

A QUALITATIVE ANALYSIS OF THE MODEL-BUILDING PROCESS

by

SAMANTHA KATHERYN HUSBANDS

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

Health Economics Unit

College of Medical and Dental Sciences

University of Birmingham

February 2016

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Decision-analytic models have an important role in healthcare funding decisions in the UK and internationally. However, errors have been reported in published models, which may indicate poor modelling practices, potentially leading to sub-optimal recommendations on cost-effectiveness. Little in-depth research has been undertaken to investigate the processes used by modellers in model development. The objective of this research was to explore the modelling methods used by modellers, with particular focus on problems encountered.

This work involved two qualitative phases of research. In the first phase, twenty-four in-depth interviews with modellers were undertaken. Constant comparative analysis was used to compare informant practices, and identify common issues in model development. The second phase involved two separate model-building case studies with teams of modellers and clinicians. Methods of non-participant observation, qualitative interviews, and think-aloud were used to investigate model development. The findings of the case studies were compared using framework analysis.

Important themes emerging from both phases of the research concerned the diversity of practices in structural development, problems with clinician involvement in modelling, and a lack of time and resources to carry out good practice methods. This work offers important recommendations for modelling practice, and suggestions for future research to improve modelling methods.

For Lillian and Eric Husbands.

Acknowledgements

First, I would like to thank my three supervisors, Joanna Coast, Susan Jowett and Pelham Barton. Thank you for your guidance, support and encouragement throughout this process, it has been invaluable and I have learnt so much from you.

This research was funded through a College Studentship at the University of Birmingham. Thank you to the funders for making this thesis possible, and for giving me the opportunity to conduct research with modellers across the UK, and to attend conferences to disseminate the findings. I would also like to express my gratitude to the Universitas 21 Doctoral Scholarship Programme, which provided funding for the research that took place in Canada.

A sincere thank you to all of the informants who gave up their valuable time to take part in this research, it could not have been possible without you. Thank you to those who supervised and assisted me during my time at the University of British Columbia, particularly Stirling Bryan and Debora Jewitt. Thank you also to everyone working in the Health Economics Unit for your support and guidance whilst I was writing up this thesis.

Aspects of this work have been presented at various seminars and conferences. These include the Health Economics Internal Seminar at the University of Birmingham, the C2E2 Seminar at the University of British Columbia, the Health Economists' Study Group conference and the International Health Economics Association conference. Thank you so much to all of those who have offered their comments and feedback on this research.

I would also like to thank those who have read drafts of this work. Thank you to Alastair and particularly Sam, who kindly gave up her time to read a full draft of this thesis.

Finally, I am so grateful to my family for their support, and want to thank them for always being my inspiration. To my Grandparents, Dad, Mum and Carla, you have continually

encouraged me and made me believe that it was all possible. Matt, a special thank you for being there for me every day, and for always offering to help me in any way that you could.

Table of Contents

Abstract.....	i
Acknowledgements	iii
List of figures	xiii
List of tables	xiv
INTRODUCTION	1
CHAPTER 1: BACKGROUND TO THE USE OF DECISION-ANALYTIC MODELS	4
1.1 Introduction	4
1.2 Positive and normative economics	5
1.3. Welfarism	6
1.3.1 Pareto optimality criterion	7
1.3.2 Criticisms of welfarism.....	8
1.4 Extra-welfarism	10
1.5 Economic evaluation in healthcare.....	13
1.5.1 Cost-benefit analysis.....	14
1.5.2 Cost-effectiveness analysis and cost-utility analysis	15
1.6 Decision-analytic modelling for economic evaluation.....	17
1.6.1 Structure.....	18
1.6.2 Data	19
1.6.2.1 Discounting	20
1.6.3 Assessment of uncertainty	20
1.6.4 The submission of decision-analytic models to decision-making bodies.....	22
1.7 Conclusion.....	23
CHAPTER 2: METHODOLOGICAL ERRORS IN MODELS AND A REVIEW OF THE MODEL-BUILDING GUIDANCE	24
2.1 Introduction	24
2.2 The definition of modelling error	24
2.3 Methodological errors in economic models	25
2.4 A systematic review of current model-building guidance.....	29
2.4.1 Review question.....	29
2.4.2 The search strategy	29
2.4.2.1 Pilot search	31
2.4.2.2 Electronic database searching	31
2.4.2.3 Reference list searching	32
2.4.2.4 Searches in textbooks.....	32

