176. Owen-Smith A, Coast J. Understanding data collection: Interviews, focus groups and observation. In: Coast J, editor. Qualitative methods in health economics. London, UK: Rowman & Littlefield Ltd; 2017. p. 59-91.
177. International Q. NVivo. 2020.
178. Coast J, Kinghorn P, Owen-Smith A. Designing qualitative research studies in health economics. In: Coast J, editor. Qualitative Methods in Health Economics. London, UK: Rowman & Littlefield Ltd; 2017. p. 19-41.
179. Public Health England. National General Practice Profiles 2019 [Available from: https://fingertips.phe.org.uk/profile/general-practice/data#page/8/page-options/map-ao-4].
180. Owen-Smith A, Coast J. Understanding sampling and recruitment. In: Coast J, editor. Qualitative methods in health economics. London, UK: Rowman & Littlefield Ltd; 2017.
181. National Institute for Health Research. Journals Library [Available from: https://www.journalslibrary.nihr.ac.uk/advancedsearch/].
182. EuroQol. EQ-5D: Support [Available from: https://euroqol.org/support/how-to-obtain-eq-5d/] Accessed on: 10th January 2021.
183. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA productivity cost questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value in Health. 2015;18(6):753-8.
184. Reilly M, Zbrozek A, Dukes E. The validity and reproducibility of a work productivity and activity impairment instrument. PharmacoEconomics. 1993;4:353-65.
185. Ipsos MORI Social Research Institute. GP patient survey 2020: technical annex. 2020.
186. Sahlqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D. Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial. BMC Medical Research Methodology. 2011;11:1-8.

187. Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. The Cochrane Database of Systematic Reviews. 2009;8(3):MR000008.
188. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data. Cham, Switzerland: Springer Nature; 2020.
189. Li J, Green M, Kearns B, Holding E, Smith C, Haywood A, et al. Patterns of multimorbidity and their association with health outcomes within Yorkshire, England: baseline results from the Yorkshire Health Study. BMC Public Health. 2016;16.
190. Huntley A, Lasserson D, Wye L, Morris R, Checkland K, England H, et al. Which features of primary care affect unscheduled secondary care use? a systematic review. BMJ Open. 2014;4:e004746.
191. Department of Health. Long term conditions compendium of information. third ed. London 2012.
192. Wammes J, van der Wees P, Tanke M, Westert G, Jeurissen P. Systematic review of high-cost patients' characteristics and healthcare utilisation. BMJ Open. 2018;8:e023113.
193. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Family Practice. 2016;17.
194. Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? an analysis of routinely collected UK general practice data BMJ Open. 2013;3.
195. Morris S, Sutton M, Gravelle H. Inequity and inequality in the use of health care in England: an empirical investigation. Social Science & Medicine. 2005;60:1251-66.
196. Steichen TJ, Cox NJ. Concordance correlation coefficient. Stata Technical Bulletin, StataCorp LP. 1999;8(43).
197. Cicchetti DV. The precision of reliability, validity estimates revisited: distinguishing between clinical and statistical significance of sample size requirements. Journal of Clinical and Experimental and Neuropsychology. 2001;23(5):695-700.
198. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 1986;327(8476):307-10.
199. NHS Digital. Mean and median length of stay of hospital inpatient episodes 2019 [Available from: https://digital.nhs.uk/data-and-information/find-data-and-publications/supplementary-information/2019-supplementary-information-files/mean-and-median-length-of-stay-of-hospital-inpatient-episodes].
200. Curtis LA. Unit Costs of Health & Social Care 2013. University of Kent: PSSRU; 2013.
201. Curtis LA, Burns A. Unit Costs of Health & Social Care 2015. University of Kent: PSSRU; 2015.
202. Turner J, O'Cathain A, Knowles E, Nicholl J, Tosh J, Sampson F, et al. Evaluation of NHS 111 pilot sites 2012 [Available from: https://www.sheffield.ac.uk/polopoly_fs/1.227404%21/file/NHS_111_final_report_August_2012.pdf].
203. NHS Business Services Authority. Prescription Cost Analysis - England 2019. 2020.
204. Office for National Statistics. Population estimates - local authority based by single year of age 2020 [Available from:

http://www.nomisweb.co.uk/query/construct/summary.asp?menuopt=200&subcomp=
].
205. Shields GE, Elvidge J. Challenges in synthesising cost-effectiveness estimates. Systematic Reviews. 2020;9(289).
206. Clinical Practice Research Datalink. [Available from: https://www.cprd.com/].
207. PECUNIA Group. PECUNIA RUM instrument 2021 [Available from: https://www.pecunia-project.eu/tools/rum-instrument Date accessed: 15/11/2021].
208. Thorn JC, Turner E, Hounsome L, Walsh E, Donovan JL, Verne J, et al. Validation of the Hospital Episode Statistics outpatient dataset in England. PharmacoEconomics. 2016;34:161-8.
209. Garfield K, Husbands S, Thorn JC, Noble S, Hollingworth W. Development of a brief, generic, modular resource-use measure: qualitative interviews with health economists. Value in Health. 2020;23(1):S292-S3, PNS48.
210. Garfield K, Husbands S, Thorn JC, Noble S, Hollingworth W. Developing items into questions for a new modular resource-use questionnaire. Trials. 2019;20(supp.1):PS1C - O3.
211. Hays RD, Weech-Maldonado R, Teresi JA, Wallace SP, Stewart AL. Commentary: copyright restrictions versus open access to survey instruments. Medical Care. 2018;56(2):107-10.
212. McKenna SP. It is time for an organised, scientific approach to the application of patient-reported outcome measures in clinical studies and trials. Clinical Rheumatology. 2018;36:611-5.

Appendix 1. Review of RUMs

This appendix contains documents related to the review of RUMs reported in **Chapter 3**.

A1.1. Search strategy

	Medline, Embase and PsycINFO via OvidSP
Part 1: Development	
develop*	(“develop*” OR “create” OR “produce” OR “write” OR "questionnaire design" OR "design a questionnaire" OR “devise” OR “prepare”).tw.
create	
produce	
write	
questionnaire design	
design a questionnaire	
devise	
prepare	
Part 2: Completion	AND
self complet*	("self complet*" OR "self report*" OR "patient report*" OR "patient complet*" OR "completion by patient*").tw.
self report*	
patient report*	
patient complet*	
completion by patient*	
Part 3: Outcome	AND
resource us*	("resource us*" OR "resource utili#ation" OR "service us*" OR "service utili#ation" OR "health care us*" OR "healthcare us*" OR "health care utili#ation" OR "healthcare utili#ation" OR "patient cost" OR "health care cost" OR "healthcare cost" OR "health data" OR "health service data" OR "health care data" OR "healthcare data").tw.
resource utili#ation	
service us*	
service utili#ation	
health care us*	
healthcare us*	
health care utili#ation	
healthcare utili#ation	
patient cost	
health care cost	
healthcare cost	
health data	
health service data	
health care data	
healthcare data	
Part 4: Measure	AND
questionnaire*	(questionnaire* OR instrument* OR survey* OR log* OR diary OR diaries OR measurement*).tw.
instrument*	
survey*	
log*	
diary	
diaries	
measurement*	

Appendix 2. Designing ModRUM

This appendix contains documents related drafting the initial versions of ModRUM (**Chapter 4)**.

A2.1. ModRUM core and depth module prototype (first version developed with the research team)

Resource use questionnaire

We would like you to answer some questions about the health care you have used [because of XXXXX/for any reason] [in the last X months/since XXXXX]. We only want you to include health care you had as an NHS patient.

It is important for us to find out what health care you have and have not used, so please answer all of the questions, even if your answer is zero. If you are unsure of an answer, please write your best guess.

1. How many times have you been to a hospital Accident and Emergency (A&E) department? ☐

2. How many times have you received care from a paramedic? ☐

3. How many times have you been to hospital for an outpatient appointment? ☐

For each outpatient appointment, please complete the table below:

Outpatient appointment	Clinic type e.g. xxxxx	Main test or surgical procedure performed e.g. xxxxx
1		
2		
3		
4		
5		
6		
7		
8		
For any other outpatient appointments, please complete Table 1 on page Y.		

Core and Depth bolt-ons v2.0 16/01/19

4. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)? ☐

For each visit to hospital for day case care, please complete the table below:

Hospital day case visit	Reason for visit e.g. xxxxx	Main test or surgical procedure performed e.g. xxxxx
1		
2		
3		
4		
For any other day case stays, please complete Table 2 on page Y.		

5. How many times have you been to hospital for an overnight stay? ☐

For each overnight stay in hospital, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Reason for stay e.g. xxxxx	Main test or surgical procedure performed e.g. xxxxx
1	☐ nights		
2	☐ nights		
3	☐ nights		
4	☐ nights		
For any other overnight stays in hospital, please complete Table 3 on page Y.			

Core and Depth bolt-ons v2.0 16/01/19

A2.1 continued

6. How many times have you visited a doctor (GP) at a GP surgery or health clinic? ☐

7. How many times have you visited another health care professional (e.g. a nurse or physiotherapist) at a GP surgery or health clinic? ☐

8. How many times has a doctor (GP) visited you at your home? ☐

9. How many times has another health care professional (e.g. a nurse or health visitor) visited you at your home? ☐

10. How many times have you had an appointment with a health care professional (e.g. GP or therapist) online or by telephone? ☐

For each health care professional, please complete the table below:

Health care professional	Health care professional seen e.g. xxxxx	Number of appointments
1		☐ appointments
2		☐ appointments
3		☐ appointments
4		☐ appointments
5		☐ appointments
For any other appointments online or by telephone, please complete Table 4 on page Y.		

11. How many medications have been prescribed for you? ☐

For each medication that has been prescribed for you, please complete the table below:

Prescribed medication	Name of prescribed medication e.g. xxxxx	Number of days you used the medication
1		☐ days
2		☐ days
3		☐ days
4		☐ days
5		☐ days
6		☐ days
7		☐ days
8		☐ days
For any other medications, please complete Table 5 on page Y.		

For the number of days you used the medication, please put:
XX if you used the medication every day for YYYY
XX if you used the medication once per week for YYYY

Do you pay for your prescribed medications? Yes ☐ No ☐

A2.2. ModRUM core module continuation table (first version developed with the research team)

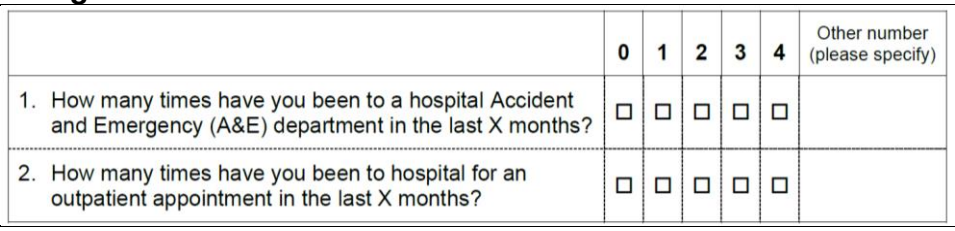
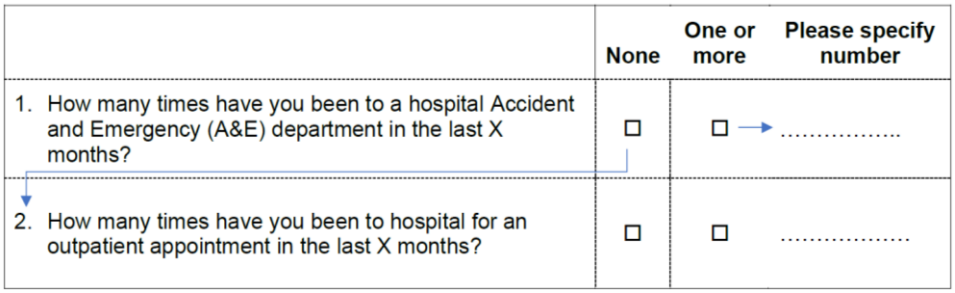
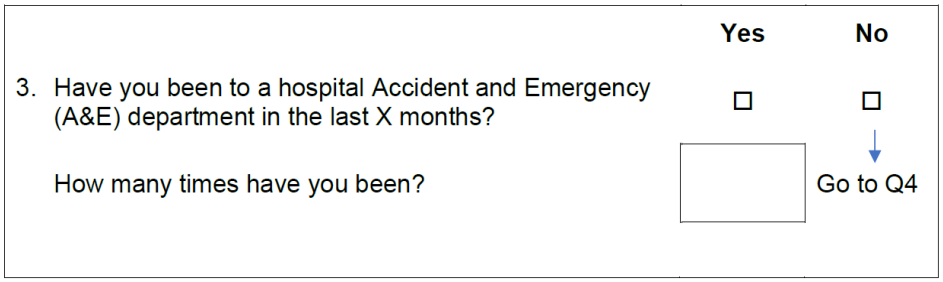
Core module continuation table

Table 1. For each overnight stay, please write down the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital
5	☐ nights
6	☐ nights
7	☐ nights
8	☐ nights
9	☐ nights
10	☐ nights
11	☐ nights
12	☐ nights
13	☐ nights
14	☐ nights
15	☐ nights

Core module continuation table, Version 2.0, 16/01/18

A2.3. Alternative ideas for response options suggested by PROM developer 1

Suggestion	Design
1	
2	
3	

A2.4. ModRUM logo

ModRUM

Modular Resource-Use Measure

Appendix 3. Qualitative interviews with health economists

This appendix contains documents related to interviews with health economists (**Chapter 5**).

A3.1. Health economics expert information sheet

University of BRISTOL

Development of a modular resource-use questionnaire
Health economics expert information sheet

What is the aim of the research?
The aim of this research is to develop a standardised resource-use questionnaire (RUQ) that can be used in clinical trials to collect information on the healthcare resources that trial participants use. The draft RUQ has a modular design, with a core module that includes a minimum set of resources that should be collected in all trials. Depth modules have been designed to be slotted into the core module to capture more details on resources used if additional detail is likely to be relevant for a trial. In the future, breadth modules will be developed to capture resources (e.g. in other sectors beyond health) not collected in the core module. Users will also have the option to add any resources that are pertinent to their trial. The draft RUQ is designed for paper completion. An online version may be developed and tested in future. By consulting both health economists and patients during the development of the RUQ, we are aiming for the RUQ to be easy for patients to complete and straightforward for health economists to analyse and cost.

What would my role be?
As a health economic expert advisor, we would like you to review and provide feedback on two draft versions of the RUQ. The first version includes the core module only and the second version includes the core module with some depth modules slotted in.

The RUQs will be sent to you in advance of a short (approximately 30 minutes) telephone call where you will be asked to provide feedback on:

- Whether the questions adequately cover the items they purport to measure (see the annotated questionnaire for questions and the items they aim to measure).
- Whether the information captured would be adequate for costing purposes in an economic evaluation.
- The presentation of the RUQ in general.

Following the telephone call, you may send any further feedback by email, and with your permission, we may contact you again, should any clarifications be required.

How was the RUQ developed?
RUQ Items were identified in a Delphi consensus study[1] where health economists agreed on 10 items for the core module of a new RUQ, that should be collected in all trials. They also identified additional items that could be included in depth or breadth bolt-on modules.

For each item identified in the Delphi, a review of the wording, formatting and layout of questions contained within existing RUQs was performed. Based on this review and the research team's opinions, questions were formulated for the core and depth module items.

Lead researcher contact details for further information
Kirsty Garfield
1-5 Whiteladies Road, Bristol, BS8 1NU
Email: Kirsty.Garfield@bristol.ac.uk

[1] Thorn, J.C., et al., *Core Items for a Standardized Resource Use Measure (ISRUM): Expert Delphi Consensus Survey*. Value in Health, 2017.

IRAS: 241489 HE expert information v1.0 17-01-19

A3.2. Health economic expert consent form

University of BRISTOL

HEALTH ECONOMIC EXPERT CONSENT FORM
Project title: Development of a modular resource-use questionnaire
Researcher: Kirsty Garfield

Please initial box

1. I confirm that I have read and understood the Health economics expert information sheet dated 17/01/19 (version 1.0). I have had the opportunity to consider the information, ask questions and have these questions answered. ☐
2. I understand that participation is voluntary and that I am free to withdraw at any time without giving a reason. ☐
3. I understand my role as an expert advisor and agree to take part in this research as a member of the health economics expert advisory group. ☐

Optional

4. I agree to having the feedback telephone call audio-recorded. ☐
5. I agree to being contacted again by the researcher, should they need to clarify any additional feedback I send them by email. ☐
6. I agree to be acknowledged by name in the outputs of this research. ☐

Name of expert advisor	Date	Signature
Name of person taking consent	Date	Signature

IRAS ID: 241489 HE expert consent form Version 1.0 17-01-19

A3.3. ModRUM core module (version sent to health economic experts for interviews)

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used, [because of XXXXX/for any reason], [in the last X months/since XXXXX]. We only want you to include healthcare you have used as an NHS patient.

Please answer all the questions, even if your answer is zero, as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

[In the last X months/since XXXXX,] [because of XXXXX,]						Other number (please specify)
1. How many times have you been to a hospital Accident and Emergency (A&E) department?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	☐
2. How many times have you been to hospital for an outpatient appointment?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	☐
3. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	☐
4. How many times have you been to hospital for an overnight stay?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	☐

For each stay, please write the number of nights you stayed in hospital:

Stay	Number of nights spent in hospital	Stay	Number of nights spent in hospital
1	☐ nights	6	☐ nights
2	☐ nights	7	☐ nights
3	☐ nights	8	☐ nights
4	☐ nights	9	☐ nights
5	☐ nights	10	☐ nights

A3.3 continued

[In the last X months/since XXXXX,] [because of XXXXX,]

						Other number (please specify)
5. How many times have you visited a GP surgery or health clinic to see:						
a doctor (GP)?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐
another healthcare professional (e.g. a nurse or physiotherapist)?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐
6. How many times have you had a home visit from:						
a doctor (GP)?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐
another healthcare professional (e.g. a nurse or physiotherapist)?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐
7. How many medications have been prescribed for you? (e.g. if you have been prescribed Medication A once and Medication B twice, your answer should be 3)	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐

A3.4. ModRUM core and depth modules (version sent to health economic experts for interviews)

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used, [because of XXXXX/for any reason], [in the last X months/since XXXXX]. We only want you to include healthcare you have used as an NHS patient.

Please answer all the questions, even if your answer is zero, as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

[In the last X months/since XXXXX,] [because of XXXXX,]						Other number (please specify)
1. How many times have you been to a hospital Accident and Emergency (A&E) department?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	
2. How many times have you received care from a paramedic?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	
3. How many times have you been to hospital for an outpatient appointment?	0 ☐	1 ☐	2 ☐	3 ☐	4 ☐	

For each appointment, please complete the table below:

Outpatient appointment	Clinic type e.g. [Dermatology]	Main test or surgical procedure performed e.g. [Removal of a skin lesion]
1		
2		
3		
4		
5		
6		
If you need more space, please use Table 1 on page 5.		

A3.4 continued

[In the last X months/since XXXXX,] [because of XXXXX,]

Other number (please specify)

4. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

For each visit, please complete the table below:

Hospital day case visit	Reason for visit e.g. [Stomach ache]	Main test or surgical procedure performed e.g. [Endoscopy with biopsy]
1		
2		
3		
4		
If you need more space, please use Table 2 on page 6.		

[In the last X months/since XXXXX,] [because of XXXXX,]

Other number (please specify)

5. How many times have you been to hospital for an overnight stay?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

For each stay, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Reason for stay e.g. [Knee pain]	Main test or surgical procedure performed e.g. [Knee replacement]
1	☐ nights		
2	☐ nights		
3	☐ nights		
4	☐ nights		
If you need more space, please use Table 3 on page 7.			

A3.4 continued

[In the last X months/since XXXXX,] [because of XXXXX,]

Other number (please specify)

6. How many times have you visited a GP surgery or health clinic to see:

	0	1	2	3	4	
a doctor (GP)?	☐	☐	☐	☐	☐	☐
another healthcare professional (e.g. a nurse or physiotherapist)?	☐	☐	☐	☐	☐	☐

7. How many times have you had a home visit from:

	0	1	2	3	4	
a doctor (GP)?	☐	☐	☐	☐	☐	☐
another healthcare professional (e.g. a nurse or physiotherapist)?	☐	☐	☐	☐	☐	☐

8. How many times have you had an online or telephone appointment with a healthcare professional (e.g. [GP or therapist])?

0	1	2	3	4	
☐	☐	☐	☐	☐	☐

For each healthcare professional, please complete the table below:

Healthcare professional	Healthcare professional title e.g. GP, therapist	Number of appointments
1		☐ appointments
2		☐ appointments
3		☐ appointments
4		☐ appointments
5		☐ appointments
6		☐ appointments
7		☐ appointments

If you need more space, please use Table 4 on page 8.

A3.4 continued

[In the last X months/since XXXXX,] [because of XXXXX,]

Other number (please specify)

9. How many medications have been prescribed for you? (e.g. if you have been prescribed Medication A once and Medication B twice, your answer should be 3)

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ ☐

For each medication, please complete the table below:

Prescribed medication	Name of prescribed medication e.g. [Tramadol] (if you don't know the name, please put the type of medication e.g. [Painkiller])	Number of days you used the medication
1		☐ days
2		☐ days
3		☐ days
4		☐ days
5		☐ days
6		☐ days
7		☐ days
8		☐ days
9		☐ days
10		☐ days
If you need more space, please use Table 5 on page 9.		

For the number of days you used the medication, please put:
[91/183] if you used the medication every day for [3/6] months
[13/26] if you used the medication once per week for [3/6] months

A3.4 continued

Tables to complete if you need more space

Table 1

For each outpatient appointment, please complete the table below:

Outpatient appointment	Clinic type e.g. [Dermatology]	Main test or surgical procedure performed e.g. [Removal of a skin lesion]
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Return to question 4 on page 2

A3.4 continued

Table 2

For each visit to hospital for day case care, please complete the table below:

Hospital day case visit	Reason for visit e.g. [Stomach ache]	Main test or surgical procedure performed e.g. [Endoscopy with biopsy]
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		

Return to question 5 on page 2.

A3.4 continued

Table 3

For each overnight stay in hospital, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Reason for stay e.g. [Knee pain]	Main test or surgical procedure performed e.g. [Knee replacement]
5	☐ nights		
6	☐ nights		
7	☐ nights		
8	☐ nights		
9	☐ nights		
10	☐ nights		
11	☐ nights		
12	☐ nights		
13	☐ nights		
14	☐ nights		
15	☐ nights		
16	☐ nights		
17	☐ nights		
18	☐ nights		

Return to question 6 on page 3.

A3.4 continued

Table 4

For each healthcare professional seen by telephone or online, please complete the table below:

Healthcare professional	Healthcare professional title e.g. [GP, therapist]	Number of appointments
8		☐ appointments
9		☐ appointments
10		☐ appointments
11		☐ appointments
12		☐ appointments
13		☐ appointments
14		☐ appointments
15		☐ appointments
16		☐ appointments
17		☐ appointments
18		☐ appointments
19		☐ appointments
20		☐ appointments
21		☐ appointments

Return to question 9 on page 4.

A3.4 continued

Table 5

For each medication that has been prescribed for you, please complete the table below:

Prescribed medication	Name of prescribed medication e.g. [Tramadol] (if you don't know the name, please put the type of medication e.g. [Painkiller])	Number of days you used the medication
11		☐ days
12		☐ days
13		☐ days
14		☐ days
15		☐ days
16		☐ days
17		☐ days
18		☐ days
19		☐ days
20		☐ days
21		☐ days
22		☐ days
23		☐ days
24		☐ days

A3.5. Topic guide for health economic expert interviews

Development of a modular resource-use questionnaire
Topic Guide for feedback calls with health economic experts

1. **Introduction**
   a. **Explain the aims of the research**
      - Thank you for acting as an advisor for this research.
      - The aim of the research is to develop a standardised resource-use questionnaire (RUQ) that can be used in clinical trials to collect information on the health care resources trial participants use. The core module provides a minimum set of resources that should be collected in all trials. The depth bolt-on version of the questionnaire allows the use to add depth to questions included in the core module. Users will also be able to add anything that is pertinent to their trial.
      - The aim of this piece of work is to find out your opinions on a draft version of the questionnaire and to find out if you think the questions are suitable for costing.
   b. **Explain what will happen during the call**
      - Firstly, I would like to go through any feedback you have on each question and then I will ask for some general feedback on the whole questionnaire.
   c. **Reminder about anonymity**
      - I may include your feedback when I write up the findings from this research. Your feedback will be completely anonymous as we will not include your name or any details that may identify you in anything we write.
      - We would however like to acknowledge your contribution in the theses, paper(s) and presentations resulting from this research. Confirm whether they agreed or disagreed to this on their consent form.
   d. **Check that the expert is OK to have the call recorded**
      - You indicated on your consent form that you agree to the call being recorded. Can I check that you are still happy for me to record this call? This will help me to record your responses accurately.
      - *Start recording*

2. **Feedback session**
   a. **Feedback on individual questions**
   Examples questions:
      - We would like to find out if the RUQ question adequately covers the item it purports to measure
        - Do you think this question adequately captures information on [item]?
          - If no, please could you explain why?
          - If no, how would you reword or extend this question to capture the correct information?
      - Are there any terms in this question that you think will be problematic?
        - If yes, why do you think [term] will be problematic?
        - If yes, is there another term you would include?
      - Would you be able to use the information captured in this question for costing in an economic evaluation?

A3.5 continued

b. Feedback on the whole questionnaire

Examples questions:

- How clear do you think the instructions are at the start of the questionnaire?
  - How would you improve the instructions?
  - Is there anything missing from the instructions?
- How did you find the order of the questions?
  - Can you think of a better question order?
- Did you find any overlap between the questions?
- Are there any questions in the core module that you think would be more appropriate as a depth-adding question?
- Are there any depth-adding questions that you think would be more appropriate in the core module?
- Are there any important aspects of NHS care that are missing from the questionnaire that would need to be added in for RCTs you work on?
- What do you think about the continuation sheets? Are they necessary?

3. Call end

- If you have any other comments or notes that you have made on the RUQ, that you would like to share with me, you can send them to me via email. If you do send me any comments or notes, are you happy for me to re-contact you if I need to clarify anything.
- Thanks for acting as an advisor for this research.

Appendix 4. Cognitive interviews with patients

This appendix contains documents related to interviews with patients (**Chapter 6**).

A4.1. Patient Information Sheet for interviews with patients

University of BRISTOL

Population Health Sciences
Bristol Medical School
University of Bristol
1-5 Whiteladies Road
Bristol
BS8 1NU

PATIENT INFORMATION SHEET
Developing a modular resource-use questionnaire for use in randomised controlled trials

A team of researchers at the University of Bristol would like to invite you to take part in a research study. The aim of this research is to develop a questionnaire that can be used in studies to ask study participants about the healthcare resources, such as GP appointments, they have used.