2.4.2.5 Online sources	33
2.4.3 The selection of literature	33
2.4.3.1 Inclusion and exclusion criteria.....	33
2.4.3.2 Quality of the papers	34
2.4.4 Data extraction	34
2.4.5 Narrative synthesis.....	34
2.4.6 Results of the review.....	35
2.4.6.1 Characteristics of the papers	36
2.4.6.2 Principal findings of the review	56
2.4.6.3 Quality of the papers	57
2.4.6.4 Other systematic reviews in the research area	58
2.4.7 Conclusion	59
CHAPTER 3: THE USE OF QUALITATIVE METHODS TO UNDERSTAND METHODOLOGICAL ISSUES IN HEALTH RESEARCH	61
3.1 Introduction	61
3.2 Qualitative research	62
3.3 The use of qualitative methods in health research.....	65
3.3.1 The use of qualitative research to understand how people complete preference-based valuation tasks.....	67
3.3.2 The use of qualitative research to improve recruitment to randomised controlled trials	72
3.3.3 The use of qualitative research to explore decision-analytic model development	77
3.3.3.1 Review question	77
3.3.3.2 The search strategy.....	77
3.3.3.3 The selection of literature.....	79
3.3.3.4 Data extraction and synthesis	80
3.3.3.5 Results of the review	80
3.3.3.6 Conclusion of review	93
3.4 Conclusion	94
CHAPTER 4: METHODS	96
4.1 Introduction	96
4.2 Overview of research design	96
4.3 Phase one: Qualitative interviews with modellers.....	97
4.3.1 Research design	97
4.3.2 The sampling and selection of informants for interview	98
4.3.2.1 Sample size.....	100

4.3.2.2 Recruitment	100
4.3.3 Data collection	102
4.3.4 Analysis	104
4.3.5 Ethical considerations	107
4.3.6 The interview informants	108
4.4 Phase two: Model-building case studies with modellers and clinicians.....	116
4.4.1 Research design	116
4.4.1.1 Case study selection	118
4.4.1.2 Recruitment	119
4.4.2 Methods and data collection	120
4.4.2.1 Non-participant observation	120
4.4.2.2 Qualitative interviews	121
4.4.2.3 Think-aloud	123
4.4.3 Analysis	124
4.4.4 Ethical considerations	129
4.4.5 The case study informants	130
4.4.6 Presentation of findings	132
CHAPTER 5: QUALITATIVE INTERVIEWS WITH MODELLERS ON THE MODEL- BUILDING PROCESS	133
5.1 Introduction	133
5.2 Experience of modelling.....	133
5.3 The modelling process.....	136
5.3.1 Defining model structure	136
5.3.2 Developing model structure.....	137
5.3.2.1 Other model structures	137
5.3.2.2 Clinical literature.....	139
5.3.2.3 Use of clinical data.....	139
5.3.2.4 Drafting the structure	141
5.3.2.5 Clinician involvement	143
5.3.2.6 Model type.....	145
5.3.2.7 Making structural assumptions	146
5.3.2.8 Implementation.....	147
5.3.2.9 Iterative.....	148
5.3.2.10 Summary: developing model structure	148
5.3.3 Populating the model	149