For this study we would like you to attend ONE interview, at a time and place most convenient to you.

Before you decide whether to take part, it is important for you to understand why the research is being done and what taking part will involve for you. This leaflet explains the purpose of the research and what we would like you to do.

1. What is the purpose of this study?
The purpose of this study is to develop a questionnaire that can be completed by study participants to report the healthcare resources (e.g. hospital outpatient appointments) they use. These details are collected in research studies so that we can estimate the cost to the NHS of the services or interventions patients use. As part of the development we want to understand how best to ask questions on healthcare use, so the answers we get are as accurate as possible.

2. Why have I been invited to take part?
You have been invited because your GP practice is supporting this study. Around 25 patients will take part in this study.

3. Do I have to take part?
No. It is up to you to decide whether to take part. If you want to take part, we will ask you to sign a consent form. You can stop taking part in the study at any time without giving a reason. Deciding not to take part or withdrawing will not affect the care you receive.

4. What will happen if I take part?
I will ask you to 'think-aloud' as you complete questions on your healthcare use. By 'thinking-aloud' I would like you to talk me through what you are thinking as you answer questions. 'Thinking-aloud' will be explained more fully at the start of the interview. At the end of the interview we will have a chat about how you found the questions in general. With your permission, I will audio-record the interview to make sure your views are accurately reported. The interview will last up to 1 hour. The audio-recording will be securely sent to a University of Bristol approved transcription company, who will transcribe the interview and securely return the transcription to the study research team at the University of Bristol.

A4.1 continued

5. Will I receive any payment for taking part?
No, we are not able to offer any expenses or payments to patients who participate in the study. However, after you have completed the interview, we will give you a £10 love2shop voucher as a thank you for taking part, which can be used in many shops and restaurants on the high street.

6. What are any possible benefits of taking part?
Your views as a patient are important and will help us understand if the questions we ask patients are easy to understand. We can't promise the study will help you, but the information you give us will improve the design of the final questions which may be used in future research studies. The questionnaire will help to provide accurate information on whether NHS services are value for money.

7. Will my taking part in this study be kept confidential?
Yes, your data will be analysed and stored electronically on a secure computer network at the University of Bristol. Any hard copies of this information will be locked in a secure filing cabinet at the research base (University of Bristol). Your personal data will be stored separately from research findings and will only be accessed by the research team. Any information that could identify you will be removed from the data before the findings are seen by others, and personal data will not be used in research reports. The handling, processing, storage, and destruction of these data are compliant with the Data Protection Act 2018.

8. How will my personal data be used and what are my rights?
The University of Bristol is the sponsor for this study based in the United Kingdom. We will be using information from you in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The University of Bristol will keep identifiable information about you for 10 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting the lead researcher (Kirsty Garfield).

The University of Bristol will collect information from you for this research study in accordance with our instructions.

The University of Bristol will use your name and contact details to contact you about the research study and to oversee the quality of the study. Individuals from University of Bristol and regulatory organisations may look at your research records to check the accuracy of the research study. The only people in the University of Bristol who will have access to information that identifies you will be people who need to contact you to discuss the research study or audit the data collection process.

A4.1 continued

9. What will happen to the results of this research study?
The results will be written up for publication and may be presented at conferences. You will not be identified in any publication or presentation.

If you would like a summary of the results, you can indicate this on the consent form and we can email you or post a leaflet to your home.

10. Who is organising and funding this study?
This study is being carried out by researchers at the University of Bristol who are members of the MRC ConDuCT-II Hub for Trials Methodology Research. This work is supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- PhD Award R9).

Research team
The research team includes:
- Miss Kirsty Garfield, the lead researcher, PhD student and Research Associate in Health Economics at the University of Bristol.
- Dr Joanna Thorn, Senior Research Associate in Health Economics at the University of Bristol.
- Dr Sian Noble, Senior Lecturer in Health Economics at the University of Bristol.
- Dr Samantha Husbands, Senior Research Associate in Qualitative Research & Health Economics at the University of Bristol.
- Professor Will Hollingworth, Professor of Health Economics at the University of Bristol.

11. Who has reviewed this study?
All research in the NHS has been reviewed and given a favourable opinion by the South Central - Berkshire B Research Ethics Committee to protect your safety, rights, wellbeing and dignity.

Contact details for further information
For further information about the research, please contact the lead researcher:

Kirsty Garfield
1-5 Whiteladies Road
Bristol
BS8 1NU
Email: Kirsty.Garfield@bristol.ac.uk
Telephone: 0117 428 3129

For complaints about the study please contact:

Will Hollingworth
Email: William.Hollingworth@bristol.ac.uk

Thank you for taking the time to read this Participant Information Sheet and for considering whether to take part in this study.
If you decide to take part you will be given a copy of this information sheet and a signed consent form to keep.

IRAS ID: 241489 Study 1 patient information sheet Version 1.1 08-07-19

A4.2. Reply Form for interviews with patients

Development of a healthcare use questionnaire

Research Study Reply Form

University of BRISTOL

If you are willing to be contacted about this study, please complete this form and return it to your GP receptionist. In addition to contact details, we are also asking for information on sex, age group, ethnic group, long-term conditions, age on leaving full time education and healthcare use. The reason we are asking for these details is to ensure we speak to a broad range of individuals.

Name: .....

Phone number: .....

1. Sex	Female	☐	Prefer not to say	☐
	Male	☐		
2. Age group	18 – 30	☐	66 – 75	☐
	31 – 45	☐	76 or older	☐
	46 – 55	☐	Prefer not to say	☐
	56 – 65	☐		
3. Ethnic group	White	☐	Asian/Asian British	☐
	Mixed/Multiple ethnic groups	☐	Black/African/ Caribbean/Black British	☐
	Other	☐	If other, please specify:	
	Prefer not to say	☐		
4. Long-term conditions, e.g. diabetes, asthma, heart disease	None	☐	More than one	☐
	One	☐	Prefer not to say	☐
5. Age on leaving full time education	16 or under	☐	19 or over	☐
	17 or 18	☐	Prefer not to say	☐
6. Number of times you have used primary healthcare services (e.g. appointments at GP practice) in the last 3 months	None	☐	Three or four times	☐
	One or two times	☐	More than four times	☐
7. Number of times you have used secondary healthcare services (e.g. hospital visits or stays) in the last 3 months	None	☐	Three or four times	☐
	One or two times	☐	More than four times	☐

8. Declaration

I am willing to be contacted by Kirsty Garfield (lead researcher) on the contact details provided above. I understand that completing this form does not obligate me to be interviewed. I understand that the personal details I have provided in this form will not be shared with anyone else and the personal contact details will not be retained following interview. ☐

IRAS: 241489 Study 1 reply form Version 1.0 17-01-19

A4.3. Consent form for interviews with patients

Participant Identification Number:

University of BRISTOL

CONSENT FORM

Project title: Developing a modular resource-use questionnaire for use in randomised controlled trials

Researcher: Kirsty Garfield

Please initial box

1. I confirm that I have read and understood the information sheet dated 08/07/19 (version 1.1) for the above study. I have had the opportunity to consider the information, ask questions and have these questions answered. ☐
2. I understand that participation is voluntary and that I am free to withdraw at any time without giving a reason, and without my medical care or legal rights being affected. ☐
3. I understand that data collected during the study may be looked at by the study research team at the University of Bristol or by individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐
4. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers. ☐
5. I understand that you will write a report about the research findings, but that it will not be possible to identify me from the report. ☐
6. I agree to the study publishing anonymous quotations from the interview in presentations and publications. ☐
7. I agree to having the interview audio-recorded. ☐
8. I agree to take part in the above study. ☐
9. Optional: I would like to receive a summary of the results once the project is complete. ☐

Name of participant	Date	Signature
Name of person taking consent	Date	Signature

One completed copy for the research team, one for the participant

IRAS ID: 241489 Study 1 consent form Version 1.1 08-07-19

A4.4. Topic guide for interviews with patients

Development of a healthcare use questionnaire
Topic Guide: Think-aloud interviews with retrospective probing with patients
To be used in conjunction with resource-use questionnaire version 1.2 dated 24/01/20.

Aim of think-aloud interviews with retrospective probing
To explore through think-aloud interviewing with retrospective probing, how people complete a new resource-use questionnaire.

1. **Introduction** (approximately 5 minutes)
   a. **Explain the aims of the study**
   - Thank you for agreeing to take part in this interview, this research is part of my PhD research that I am completing at the University of Bristol.
   - Aim of the research: develop a questionnaire for clinical trials to collect information on healthcare resources (e.g. GP visits or hospital appointments).
   - Rationale for the research: develop a questionnaire that is easy for patients to complete and captures accurate information.
   - Aim of this study: understand what patients are thinking when they answer the questionnaire in order to improve the questionnaire.

   b. **Explain what will be asked about during this interview**
   - You will be asked to:
     - Fill in a questionnaire, 'Think-aloud' while filling in the questionnaire: explain 'think-aloud', warm-up, Answer questions on how you found the questionnaire.

   c. **Check that the participant is OK to have the interview recorded**
   - Reiterate why the interview is recorded, check participant is OK to be recorded.

   d. **Reminder about anonymity and withdrawing from the study**
   - Quotes may be used when we write-up the study, but they will be anonymous.
   - You can stop the interview at any time or withdraw.
   - Explain option to destroy recordings and questionnaires if participant withdraws.

   **Complete consent forms*

2. **Warm-up exercise** (approximately 5 minutes)
   - 'Thinking-aloud' can feel unfamiliar, start with warm-up.
   - Warm-up: counting windows in your home, tell me what you are seeing and thinking, I won't interrupt unless you are silent for more than 10 seconds.
   - Prompts to 'think-aloud' as required, provide feedback following warm-up.

   **Start recording*

3. **Think-aloud exercise** (approximately 20 minutes)
   - Reminders to 'think-aloud', say aloud your thought process from reading each question to writing in your final answer, don't plan what you are going to say,
   - Act as though you are alone and are constantly 'thinking-aloud' your thoughts, try to speak as clearly as possible, as I will be recording what you say.
   - So that I don't distract you, I will sit out of your line of sight. If you are silent for a long time, I will prompt you to continue 'thinking-aloud'.
   - Start 'think-aloud' exercise.
   - Prompts if necessary: 'think-aloud' from beginning to end, speak as clearly as possible, no need to plan, rush, or go any slower than normal.

A4.4 continued

4. **Follow-up questions** (approximately 10 minutes)
   - Thank you, now I would like to ask question about how you found completing the questionnaire.
   - [Examples]
     - How did you find completing the questionnaire?
     - You seemed a bit unclear about the instructions. What was it about them that you didn't understand?
     - You seemed to hesitate on [question x]. Why was that?
       - Did you find answering the question difficult?
         - If yes, what made the question difficult to answer?
     - You seemed a bit unsure by what was meant by the term [term]
       - What do you understand by that word?
       - How did you interpret the word in your answer?
       - How did that affect how you answered the question?
     - We would like to find out more about what the term [term] means to you
       - What is your understanding of the term [term]?
       - Is there an alternative term you would use to better describe [term]?
5. **Semi-structured interview** (approximately 10 minutes)
   - Thank you, now I would like to ask some general questions about the questionnaire.
   - [Examples]
     - Do you think the instructions at the beginning of the questionnaire were easy to follow?
     - Did you find it helpful to have the instructions in the box?
     - Did you find it helpful to have important terms in bold?
     - What did you think about the text at the end of the question?
     - What do you think about the layout of the questionnaire?
     - Did you find any questions off-putting?
     - Did you find any questions confusing?
     - How easy or difficult did you find answering the questionnaire overall?
     - How did you find the order of the questions?
     - Did you find any overlap between the questions?
     - How easy or difficult was it for you to remember the healthcare you have received in the last [X] months?
       - Were any healthcare resources easier or more difficult to remember than others over the time period?
       - Can you think of a recall period that might make it easier for you to answer the questions?
     - Do you think we should add section headings?
     - Were you comfortable with the length and time it took you to complete the questionnaire? *Acceptability check*
     - Is there anything else you would change?
     - Is there anything else that you would like to add (which you think might be useful to the research)?
6. **Interview end, greeting and leave**

A4.5. Core module with depth questions (pre-cognitive interviewing)

Participant Identification Number:

University of BRISTOL

Development of a healthcare use questionnaire

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used, for any reason, in the last 3 months. We only want you to include healthcare you have used as an NHS patient.

Please answer all the questions, even if your answer is zero, as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

In the last 3 months, Please tick or write the number of times

1. How many times have you been to a hospital Accident and Emergency (A&E) department? 0 1 2 3 4 ☐

2. How many times have you received care from the ambulance service:
   and were taken to hospital? 0 1 2 3 4 ☐
   but were not taken to hospital? 0 1 2 3 4 ☐

DRAFT

In the last 3 months, Please tick or write the number of times

3. How many times have you been to hospital for an outpatient appointment? 0 1 2 3 4 ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type e.g. Dermatology	Reason for visit e.g. Mole removal	Tests or surgical procedures performed (if applicable) e.g. Removal of a skin lesion
1			
2			
3			
4			
5			
6			

If you need more space, please use the box on page 8.

DRAFT

In the last 3 months, Please tick or write the number of times

4. How many times have you been to hospital for day case care (used a bed, but did not stay overnight)? 0 1 2 3 4 ☐

For each visit, please complete the table below:

Hospital day case visit	Department e.g. Gastroenterology	Reason for visit e.g. Stomach ache	Tests or surgical procedures performed (if applicable) e.g. Endoscopy with biopsy
1			
2			
3			
4			
5			

If you need more space, please use the box on page 8.

DRAFT

In the last 3 months, Please tick or write the number of times

5. How many times have you been to hospital for an overnight stay? 0 1 2 3 4 ☐

For each stay, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Department e.g. Orthopaedics	Reason for stay e.g. knee replacement	Tests or surgical procedures performed (if applicable) e.g. Knee replacement
1	☐ nights			
2	☐ nights			
3	☐ nights			
4	☐ nights			
5	☐ nights			

If you need more space, please use the box on page 8.

DRAFT

A4.5 continued

In the last 3 months,

Please tick or write the number of times

6. How many times have you had an appointment with a doctor (GP):

at a GP surgery or health centre? 0 1 2 3 4 ☐

at home? 0 1 2 3 4 ☐

over the telephone or online? 0 1 2 3 4 ☐

7. How many times have you had an appointment with a nurse:

at a GP surgery or health centre? 0 1 2 3 4 ☐

at home? 0 1 2 3 4 ☐

over the telephone or online? 0 1 2 3 4 ☐

DRAFT

In the last 3 months,

Please tick or write the number of times

8 How many times have you had contact with any other healthcare professionals (e.g. physiotherapist) or healthcare services (e.g. NHS 111):

at a GP surgery or health centre? 0 1 2 3 4 ☐

at home? 0 1 2 3 4 ☐

over the telephone or online? 0 1 2 3 4 ☐

For each healthcare professional you have seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or healthcare service contact	Healthcare professional seen (e.g. physiotherapist) or healthcare service used (e.g. NHS 111)
1	
2	
3	
4	
5	

If you need more space, please use the box on page 8.

DRAFT

In the last 3 months,

Please tick

9. Have you picked up or received any prescribed medications? Yes ☐ No ☐

For each prescribed medication, please complete the table:

Prescribed medication	Name of prescribed medication e.g. Tramadol (if you don't know the name, please put the type of medication e.g. Painkiller)	Number of times the medication has been picked up and/or received
1		☐ times
2		☐ times
3		☐ times
4		☐ times
5		☐ times
6		☐ times
7		☐ times
8		☐ times
9		☐ times
10		☐ times

If you need more space, please use the box on page 8.

DRAFT

If you run out of space in the tables, please use the box below to write about the other healthcare you have used:

DRAFT

A4.6. Core module scoring sheet

RUQ development study 'think-aloud' scoring sheet: core module

Participant ID			

Core module	1. A&E	2. OP	3.DC	4.1. IP	4.2. IP nights	5.1. GP at practice	5.2 GP at home	6.1. oHCP at practice	6.2 oHCP at home

Rater	KG[1]	JT[2]	SH[3]

No	Item	(a) Comp	(b) Retrieval	(c) Judge	(d) Response	(e) Struggle	(f) Comments
1	A&E						
2	Outpatient						
3	Day case						
4.1	Inpatient number						
4.2	Inpatient nights per stay						
5.1	GP at GP practice						
5.2	GP at home						
6.1	Other healthcare professional at GP practice						
6.2	Other healthcare professional at home						

A4.6 continued

a. *Comprehension*
   - any misunderstanding of a word, phrase, or response option
   - any words or phrase that the participant does not understand.

b. *Retrieval*
   - a recall error or a miscalculation of the time frame stated in the question,

c. *Judgement*
   - the participants response does not match that of the investigators intent for the question item,
   - any recalled relevant experiences that the participant questions as irrelevant or inadequate,

d. *Response*
   - participants desired response is missing from the written survey response choices,
   - any response which is felt to be socially desirable answer
   - participants response is inconsistent with the personal experience expressed
   - Participant's answer is inconsistent with previous answers

e. Struggle
   - the participant has had difficulty answering the question, even when their final response is correct.

	Error	Not error
Comprehension	Day care? I don't have any help at home.	Day care? Aah…day case care...no I have not had any day case visits.
Retrieval	I just can't remember whether I've had any outpatient appointments.	Outpatient appointments…let me see… there's a few to remember [pause] one for hip, one for physio…
Judgement	Other healthcare professionals at the GP practice, I've seen a healthcare assistant at the GP, but that isn't one of the examples, so I won't include it.	GP at the practice, I had one appointment with a GP from the GP practice, so one, but it was over the phone and not actually at the GP, so I won't include it.
Response	I've had 6 GP appointments, but there isn't an option for that, I'll just tick the big box [big box ticked]	I've had 6 GP appointments, but there isn't an option for that… oh, you're meant to write the number if more than 4. ['6' written in big box]
Struggle	Day case care? That's a difficult one because they are like outpatient appointments. But the part in brackets says, 'used a bed', I only did this on one occasion for my day surgery, the other visit was a check-up with my consultant. So, I'll just include the day surgery as I have included the consultant visit under outpatients. ['1' ticked]	Day case care. I know they are like outpatients but not quite the same as you usually have a procedure. The last day case visit I had was a year ago, so zero for that. ['0' ticked]

Appendix 5. Piloting with health economists

This appendix contains documents related to the health economist pilot of ModRUM (**Chapter 7**).

A5.1. Health economist initial contact email

Invitation to take part in a pilot for a PhD to develop a new generic RUQ

Dear [Expert name],

I am currently undertaking a PhD to develop and validate a new modular resource-use questionnaire (RUQ) called ModRUM. ModRUM is a brief RUQ that can be used as a basis to capture healthcare resource-use data from patients for economic evaluations in a wide range of randomised controlled trials of healthcare.

I have identified, from a search of the NIHR journals library, that you have recently had a grant funded that involves a trial based economic evaluation. I am emailing to ask if you would be willing to take part in a pilot study of ModRUM.

In the pilot, you would be asked to adapt ModRUM for [grant name and reference number] as you would normally prepare RUQs for your trials. This may involve working with a colleague (e.g. a more junior health economist working on the trial). We would then ask you to share the adapted ModRUM you have prepared and provide feedback in an online survey about the suitability of ModRUM for your trial. As this is a pilot and we have not yet completed validation work, we are not expecting you to use ModRUM in your trial. If you would like to take part, I will send you ModRUM, a User Guide and a link to the brief online survey.

In the information sheet attached, you will find some more information about the project, what your role would be, how your data would be used and some details on the development of ModRUM so far.

If you are willing to take part, please contact me by email (Kirsty.Garfield@bristol.ac.uk).

Kind regards
Kirsty

PhD supervisors
Joanna Thorn
Sian Noble
Samantha Husbands
Will Hollingworth

A5.2. Health economist pilot information sheet

University of BRISTOL

Development of a modular resource-use questionnaire for use in randomised controlled trials

Health economist pilot information sheet

What is the aim of the research?

The aim of this research is to develop a standardised resource-use questionnaire (RUQ) that can be used in clinical trials to collect information on the healthcare resources that trial participants use. ModRUM (Modular Resource-Use Measure) has a modular design, with a core module that includes a minimum set of resources that should be collected in all trials. If additional detail is likely to be relevant for a trial, depth questions can be slotted into the core module to capture more details on resources used. ModRUM has adaptable features including the recall period and the types of resources collected (all-cause versus condition-specific).

In the future, breadth modules will be developed to capture resources not collected in the core module (e.g. in other sectors beyond health). ModRUM is currently developed for paper completion. An online version may be developed and tested in future. By consulting both health economists and patients during the development of ModRUM, we are aiming for it to be easy for patients to complete and straightforward for health economists to prepare, analyse and cost.

What would my role in the research be?

We would like you to:

- Adapt and return ModRUM, using an NHS perspective, for your recently funded trial using the core module, depth questions and User Guide. We would like you to adapt ModRUM as you would usually prepare RUQs for trials. This may mean that you prepare it yourself, with a colleague or ask a colleague to prepare it.
- Complete an online survey about your experience with ModRUM. The survey includes questions on ease of preparation, suitability of ModRUM questions, experience of the User Guide and future use of ModRUM.
- As ModRUM is still in development, we request that you do not share or use the draft versions of ModRUM in your work.

Your contribution will:

- Allow us to assess the suitability of ModRUM for use in economic evaluations.
- Inform modifications to the formatting, structure and content of ModRUM.

A5.2 continued

How will my data be used?
Your data will be analysed and stored electronically on a secure computer network at the University of Bristol. Your personal data will be stored separately from research findings and will only be accessed by the research team. Any information that could identify you will be removed from the data before the findings are seen by others. The handling, processing, storage, and destruction of these data are compliant with the Data Protection Act 2018.

Bone fide researchers will be able to request access to research data. Personal data will not be made available. The research findings will be written up for publication and may be presented at conferences. You will not be identifiable in any publication or presentation.

Can I withdraw from the study?
You can withdraw from the study at any time. If you choose to withdraw from the study, we will keep the information about you that we have already obtained.

How was ModRUM developed?
RUQ items were identified in a Delphi consensus study[1] where health economists agreed on items for the core module of a new RUQ, that should be collected in all trials. They also identified additional items that could be included in depth or breadth bolt-on modules.

For each item identified in the Delphi, a review of the wording, formatting and layout of questions contained within existing RUQs was performed. Based on this review, the Delphi results and the research team's opinions, a prototype of ModRUM was developed.

ModRUM has been tested in interviews with health economists and patients, with refinements continually made in response to research findings. Interviews with health economists allowed us to assess the validity and appropriateness of ModRUM for costing in economic evaluations. 'Think-aloud' interviews with patients allowed us to assess the validity and acceptability of ModRUM.

Lead researcher contact details for further information
Kirsty Garfield
1-5 Whiteladies Road, Bristol, BS8 1NU
Email: Kirsty.Garfield@bristol.ac.uk

---

[1] Thorn, J.C., et al., *Core Items for a Standardized Resource Use Measure (ISRUM): Expert Delphi Consensus Survey*. Value in Health, 2017.

A5.3. ModRUM core module with depth questions (pre-health economist piloting)

ModRUM

Modular Resource-Use Measure

Participant Identification Number: _____

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used [because of XXXXX/for any reason], **[in the last X months/since XXXXX]. We only want you to include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

Questions 1 to 3 ask about emergency healthcare: A&E and ambulance

		Please tick or write the number of times
1	In the last [X] months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
2	In the last [X] months, how many times have you received care from the **ambulance service and were taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
3	In the last [X] months, how many times have you received care from the **ambulance service but were not taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

Questions 4 to 6 ask about healthcare at hospitals: outpatients, day case and overnight stays

		Please tick or write the number of times
4	In the last [X] months, how many times have you been to **hospital for an outpatient appointment** (e.g. [to see a consultant or hospital physiotherapist, or to have an x-ray])?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type	Tests or surgical procedures performed (if applicable)	Reason for visit
Example 1	[radiology]	[x-ray]	[x-ray for pain after a fall]
Example 2	[dermatology]	[removal of a skin lesion]	[mole removal]
Appointment 1			
Appointment 2			
Appointment 3			
Appointment 4			
Appointment 5			
Appointment 6			

If you need more space, please use the box on page 10.

A5.3 continued

Please tick or write the number of times

5 In the last [X] months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Hospital day case	Department	Tests or surgical procedures performed (if applicable)	Reason for day case
Example	[gastroenterology]	[endoscopy with biopsy]	[stomach ache]
Day case 1			
Day case 2			
Day case 3			
Day case 4			
Day case 5			

If you need more space, please use the box on page 10.

Please tick or write the number of times

6 In the last [X] months, how many times have you been to **hospital for an overnight stay**?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Department	Tests or surgical procedures performed (if applicable)	Reason for stay
Example	[4]	[orthopaedics]	[knee replacement]	[knee replacement]
Stay 1				
Stay 2				
Stay 3				
Stay 4				
Stay 5				

If you need more space, please use the box on page 10.

A5.3 continued

Questions 7 to 15 ask about healthcare outside of hospitals: GP (doctor), nurse and any other NHS healthcare professionals or services

		Please tick or write the number of times
7	In the last [X] months, how many times have you had an appointment with a doctor (GP) at a GP surgery, health centre or walk-in centre?	0 1 2 3 4 How many?
8	In the last [X] months, how many times have you had an appointment with a doctor (GP) at home?	0 1 2 3 4 How many?
9	In the last [X] months, how many times have you had an appointment with a doctor (GP) over the telephone or online?	0 1 2 3 4 How many?
10	In the last [X] months, how many times have you had an appointment/contact with [pre-specified healthcare professional or healthcare service] at a GP surgery, health centre or walk-in centre?	0 1 2 3 4 How many?
11	In the last [X] months, how many times have you had an appointment/contact with [pre-specified healthcare professional or healthcare service] at home?	0 1 2 3 4 How many?
12	In the last [X] months, how many times have you had an appointment/contact with [pre-specified healthcare professional or healthcare service] over the telephone or online?	0 1 2 3 4 How many?

Please tick or write the number of times

13	In the last [X] months, how many times have you had contact with any other NHS healthcare professionals (e.g. [nurse]) or NHS healthcare services at a GP surgery, health centre or walk-in centre?	0 1 2 3 4 How many?

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[community physiotherapist]	[3]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 10.

A5.3 continued

Please tick or write the number of times

14 In the last [X] months, how many times have you had contact with **any other NHS healthcare professionals** (e.g. [district nurse]) **or NHS healthcare services at home?**

0 1 2 3 4 How many? ☐ ☐ ☐ ☐ ☐ ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[NHS chiropodist]	[3]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 10.