5.3.3.1 Identifying data	149
5.3.3.2 Making data assumptions	151
5.3.3.3 Clinician involvement	152
5.3.3.4 Summary: populating the structure	153
5.3.4 Model checking.....	154
5.3.4.1 Sensitivity analysis.....	154
5.3.4.2 Internal validation	155
5.3.4.3 External validation	157
5.3.4.4 Structural checks	158
5.3.4.5 Clinician involvement	158
5.3.4.6 Other involvement.....	160
5.3.4.7 Summary: model checking.....	161
5.4 Clinician involvement	162
5.4.1 Recruitment.....	162
5.4.2 Number recruited	163
5.4.3 Communication.....	166
5.4.3.1 Meetings	166
5.4.4 Relationship	167
5.4.5 Engagement	168
5.4.5.1 Understanding	168
5.4.5.2 Personality.....	170
5.4.5.3 Vested interest.....	170
5.4.5.4 Methods of engagement	171
5.4.6 Summary: clinician involvement.....	172
5.5 Model reflection	173
5.5.1 Good and poor modelling practice.....	174
5.5.1.1 Expertise.....	174
5.5.1.2 Rigour.....	175
5.5.2 Good and poor modelling outcomes	176
5.5.2.1 Integrity	176
5.5.2.2 Representative structure	177
5.5.2.3 Answer	178
5.5.3 Ideal model	179
5.5.4 Summary: model reflection.....	180
5.6 Modelling guidance	180

5.6.1 Use of guidance	181
5.6.2 Attitude towards guidance	182
5.6.3 Summary: modelling guidance	183
5.7 Future research	184
5.8 Conclusion	186
CHAPTER 6: RESULTS OF MODEL-BUILDING CASE STUDIES WITH MODELLERS AND CLINICIANS	189
6.1 Introduction	189
6.2 Case Study A – Diagnostic model in Condition 1	190
6.2.1 Background	190
6.2.2 Structural development	191
6.2.2.1 Beginning structural development	191
6.2.2.2 Pathway plotting	191
6.2.2.3 Boundaries of model structure	193
6.2.2.4 Basis of model structure	194
6.2.2.5 Making structural assumptions	196
6.2.2.6 Communication in structural development	197
6.2.2.7 Role of the clinical collaborators	198
6.2.2.8 Location of treatment pathway	199
6.2.2.9 Implementation	200
6.2.2.10 Reflections on structure	200
6.2.2.11 Summary of structural development	201
6.2.3 Data and populating the model	202
6.2.3.1 Clinical data	202
6.2.3.2 Economic data	202
6.2.3.3 Data and clinician involvement	205
6.2.3.4 Summary of data and populating the model	205
6.2.4 Results	206
6.2.5 Model checking	207
6.2.5.1 External validation	209
6.2.5.2 Clinician involvement in model checking	209
6.2.5.3 Summary of results and model checking	209
6.2.6 Clinician involvement in model development	210
6.2.6.1 Number of clinicians	210
6.2.6.2 Clinician understanding of modelling	211

6.2.6.3	Guidance for clinicians.....	211
6.2.6.4	Summary of clinician involvement in model development	212
6.2.7	Conclusion	213
6.3	Case Study B – Diagnostic model in Condition 2	216
6.3.1	Background	216
6.3.1.1	Clinician training.....	217
6.3.2	Structural development	218
6.3.2.1	Beginning structural development	218
6.3.2.2	Pathway plotting.....	219
6.3.2.3	Boundaries of model structure	221
6.3.2.4	Basis of model structure.....	222
6.3.2.5	Making structural assumptions	223
6.3.2.6	Communication in structural development	224
6.3.2.7	Implementation.....	226
6.3.2.8	Reflections on model structure.....	226
6.3.2.9	Summary of structural development	226
6.3.3	Data and populating the model	227
6.3.3.1	Clinical data.....	227
6.3.3.2	Economic data	228
6.3.3.3	Data and clinician involvement.....	229
6.3.3.4	Summary of data and populating the model.....	230
6.3.4	Results.....	231
6.3.5	Model checking.....	233
6.3.5.1	External validation	234
6.3.5.2	Clinician involvement in model checking.....	234
6.3.5.3	Summary of results and model checking	235
6.3.6	Clinician involvement in model development	236
6.3.6.1	Number of clinicians	236
6.3.6.2	Clinician understanding of modelling.....	237
6.3.6.3	Guidance for clinicians.....	238
6.3.6.4	Summary of clinician involvement	239
6.3.7	Conclusion	239
7.1	Introduction	243
7.2	Summary of findings of the qualitative interviews with modellers	244
7.3	Summary of findings of the case studies with modellers and clinicians	246

7.3.1 Comparison of the case studies.....	247
7.4 Synthesis of findings from the different aspects of the research.....	251
7.5 Findings compared to existing literature.....	254
7.6 Implications of the findings for model building practice.....	257
7.6.1 Implications for structural development.....	257
7.6.1.1 Model conceptualisation: planning the structure and determining model boundaries.....	258
7.6.1.2 Basis of structure.....	260
7.6.2 Implications for clinician involvement.....	263
7.6.2.1 Number of clinicians.....	263
7.6.2.2 Communication and training for clinicians.....	264
7.6.3 Implications of a lack of time and resources for model development.....	267
7.6.4 Exploration and documentation of the practices of modellers.....	268
7.7 Future research.....	270
7.8 Reflections on the method.....	270
7.9 Conclusion.....	277
APPENDICES.....	280
Appendix 1: Search strategy for the systematic review of modelling guidance.....	281
Appendix 2: Data extraction form - systematic review of modelling guidance.....	285
Appendix 3: Research protocol for the systematic review of modelling guidance.....	286
Appendix 4: Search strategy for the systematic review of qualitative modelling studies.....	288
Appendix 5: Data extraction form - systematic review of qualitative modelling studies.....	291
Appendix 6: Research protocol for the systematic review of qualitative modelling studies.....	292
Appendix 7: Ethical Approval for phase one of the research.....	294
Appendix 8: Ethical Approval for phase two of the research.....	296
Appendix 9: Ethical approval for the Canadian phase of the research.....	298
Appendix 10: Sample invitation letter to modellers – phase one.....	300
Appendix 11: Informant information sheet – phase one.....	302
Appendix 12: Interview guide – phase one.....	305
Appendix 13: Initial version of coding structure – phase one.....	307
Appendix 14: Final version of coding structure – phase one.....	310
Appendix 15: Final version of coding structure definitions.....	314
Appendix 16: Excerpt from matrices comparing responses of interview informants – phase one.....	322
Appendix 17: Consent form for phase one of the research.....	324

Appendix 18: Informant information sheet - phase two.....	326
Appendix 19: Interview guide for phase two	329
Appendix 20: Informant interview information sheet – phase two	331
Appendix 21: Consent form for phase two of the research (consent for interview)	334
Appendix 22: Thematic framework for Case Study A	336
Appendix 23: Thematic framework for Case Study B	339
REFERENCES	342

List of figures

Figure 1: QUOROM flow diagram of study selection (updated review).....	36
Figure 2: QUOROM flow diagram of study selection	82
Figure 3: Excerpt of framework for Case Study A.....	128
Figure 4: Diagram of the typical model-building process with questions raised	187
Figure 5: Modelling process for Case Study A, with issues highlighted	214
Figure 6: Modelling process for Case Study B, with issues highlighted.....	241
Figure 7: A suggested hierarchy of sources to inform model structure.....	262

List of tables

Table 1: Summary of paper characteristics	38
Table 2: Summary of paper characteristics	84
Table 3: Interview informant characteristics	110
Table 4: Case study informants with their roles in model development	131
Table 5: Informant's backgrounds and modelling work (by number of informants).....	134
Table 6: Summary of type of guidance used or referred to by informants.....	181
Table 7: Informant suggestions for future research.....	185