Please tick or write the number of times

15 In the last [X] months, how many times have you had contact with **any other NHS healthcare professionals or NHS healthcare services** (e.g. [NHS 111 telephone call]) over the **telephone or online?**

0 1 2 3 4 How many? ☐ ☐ ☐ ☐ ☐ ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[NHS 111 telephone call]	[1]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 10.

A5.3 continued

Question 16 asks about prescribed medications

		Please tick yes	Please tick no
16	In the last [X] months, have you picked up or received any **prescribed medications?**	☐	☐

For each prescribed medication, please complete the table:

Prescribed medication	Name of prescribed medication If you don't know the name, please put the type of medication	Number of times the medication has been picked up and/or received
Example 1	[Tramadol / Painkiller]	[3]
Example 2	[Diprobase / Eczema cream]	[1]
Medication 1		
Medication 2		
Medication 3		
Medication 4		
Medication 5		
Medication 6		
Medication 7		
Medication 8		
Medication 9		
Medication 10		

If you need more space, please use the box on page 10.

If you run out of space in the tables, please use the box below to write about the other healthcare you have used:

Please check you have answered every question.

Thank you for completing the questionnaire.

A5.4. ModRUM User Guide (pre-health economist piloting)

ModRUM

Modular Resource-Use Measure

ModRUM User Guide

Instructions for using ModRUM

Quick start guide

To use ModRUM for collecting resource-use data you will need to:

(1) Assess whether the core module is adequate for your needs.

(2) If not, choose the depth version of ModRUM in its entirety, or edit the core module and replace core questions with chosen depth questions.

(3) If necessary, add questions for resources that are not included in ModRUM, that are important to capture in your trial.

(4) Edit adaptable aspects of ModRUM to suit your needs, including:

- (a) recall period
- (b) resource-use type
- (c) examples
- (d) table rows

(5) Update the wording and formatting (see Section 7 for tips) of ModRUM to ensure that:

- (a) each question is contained within one page, where feasible
- (b) question numbers are updated, if edits have been made
- (c) square brackets and highlighting are removed
- (d) section instructions and the extra space box has been added at the end, if ModRUM has been extended beyond the core module

1. Introduction

This document provides guidance on how to use ModRUM. ModRUM is a brief, generic, standardised resource-use questionnaire (RUQ) designed to collect patient-reported data. ModRUM was developed with the aim of alleviating some of the issues associated with resource-use measurement. Despite wide use of self-report RUQs, research efforts are often duplicated as new RUQs are designed on a trial-by-trial basis, with limited time for validity and reliability testing. A well-validated, standardised RUQ would reduce unnecessary work and make study results more transparent and directly comparable with one another.

ModRUM has a modular design, with a core module that includes a minimum set of healthcare resource items that could be collected in all trials. If needed, the core module can be expanded to ask patients for additional details by substituting 'core' questions for 'depth' questions. Resources in sectors beyond healthcare can be measured by adding additional 'breadth' modules (in development). Users may also adapt ModRUM to give examples of resources pertinent to their trial. ModRUM is designed to capture general healthcare use over the study follow-up period. Additional micro-costing methods may be needed to cost the intervention and comparator procedures/services.

While some flexibility is permitted (e.g. the recall period or the examples of resource-use items given) to ensure ModRUM is appropriate to each study, users are discouraged from editing parts of questions that are not specified as adaptable (Section 2). This will maximise the potential for ModRUM-based costs to be compared across studies.

A5.4 continued

2. Implementation of ModRUM

2.1. Core module

The core module includes 8 items that were identified in an expert Delphi consensus survey as potentially relevant to most trials, conditions and patient groups. The items include A&E, outpatients, day cases, inpatients, GP at the surgery/health centre or home and other healthcare professionals at the surgery/health centre. To maintain standardisation in the implementation of ModRUM, editing of the core questions, other than extending them to depth questions, is discouraged.

2.2. Depth questions

The core module captures enough information on important items to allow for cost estimation of healthcare utilisation. However, depth questions allow users to capture more detail for questions that are potentially key cost drivers or are highly utilised resources in their study, so that costs can be estimated with greater precision. Depth questions are available for outpatients, day cases, inpatients, GP at the surgery/health centre or home and other healthcare professionals at the surgery/health centre. Additional depth information includes free-test fields, such as clinic type, tests/procedures and reason for appointment. Depth questions also include questions on the ambulance service and prescribed medications. When using depth questions users may select how many rows are appropriate to include in tables for their patient group. Section instructions will also need to be revised to reflect the questions included.

2.3. Scope, perspective and items beyond ModRUM

The scope of the core module is healthcare and the questions are worded from a public payer (i.e. NHS) perspective. The wording could be adapted to take a different perspective (e.g. healthcare provided by public or private healthcare providers) or to make it internationally applicable (substitute NHS for equivalent international body). Users may also adapt ModRUM to add resources that are pertinent to their trial, that are not included in ModRUM.

2.4. Examples

Examples to aid patients understanding are highlighted in the questionnaire template in in square brackets. They have gone through patient testing; however, the user can assess whether they are relevant to their patient group and use different examples if appropriate.

2.5. Administration mode

In piloting, adults aged 18 and over self-completed paper versions of ModRUM. It may be appropriate for future adaptations of the ModRUM to be administered by an interviewer or completed by a proxy. The user can decide which administration mode is most suitable for their population. ModRUM is currently available in paper format; however, in future there are plans to develop an online version.

2.6. Recall period

While it is acknowledged that recall periods for RUQs are often driven by outcome measure collection points, ModRUM has undergone initial piloting with a 3-month recall period, which most patients found acceptable. Patients provided some indication that a 6-month recall period may be acceptable; however, recall periods beyond that would require further testing as they are potentially problematic for less salient items.

In testing, an 'anchor' point was not included; however, the user may use one as evidence suggests they can help with recall (e.g. Since your knee replacement).

2.7. All-cause or condition-specific resource-use

ModRUM has been piloted with patients who completed their all-cause (i.e. for any reason) healthcare use. If appropriate, the user can specify in the introduction that only resources related to a specific condition should be included.

A5.4 continued

3. Analysis of ModRUM

The design of ModRUM questions aims to make responses straightforward to cost using publicly available unit cost data. In future, we aim to develop a cost compendium and software code to semi-automate the costing of ModRUM responses.

4. Reporting use of ModRUM

To increase the transparency in how resource-use data are collected, in the write up of research, users should report which modules, depth questions and resources beyond ModRUM they have included. They should also report which parts they have adapted (e.g. recall period).

5. Registration and copyright of ModRUM

Use of ModRUM should be reported on the short, online registration form. *Further details will be included in the final version of this document. We intend to make the final version of ModRUM available for researchers using a registration system and Creative Commons License.*

6. Validation of ModRUM

Development and validation of ModRUM has incorporated:

- A Delphi consensus survey with health economists to identify items.
- Semi-structured interviews with experienced health economists to assess face and content validity.
- Patient 'think-aloud' interviews with retrospective probing to assess content validity and acceptability.

Research currently being undertaken to further validate ModRUM includes:

- A pilot with health economists to assess feasibility.

Future development and validation are needed for ModRUM, including:

- A pilot with patients to assess feasibility, acceptability, construct validity, criterion validity and reliability.
- Development and validation of breadth modules including social care and informal care module, educational support, vocational support and criminal justice.
- Development and validation of online, interviewer administered, and proxy versions.

7. Formatting tips

(a) How do I move a depth question into the core module, to replace a core question?

- If the depth question takes up 1 page, click the top of the page where you wish to add a new page. The curser should blink.
- In the task bar, go to "Insert" tab and click "Pages, then click "Blank page". Everything below the curser should move to the page below and a new page will be added.
- Copy the depth question and paste it on the new page. Delete the core question.

(b) I want to capture data on nurses, separate from other healthcare professionals. How do I add a question on nurses?

- Data on nurses and any other healthcare professionals seen or services used in the community, can be captured using prespecified questions in Questions 10 to 13 or using free-text questions in Questions 14 to 16.
- If you include 10 to 13, you can replace [pre-specified healthcare professional or healthcare service] with the name of the professional or service you wish to capture.
- To avoid double-counting, if you are also including questions 14 to 16, you will need to remove professionals or services named in 10 to 13 from examples in 14 to 16.
- Questions 10-13 can be repeated to collect to collect data on different healthcare professionals/services.

(c) How do I quickly change the recall periods in ModRUM?

- If you want to change the recall period to 3 months. In the task bar, go to "Home" and in "Editing" click "Replace".
- The "Find and replace" box will appear. In "Find what", write "[X]" and in "Replace with" write "3". Click "Replace all" and all recall periods should now state "3".

(d) I want to use the examples provided; how do I remove the square brackets?

- As above, use the "Find and replace" box. In "Find what", write "[" and leave "Replace with" blank. Click "Replace all". Repeat for "]" and all square brackets will be removed.

A5.5. ModRUM User Guide (post-health economist piloting)

ModRUM

Modular Resource-Use Measure

ModRUM User Guide

Instructions for using ModRUM

Quick start guide

To use ModRUM for collecting resource-use data you will need to:

(1) Assess whether the core module is adequate for your needs.

(2) If not, choose the depth version of ModRUM in its entirety, or edit the core/depth version and replace core/depth questions with chosen depth/core questions (see Section 8 for equivalent questions in each version).

(3) If necessary, add questions for resources that are not included in ModRUM, that are important to capture in your trial.

(4) Edit adaptable aspects of ModRUM to suit your needs, including:
- (a) recall period
- (b) resource-use type
- (c) examples
- (d) table rows

(5) Update the wording and formatting (see Section 7 for tips) of ModRUM to ensure that:
- (a) each question is contained within one page, where feasible
- (b) question numbers are updated, if edits have been made
- (c) square brackets are removed
- (d) section instructions and the extra space box has been added at the end, if ModRUM has been extended beyond the core module

1. Introduction

This document provides guidance on how to use ModRUM. ModRUM is a brief, generic, standardised resource-use measure (RUM) designed to collect patient-reported data. ModRUM was developed with the aim of alleviating some of the issues associated with resource-use measurement. Despite wide use of self-report RUQs, research efforts are often duplicated as new RUQs are designed on a trial-by-trial basis, with limited time for validity and reliability testing. A well-validated, standardised RUQ would reduce unnecessary work and make study results more transparent and directly comparable with one another.

ModRUM has a modular design, with a core module that includes a minimum set of healthcare resource items that could be collected in all trials. If needed, the core module can be expanded to ask patients for additional details by substituting 'core' questions for 'depth' questions. Resources in sectors beyond healthcare can be measured by adding additional 'breadth' modules (in development). Users may also adapt ModRUM to give examples of resources pertinent to their trial. ModRUM is designed to capture general healthcare use over the study follow-up period. Additional micro-costing methods may be needed to cost the intervention and comparator procedures/services.

While some flexibility is permitted (e.g. the recall period or the examples of resource-use items given) to ensure ModRUM is appropriate to each study, users are discouraged from editing parts of questions that are not specified as adaptable (see Section 2 for permitted adaptations). This will maximise the potential for ModRUM-based costs to be compared across studies.

A5.5 continued

2. Implementation of ModRUM

2.1. Core module

The core module includes 11 items that are potentially relevant to most trials, conditions and patient groups. The items cover A&E, outpatients, day cases, inpatients, GP and other healthcare professionals. The items can be seen in Section 8. To maintain standardisation in the implementation of ModRUM, users are not permitted to edit or omit core questions, other than extending them to depth questions and making adaptations that are listed as permitted in this document.

2.2. Depth questions

The core module captures enough information on important items to allow for cost estimation of healthcare utilisation. However, depth questions allow users to capture more detail for questions that are potentially key cost drivers or are highly utilised resources in their study, so that costs can be estimated with greater precision. Depth questions are available for outpatients, day cases, inpatients and other healthcare professionals at the surgery/health centre. Additional depth information includes free-text fields, such as clinic type, tests/procedures and reason for appointment. Depth questions also include questions on the ambulance service and prescribed medications. When using depth questions users may select how many rows are appropriate to include in tables for their patient group. If adaptations are made, users also need to revise section instructions to reflect the questions included. When depth questions are included the extra space box should be included at the end of the questions.

2.3. Scope, perspective and items beyond ModRUM

The scope of the core module is healthcare and the questions are worded from a public payer (i.e. NHS) perspective. The wording could be adapted to take a different perspective (e.g. healthcare provided by public or private healthcare providers) or to make it internationally applicable (substitute NHS for equivalent international body). Users may also adapt ModRUM to add resources that are pertinent to their trial, that are not included in ModRUM.

2.4. Examples

Examples to aid patients understanding are highlighted in the questionnaire template in in square brackets. They have gone through patient testing; however, the user can assess whether they are relevant to their patient group and use different examples if appropriate.

2.5. Administration mode

In piloting, adults aged 18 and over self-completed paper versions of ModRUM. It may be appropriate for future adaptations of the ModRUM to be administered by an interviewer or completed by a proxy. The user can decide which administration mode is most suitable for their population. ModRUM is currently available in paper format; however, in future there are plans to develop an online version.

2.6. Recall period

While it is acknowledged that recall periods for RUQs are often driven by outcome measure collection points, ModRUM has undergone initial piloting with a 3-month recall period, which most patients found acceptable. Patients provided some indication that a 6-month recall period may be acceptable; however, recall periods beyond that would require further testing as they are potentially problematic for less salient items.

In testing, an 'anchor' point was not included; however, the user may use one as evidence suggests they can help with recall (e.g. Since your knee replacement). To minimise respondent burden, the same recall period should be used for all questions.

2.7. All-cause or condition-specific resource-use

ModRUM has been piloted with patients who completed their all-cause (i.e. for any reason) healthcare use. If appropriate, the user can specify in the introduction that only resources related to a specific condition should be included. The condition can also be repeated in each question (e.g. 'In the last 3 months, because of your knee pain, how many times…'.

A5.5 continued

3. Analysis of ModRUM

The design of ModRUM questions aims to make responses straightforward to cost using publicly available unit cost data. We recommend the following unit cost sources: the National Schedule of NHS Costs (emergency, secondary and community care), the Unit Costs of Health and Social Care (primary and community care), and the Prescription Cost Analysis or British National Formulary (prescribed medications).

4. Reporting use of ModRUM

To increase the transparency in how resource-use data are collected, in the write up of research, users should report which modules, depth questions and resources beyond ModRUM they have included. They should also report which parts they have adapted (e.g. recall period).

5. Registration and copyright of ModRUM

To obtain and use ModRUM, prospective users should register using the short, registration form, which can be found here: https://www.bristol.ac.uk/population-health-sciences/centres/healthecon/[to_be_confirmed]. Following registration, ModRUM is free to use for non-commercial use.

6. Development of ModRUM

Development and validation of ModRUM has incorporated:

- A Delphi consensus survey with health economists to identify items (https://doi.org/10.1016/j.jval.2017.06.011).
- Interviews with health economists to assess face and content validity.
- Interviews with patients to assess content validity and acceptability (https://doi.org/10.1186/s12913-021-06364-w).
- An online survey of health economists to assess feasibility and suitability.
- A pilot study with patients to assess feasibility, acceptability, construct validity and criterion validity.

Ongoing work includes:

- Development and validation of social care and informal care modules.

7. Formatting tips

(a) How do I move a depth question into the core module, to replace a core question?

- If the depth question takes up 1 page, click the top of the page where you wish to add a new page. The curser should blink.
- In the task bar, go to "Insert" tab and click "Pages, then click "Blank page". Everything below the curser should move to the page below and a new page will be added.
- Copy the depth question and paste it on the new page. Delete the core question.

(b) I want to capture data on nurses, separate from other healthcare professionals. How do I add a question on nurses?

- Data on nurses and any other healthcare professionals seen or services used in the community, can be captured using prespecified questions in depth questions 11 to 13 or using free-text questions in depth questions 14 to 16.
- If you include 11 to 13, you can replace [pre-specified healthcare professional or healthcare service] with the name of the professional or service you wish to capture.
- To avoid double-counting, if you are also including questions 14 to 16, you will need to remove professionals or services named in 11 to 13 from examples in 14 to 16.
- Questions 11 to 13 can be repeated to collect to collect data on different healthcare professionals/services.

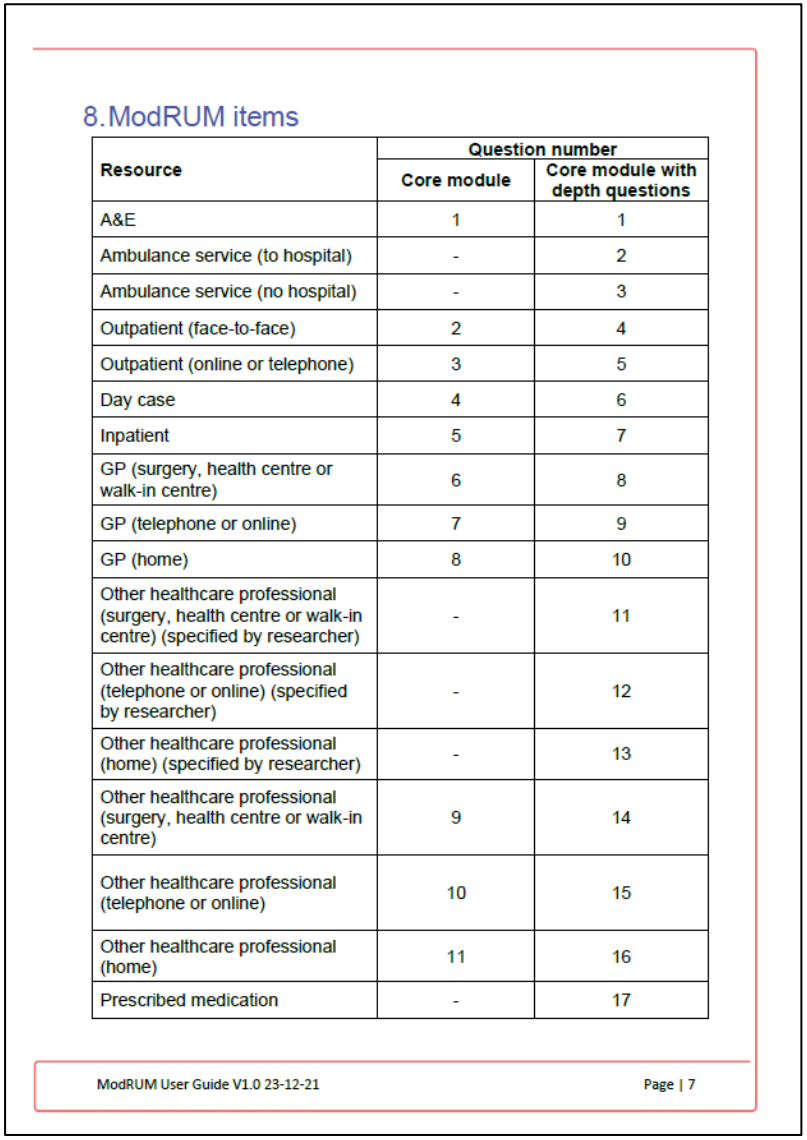
(c) How do I quickly change the recall periods in ModRUM?

- If you want to change the recall period to 3 months. In the task bar, go to "Home" and in "Editing" click "Replace".
- The "Find and replace" box will appear. In "Find what", write "[X]" and in "Replace with" write "3". Click "Replace all" and all recall periods should now state "3".

(d) I want to use the examples provided; how do I remove the square brackets?

- As above, use the "Find and replace" box. In "Find what", write "[" and leave "Replace with" blank. Click "Replace all". Repeat for "]" and all square brackets will be removed.

A5.5 continued



8. ModRUM items

Resource	Question number	
	Core module	Core module with depth questions
A&E	1	1
Ambulance service (to hospital)	-	2
Ambulance service (no hospital)	-	3
Outpatient (face-to-face)	2	4
Outpatient (online or telephone)	3	5
Day case	4	6
Inpatient	5	7
GP (surgery, health centre or walk-in centre)	6	8
GP (telephone or online)	7	9
GP (home)	8	10
Other healthcare professional (surgery, health centre or walk-in centre) (specified by researcher)	-	11
Other healthcare professional (telephone or online) (specified by researcher)	-	12
Other healthcare professional (home) (specified by researcher)	-	13
Other healthcare professional (surgery, health centre or walk-in centre)	9	14
Other healthcare professional (telephone or online)	10	15
Other healthcare professional (home)	11	16
Prescribed medication	-	17

Appendix 6. Piloting with patients

This appendix contains documents related to piloting with patients (**Chapter 8**).

A6.1. Cover letter for pilot with patients

<<Custom 3>>
<<Custom 4>>
<<Custom 5>>
<<Custom 6>>
<<Custom 7>>
<<Custom 8>>

<<Title>> <<Firstname>> <<Surname>>
<<Address 1>>
<<Address 2>>
<<Address 3>>
<<Address 4>>
<<Address 5>>

<<date4>>

Dear <<Title>> <<Surname>>

Our surgery is supporting a research study to develop a questionnaire that can be used in studies to ask study participants about the healthcare resources, such as GP appointments, they have used. This is independent research being carried out by researchers at the University of Bristol. I am writing on their behalf to invite you to take part.

Details on healthcare resource use are collected in research studies so that the cost to the NHS of services patients use can be estimated. As part of the questionnaire development, the researchers want to understand how best to ask questions on healthcare use, so patients can answer the questions as accurately as possible.

I am writing to you because the researchers are looking for adults to take part from our GP practice. Please read the enclosed information sheet to learn more about the study and what taking part would involve. You are under no obligation to take part in this study.

If you would like to take part, please complete the consent form, health questionnaire, use of healthcare services questionnaire, patient characteristics form and contact details form (optional) enclosed in this letter and return them using the pre-paid envelope.

If you do not want to take part, then you do not need to do anything else.

If you have any questions about the research, please contact the lead researcher, Kirsty Garfield, at the University of Bristol by email (kirsty.garfield@bristol.ac.uk).

Thank you for giving the invitation to take part some thought.

Yours sincerely,

<<Custom 2>>

IRAS ID: 241489 GP patient invite letter, Version 1.0 23-09-20

A6.2. ModRUM core module with depth questions - patient pilot version

ModRUM

Modular Resource-Use Measure

Participant Identification Number:

<<Custom 1>>

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used **in the last 3 months.** We only want you to **include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

Questions 1 to 3 ask about emergency healthcare: A&E and ambulance

		Please tick or write the number of times
1	In the last 3 months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
2	In the last 3 months, how many times have you received care from the **ambulance service and were taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
3	In the last 3 months, how many times have you received care from the **ambulance service but were not taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

Questions 4 to 7 ask about healthcare at hospitals: outpatients (at the hospital), outpatients (over the telephone or online), day case and overnight stays

		Please tick or write the number of times
4	In the last 3 months, how many times have you been to **hospital for an outpatient appointment** (e.g. to see a consultant or hospital physiotherapist, or to have an x-ray)?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type	Tests or surgical procedures performed (if applicable)	Reason for appointment
Example 1	radiology	x-ray	x-ray for pain after a fall
Example 2	dermatology	removal of a skin lesion	mole removal
Appointment 1			
Appointment 2			
Appointment 3			
Appointment 4			
Appointment 5			
Appointment 6			

If you need more space, please use the box on page 11.

A6.2 continued

Please tick or write the number of times

5 In the last 3 months, how many times have you had an **online or telephone hospital outpatient appointment** (e.g. with a consultant)?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type	Reason for appointment
Example 1	physiotherapy	update treatment plan and exercises
Example 2	rheumatology	routine arthritis check-up
Appointment 1		
Appointment 2		
Appointment 3		
Appointment 4		
Appointment 5		
Appointment 6		

If you need more space, please use the box on page 11.

Please tick or write the number of times

6 In the last 3 months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Hospital day case	Department	Tests or surgical procedures performed (if applicable)	Reason for day case
Example	gastroenterology	endoscopy with biopsy	stomach ache
Day case 1			
Day case 2			
Day case 3			
Day case 4			
Day case 5			

If you need more space, please use the box on page 11.

A6.2 continued

		Please tick or write the number of times					
7	In the last 3 months, how many times have you been to **hospital for an overnight stay?**	0	1	2	3	4	How many?

For each appointment, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Department	Tests or surgical procedures performed (if applicable)	Reason for stay
Example	4	orthopaedics	knee replacement	knee replacement
Stay 1				
Stay 2				
Stay 3				
Stay 4				
Stay 5				

If you need more space, please use the box on page 11.

Questions 8 to 16 ask about healthcare outside of hospitals: GP (doctor), nurse and any other NHS healthcare professionals or services

		Please tick or write the number of times					
8	In the last 3 months, how many times have you had an appointment with a **doctor (GP) at a GP surgery, health centre or walk-in centre?**	0	1	2	3	4	How many?
9	In the last 3 months, how many times have you had an appointment with a **doctor (GP) over the telephone or online ?**	0	1	2	3	4	How many?
10	In the last 3 months, how many times have you had an appointment with a **doctor (GP) at home?**	0	1	2	3	4	How many?
11	In the last 3 months, how many times have you had an appointment/contact with **a nurse at a GP surgery, health centre or walk-in centre?**	0	1	2	3	4	How many?
12	In the last 3 months, how many times have you had an appointment/contact with **a nurse over the telephone or online?**	0	1	2	3	4	How many?
13	In the last 3 months, how many times have you had an appointment/contact with **a nurse at home?**	0	1	2	3	4	How many?

A6.2 continued

Please tick or write the number of times

14 In the last 3 months, how many times have you had contact with **any other NHS healthcare professionals** (e.g. community physiotherapist) or **NHS healthcare services** at a GP surgery, health centre or walk-in centre?

0 1 2 3 4 How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	community physiotherapist	3
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

Please tick or write the number of times

15 In the last 3 months, how many times have you had contact with **any other NHS healthcare professionals or NHS healthcare services** (e.g. NHS 111 telephone call) over the **telephone or online?**

0 1 2 3 4 How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	NHS 111 telephone call	1
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

A6.2 continued

Please tick or write the number of times

16 In the last 3 months, how many times have you had contact with **any other NHS healthcare professionals** (e.g. health visitor) or **NHS healthcare services** at home?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	NHS chiropodist	3
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

Question 17 asks about prescribed medications

Please tick

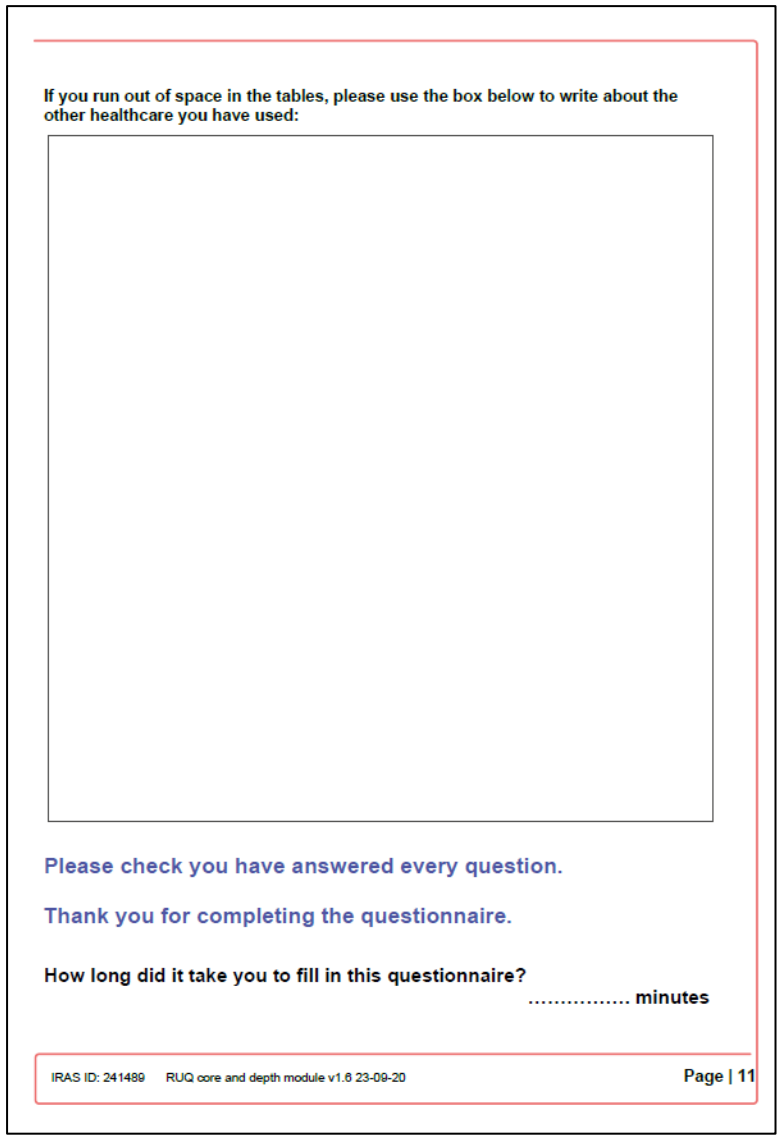
17 In the last 3 months, have you picked up or received any **prescribed medications**? yes ☐ no ☐

For each prescribed medication, please complete the table:

Prescribed medication	Name of prescribed medication If you don't know the name, please put the type of medication	Number of times the medication has been picked up and/or received
Example 1	Tramadol / Painkiller	3
Example 2	Diprobase / Eczema cream	1
Medication 1		
Medication 2		
Medication 3		
Medication 4		
Medication 5		
Medication 6		
Medication 7		
Medication 8		
Medication 9		
Medication 10		

If you need more space, please use the box on page 11.

A6.2 continued



If you run out of space in the tables, please use the box below to write about the other healthcare you have used:

Please check you have answered every question.

Thank you for completing the questionnaire.

How long did it take you to fill in this questionnaire?

..... minutes

A6.3. Patient information sheet – patient pilot

University of BRISTOL

Population Health Sciences
Bristol Medical School
University of Bristol
1-5 Whiteladies Road
Bristol
BS8 1NU

PATIENT INFORMATION SHEET
Developing a modular resource-use questionnaire for use in randomised controlled trials

A team of researchers at the University of Bristol would like to invite you to take part in a research study. The aim of this research is to develop a questionnaire that can be used in studies to ask study participants about the healthcare resources, such as GP appointments, they have used.

For this study we would like you to fill in a questionnaire about your healthcare use. We would then like to compare your responses to your GP medical records.

Before you decide whether to take part, it is important for you to understand why the research is being done and what taking part will involve for you. This leaflet explains the purpose of the research and what we would like you to do.

1. What is the purpose of this study?
The purpose of this study is to develop a questionnaire that can be completed by study participants to report the healthcare resources (e.g. hospital outpatient appointments) they use. These details are collected in research studies so that we can estimate the cost to the NHS of the services or interventions patients use. As part of the development we want to understand how best to ask questions on healthcare use, so the answers we get are as accurate as possible.

2. Why have I been invited to take part?
You have been invited because your GP practice is supporting this study. Around 400 patients will take part in this study.

3. Do I have to take part?
No, it is up to you to decide whether to take part. If you want to take part, please complete and sign the enclosed consent form. You can stop taking part in the study at any time without giving a reason. Deciding not to take part or withdrawing will not affect the care you receive.

4. What will happen if I take part?
If you decide to take part, I would like you to fill in some questions on your healthcare use, your quality-of-life and your characteristics (e.g. age group). They will take around 15 minutes to complete. Once you have completed the forms, please return the consent form, health questionnaire, use of healthcare services questionnaire, patient characteristics form and contact details form (optional) in the enclosed pre-paid envelope. With your permission I will obtain information on your healthcare use from your GP medical records. These details will be compared to your answers.

A6.3 continued

5. Will I receive any payment for taking part?
No, we are not able to offer any expenses or payments to patients who participate in the study. However, as a thank you for taking part you can enter a competition to win a £100 love2shop voucher which can be used in many shops and restaurants on the high street. If you would like to enter the competition, please provide your postal address on the contact details form.

6. What are any possible benefits of taking part?
Your views as a patient will provide important research input. You will help us to find out if the questions we ask are easy to understand and the answers we get are accurate. We can't promise the study will help you, but the information you give us will improve the design of the final questions which may be used in future research. The questionnaire will help to provide accurate information on whether NHS services are value for money.

7. Will my taking part in this study be kept confidential?
Yes, your data will be analysed and stored electronically on a secure computer network at the University of Bristol. Any hard copies of this information will be locked in a secure filing cabinet at the research base (University of Bristol). Your personal data will be stored separately from research findings and will only be accessed by the research team. Any information that could identify you will be removed from the data before the findings are seen by others, and personal data will not be used in research reports. The handling, processing, storage, and destruction of these data are compliant with the Data Protection Act 2018.

8. How will my personal data be used and what are my rights?
The University of Bristol is the sponsor for this study based in the United Kingdom. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The University of Bristol will keep identifiable information about you for 10 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information by contacting the lead researcher (Kirsty Garfield).

The University of Bristol will use your name and contact details to contact you about the research study and to oversee the quality of the study. Individuals from the University of Bristol and regulatory organisations may look at your medical and research records to check the accuracy of the research study. Your GP practice will pass these details to the University of Bristol along with the information collected from you and your medical records. The only people in the University of Bristol who will have access to information that identifies you will be people who need to contact you to discuss the research study or audit the data collection process.

The University of Bristol will collect information about you for this research study from your GP practice. This information will include health information, which is regarded as a special category of information. We will use this information to compare your questionnaire responses to the healthcare use reported in your GP medical records.

9. What will happen to the results of this research study?
The results will be written up for publication and may be presented at conferences. You will not be identified in any publication or presentation.

If you would like a summary of the results, you can indicate this on the consent form and provide your email or postal address on the contact details form and I will email or post a leaflet to your home.

10. Who is organising and funding this study?
This study is being carried out by researchers at the University of Bristol who are members of the MRC ConDuCT-II Hub for Trials Methodology Research. This work is supported by the MRC Network of Hubs for Trials Methodology Research (MR/L004933/2- PhD Award R9).

University of BRISTOL

MRC Hub for Trials Methodology Research

MRC Hubs for Trials Methodology Research

Research team
The research team includes:
- Miss Kirsty Garfield, the lead researcher, PhD student and Research Associate in Health Economics at the University of Bristol.
- Dr Joanna Thom, Senior Research Associate in Health Economics at the University of Bristol.
- Dr Sian Noble, Senior Lecturer in Health Economics at the University of Bristol.
- Dr Samantha Husbands, Senior Research Associate in Qualitative Research & Health Economics at the University of Bristol.
- Professor Will Hollingworth, Professor of Health Economics at the University of Bristol.

11. Who has reviewed this study?
All research in the NHS has been reviewed and given a favourable opinion by the South Central - Berkshire B Research Ethics Committee to protect your safety, rights, wellbeing and dignity.

Contact details for further information
For further information about the research, please contact the lead researcher:

Kirsty Garfield
1-5 Whiteladies Road
Bristol
BS8 1NU
Email: Kirsty.Garfield@bristol.ac.uk

For complaints about the study please contact:

Will Hollingworth
Email: William.Hollingworth@bristol.ac.uk

Thank you for taking the time to read this Participant Information Sheet and for considering whether to take part in this study.

If you decide to take part, please keep this information sheet and one signed copy of the consent form.

A6.4. Patient consent form – patient pilot

Participant Identification Number: <<Custom 1>>

University of BRISTOL

CONSENT FORM

Project title: Developing a modular resource-use questionnaire for use in randomised controlled trials

Researcher: Kirsty Garfield

Please initial box

1. I confirm that I have read and understood the information sheet dated 23/09/20 (version 1.2) for the above study. I have had the opportunity to consider the information, contact the researcher to ask questions if needed and have these questions answered. ☐
2. I understand that participation is voluntary and that I am free to withdraw at any time without giving a reason, and without my medical care or legal rights being affected. ☐
3. I understand that relevant sections of my medical notes and data collected during the study may be looked at by the study research team at the University of Bristol or by individuals from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records. ☐
4. I understand that you will write a report about the research findings, but that it will not be possible to identify me from the report. ☐
5. I understand that the information collected about me will be used to support other research in the future, and may be shared anonymously with other researchers. ☐
6. I agree to take part in the above study. ☐

Optional:

7. I would like to receive a summary of the results once the project is complete. ☐
8. I would like to enter the competition to win a £100 Love2shop voucher. ☐

Name of participant _____ Date _____ Signature _____

Please return one completed copy to the research team and keep one completed copy for yourself.

IRAS ID: 241489 Study 2 consent form Version 1.2 23-09-20

A6.5. Patient characteristics form – patient pilot

Participant Identification Number: <<Custom 1>>

University of BRISTOL

Development of a healthcare use questionnaire

Patient Characteristics Form

We would like to ask you for some information on your sex, age group, ethnic group, long-term conditions and age on leaving full time education. The reason we are asking for these details is to find out how different patients answer the questionnaire. This will help us to make sure the questionnaire is suitable for a broad range of patients.

1. Sex	Female	☐	
	Male	☐	
	Prefer not to say	☐	
2. Age group	18 – 30	☐	
	31 – 45	☐	
	46 – 55	☐	
	56 – 65	☐	
	66 – 75	☐	
	76 or older	☐	
	Prefer not to say	☐	
3. Ethnic group	White	☐	
	Mixed/Multiple ethnic groups	☐	
	Asian/Asian British	☐	
	Black/African/Caribbean/ Black British	☐	
	Other	☐	If other, please specify:
	Prefer not to say	☐	
4. Long-term conditions, e.g. diabetes, asthma, heart disease	None	☐	
	One	☐	
	More than one	☐	
	Prefer not to say	☐	
5. Age on leaving full time education	16 or under	☐	
	17 or 18	☐	
	19 or over	☐	
	Prefer not to say	☐	

IRAS ID: 241489 Study 2 patient characteristics form Version 1.1 23-09-20

A6.6. Unit costs for GP practice electronic medical record data

Healthcare resource	Unit cost (£)	Source detail
General practitioner		
GP surgery	33.19	9.22 minute consultation (including direct care costs, excluding qualification costs) (66).
Online/telephone	25.56	7.1 minute telephone call, 3.60 per minute of patient contact (66, 200).
Practice nurse		
GP surgery	16.28	Assume Band 5 nurse: 63 per hour patient related work, 15.5 minutes (66, 201).
Online/telephone	6.30	Assume Band 5 nurse: same duration as specialist nurse, 6 minutes (66, 201).
Specialist Nurse practitioner		
GP surgery	30.00	Assume Band 7 nurse: 120 per hour patient related work, 15 minutes (66, 201).
Online/telephone	12.00	Assume Band 7 nurse: 6 minutes (66, 201).
Healthcare assistant		
GP surgery	8.53	Healthcare assistant -18,688 per annum plus oncosts (additional 0.29 same as a Band 4 nurse), assume same ratio (salary to hours) and duration as practice nurse: 33 per hour, 15.5 minutes (66, 201).
Online/telephone	3.30	Healthcare assistant: same duration as specialist nurse, 6 minutes (2015)
Paramedic Specialist Practitioner		
GP surgery	26.13	Assume Band 6/7: 104.50 per hour patient related work, assume same duration as specialist nurse, 15 minutes (66, 201).
Online/telephone	10.45	Assume Band 6/7: same duration as specialist nurse, 6 minutes (66, 201).
Pharmacist		
Online/telephone	8.90	Assume Band 6: 89 per hour patient related work, assume same duration as specialist nurse, 6 minutes (66, 201).

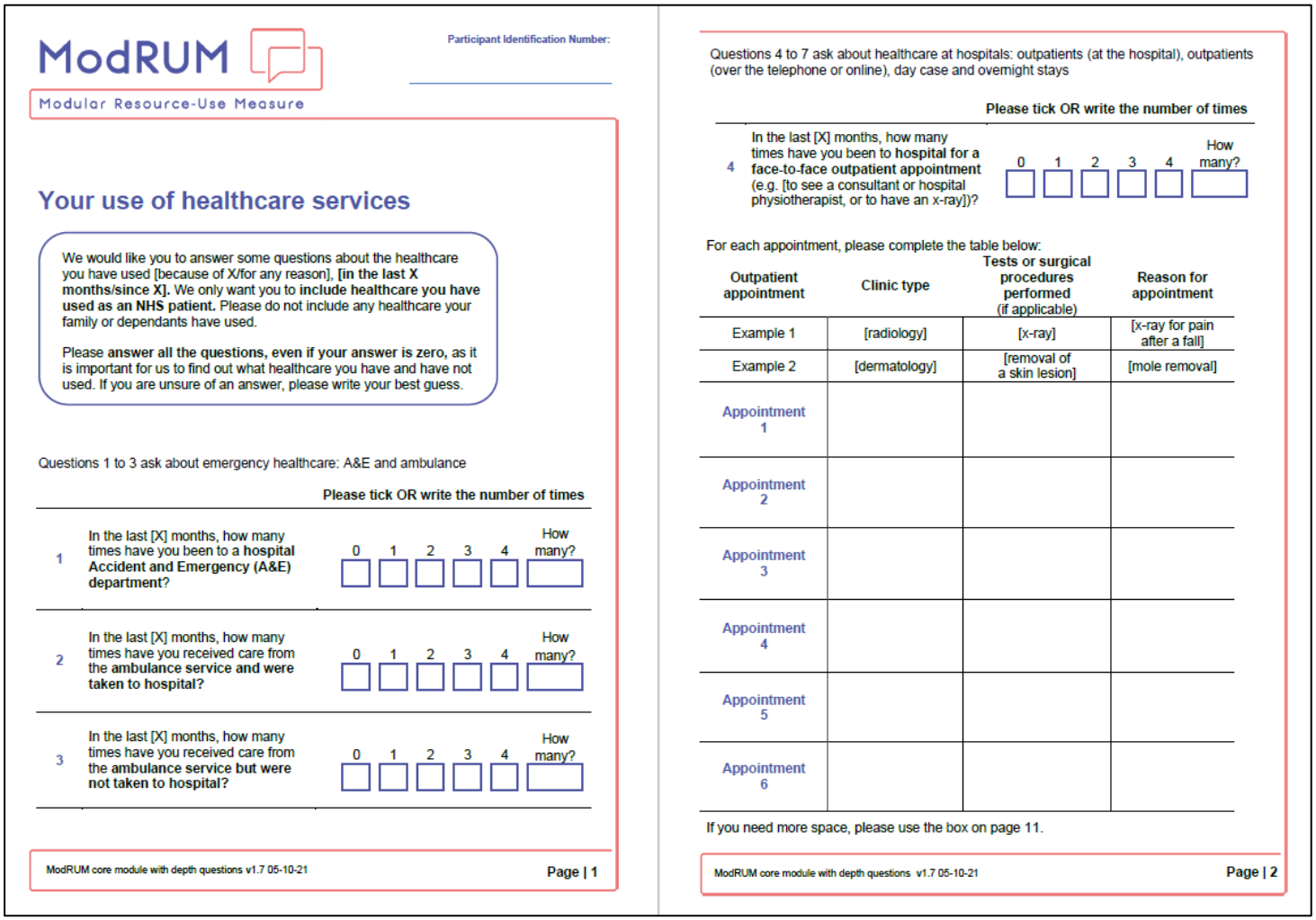
A6.6 continued

Healthcare resource	Unit cost (£)	Source detail
Phlebotomist		
GP surgery	3.71	DAPS tab: phlebotomy (48).
Physiotherapist		
GP surgery	62.90	AHP tab: Physiotherapist, Adult, One to One (48).
Online/telephone	62.90	AHP tab: Physiotherapist, Adult, One to One (48).
Prescribed medications	Varies	Name, dose and quantity costed using cost per quantity (203).

Appendix 7. Discussion and conclusion

This appendix contains documents related to the discussion and conclusion chapter (**Chapter 9**).

A7.1. ModRUM core module with depth questions (final version)



ModRUM

Modular Resource-Use Measure

Participant Identification Number:

Your use of healthcare services

We would like you to answer some questions about the healthcare you have used [because of X/for any reason], **[in the last X months/since X].** We only want you to **include healthcare you have used as an NHS patient.** Please do not include any healthcare your family or dependants have used.

Please **answer all the questions, even if your answer is zero,** as it is important for us to find out what healthcare you have and have not used. If you are unsure of an answer, please write your best guess.

Questions 1 to 3 ask about emergency healthcare: A&E and ambulance

		Please tick OR write the number of times
1	In the last [X] months, how many times have you been to a **hospital Accident and Emergency (A&E) department?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
2	In the last [X] months, how many times have you received care from the **ambulance service and were taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐
3	In the last [X] months, how many times have you received care from the **ambulance service but were not taken to hospital?**	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

ModRUM core module with depth questions v1.7 05-10-21 Page | 1

Questions 4 to 7 ask about healthcare at hospitals: outpatients (at the hospital), outpatients (over the telephone or online), day case and overnight stays

		Please tick OR write the number of times
4	In the last [X] months, how many times have you been to **hospital for a face-to-face outpatient appointment** (e.g. [to see a consultant or hospital physiotherapist, or to have an x-ray])?	0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type	Tests or surgical procedures performed (if applicable)	Reason for appointment
Example 1	[radiology]	[x-ray]	[x-ray for pain after a fall]
Example 2	[dermatology]	[removal of a skin lesion]	[mole removal]
Appointment 1			
Appointment 2			
Appointment 3			
Appointment 4			
Appointment 5			
Appointment 6			

If you need more space, please use the box on page 11.

ModRUM core module with depth questions v1.7 05-10-21 Page | 2

A7.1 continued

Please tick OR write the number of times

5 In the last [X] months, how many times have you had an **online or telephone hospital outpatient appointment** (e.g. with a consultant)? 0 1 2 3 4 How many? ☐

For each appointment, please complete the table below:

Outpatient appointment	Clinic type	Reason for appointment
Example 1	[physiotherapy]	[update treatment plan and exercises]
Example 2	[rheumatology]	[routine arthritis check-up]
Appointment 1		
Appointment 2		
Appointment 3		
Appointment 4		
Appointment 5		
Appointment 6		

If you need more space, please use the box on page 11.

Please tick OR write the number of times

6 In the last [X] months, how many times have you been to **hospital for a day case** (used a bed, but did not stay overnight)? 0 1 2 3 4 How many? ☐

For each appointment, please complete the table below:

Hospital day case	Department	Tests or surgical procedures performed (if applicable)	Reason for day case
Example	[gastroenterology]	[endoscopy with biopsy]	[stomach ache]
Day case 1			
Day case 2			
Day case 3			
Day case 4			
Day case 5			

If you need more space, please use the box on page 11.

A7.1 continued

Please tick OR write the number of times

		0	1	2	3	4	How many?
7	In the last [X] months, how many times have you been to **hospital for an overnight stay?**	☐	☐	☐	☐	☐	☐

For each appointment, please complete the table below:

Hospital overnight stay	Number of nights spent in hospital	Department	Tests or surgical procedures performed (if applicable)	Reason for stay
Example	[4]	[orthopaedics]	[knee replacement]	[knee replacement]
Stay 1				
Stay 2				
Stay 3				
Stay 4				
Stay 5				

If you need more space, please use the box on page 11.

Questions 8 to 13 ask about GPs (doctors) and nurses, seen outside of hospitals. Questions 14 to 16 ask about any other NHS healthcare professionals or services seen outside of hospitals.

Please tick OR write the number of times

		0	1	2	3	4	How many?
8	In the last [X] months, how many times have you had an appointment with a **doctor (GP) at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
9	In the last [X] months, how many times have you had an appointment with a **doctor (GP) over the telephone or online ?**	☐	☐	☐	☐	☐	☐
10	In the last [X] months, how many times have you had an appointment with a **doctor (GP) at home?**	☐	☐	☐	☐	☐	☐
11	In the last [X] months, how many times have you had an [appointment/contact with] **[pre-specified healthcare professional or healthcare service] at a GP surgery, health centre or walk-in centre?**	☐	☐	☐	☐	☐	☐
12	In the last [X] months, how many times have you had an [appointment/contact with] **[pre-specified healthcare professional or healthcare service] over the telephone or online?**	☐	☐	☐	☐	☐	☐
13	In the last [X] months, how many times have you had an [appointment/contact with] **[pre-specified healthcare professional or healthcare service] at home?**	☐	☐	☐	☐	☐	☐

A7.1 continued

Please tick OR write the number of times

14 In the last [X] months, how many times have you had contact with **any other NHS healthcare professionals** (e.g. [community physiotherapist]) or **NHS healthcare services** at a GP surgery, health centre or walk-in centre?

0 1 2 3 4 How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[community physiotherapist]	[3]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

Please tick OR write the number of times

15 In the last [X] months, how many times have you had contact with **any other NHS healthcare professionals or NHS healthcare services** (e.g. [NHS 111 telephone call]) over the **telephone or online**?

0 1 2 3 4 How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[NHS 111 telephone call]	[1]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

A7.1 continued

Please tick OR write the number of times

16 In the last [X] months, how many times have you had contact with **any other NHS healthcare professionals** (e.g. [health visitor]) **or NHS healthcare services** at home?

0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ How many? ☐

For each healthcare professional seen and/or healthcare service you have used, please complete the table below:

Healthcare professional or service	Healthcare professional seen or healthcare service used	Number of times
Example	[NHS chiropodist]	[3]
Healthcare professional or service 1		
Healthcare professional or service 2		
Healthcare professional or service 3		
Healthcare professional or service 4		
Healthcare professional or service 5		

If you need more space, please use the box on page 11.

Question 17 asks about prescribed medications

Please tick

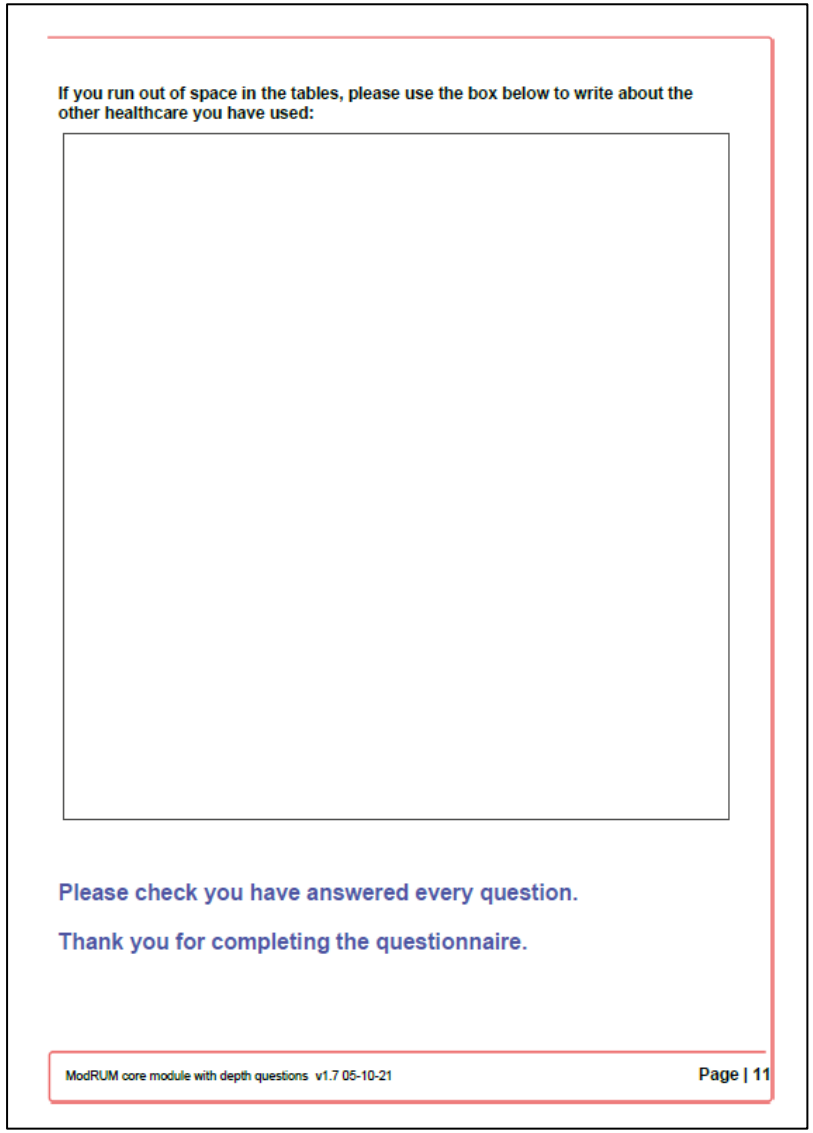
17 In the last [X] months, have you picked up or received any **prescribed medications**? yes ☐ no ☐

For each prescribed medication, please complete the table:

Prescribed medication	Name of prescribed medication If you don't know the name, please put the type of medication	Number of times the medication has been picked up and/or received
Example 1	[Tramadol / Painkiller]	[3]
Example 2	[Diprobase / Eczema cream]	[1]
Medication 1		
Medication 2		
Medication 3		
Medication 4		
Medication 5		
Medication 6		
Medication 7		
Medication 8		
Medication 9		
Medication 10		

If you need more space, please use the box on page 11.

A7.1 continued



If you run out of space in the tables, please use the box below to write about the other healthcare you have used:

Please check you have answered every question.

Thank you for completing the questionnaire.