INTRODUCTION

This thesis has used qualitative research to explore the process of decision-analytic modelling from the perspective of those involved. The research was undertaken on the basis of literature that highlighted methodological errors in models, and the findings of a systematic review suggesting that there was limited detailed guidance available to support modellers in model development. Exploratory qualitative methods were chosen, specifically in-depth interviews with modellers, and case studies with modellers and clinicians, which used non-participant observation, think-aloud and qualitative interviews. The aim of the research contained within this thesis was to understand where problems were occurring in modelling processes, and provide recommendations around good practice in decision-analytic model development.

The first chapter sets the context for this thesis in the use of decision-analytic models for economic evaluation, and in healthcare funding decisions. The chapter provides a summary of the theoretical and practical basis for decision-analytic models, and outlines their use by decision-making organisations such as the National Institute for Health and Care Excellence (NICE) in the UK and the Canadian Agency for Drugs and Technologies in Health (CADTH) in Canada. Decision-analytic models are used within these agencies to compare the costs and health consequences of competing healthcare interventions to determine their relative cost-effectiveness.

The second chapter presents a review of the modelling literature, demonstrating methodological errors present in both industry submission and published decision-analytic models. The latter part of the chapter presents a systematic review of model-building guidance, which aimed to determine what guidance was currently available to assist modellers

in the modelling process, and where further guidance and research on modelling methods might be required.

The third chapter explores how qualitative research has been used in health services research and health economics to understand methodological issues in healthcare processes. The chapter begins by outlining the methodological basis for qualitative research, including its theoretical underpinnings, the nature of the methods applied, and how qualitative methods have generally been used in health research. A review was then undertaken of the use of qualitative methods to understand and improve processes in two other areas, to ascertain whether qualitative research could also be used to understand problems in the development of decision-analytic models. The areas chosen were individual's completion of valuation tasks, and issues with recruitment to randomised controlled trials. The latter part of the chapter presents a systematic review of the small number of existing studies that have used qualitative methods to explore decision-analytic model development.

The fourth chapter concerns the design of the empirical research reported within this thesis. The first part of this chapter outlines the research design for the first phase of the research, detailing the sampling, recruitment, data collection and analysis undertaken for the in-depth interviews with modellers. The latter part of the chapter details the research design chosen for the two model-building case studies undertaken within this thesis. Case study selection, recruitment of informants, data collection, analysis and the qualitative methods used within each of the case studies are discussed.

The fifth and sixth chapters present the results of the empirical work undertaken within this thesis. The fifth chapter focuses on the findings of the in-depth interviews with modellers, and is organised under the themes that emerged from the qualitative analysis, including the modelling process, clinician involvement, model reflection, modelling guidance and future research. The findings of the two case studies with modellers and clinicians are presented

individually in chapter six, with the first case study focusing on the processes and methods used to develop an academic model, and the second the development of a model within a healthcare policy context.

The final chapter discusses the findings of the thesis, as presented in the previous two chapters. The results of the first phase of the research are summarised, and the findings of the two case studies are compared for similarities and differences. Findings across both phases of the research are then considered and synthesised into common themes. Later in the chapter, the findings of this research are compared to existing relevant literature, and reflections are given on the appropriateness of the methods used. The final part of the chapter discusses the implications of research conclusions for model building practice, and offers recommendations for improvements to modelling practice and suggestions for future research.

CHAPTER 1: BACKGROUND TO THE USE OF DECISION-ANALYTIC MODELS

1.1 Introduction

According to the World Health Organization (2015), ‘the ultimate goal of primary health care is better health for all’, of which a key element is ‘organizing health services around people’s need and expectations’. The organisation of healthcare typically falls to those who determine and manage healthcare systems, and requires decisions to be made about how healthcare resources are allocated (Wonderling et al. 2005). However, the resources used to produce healthcare services, in a similar sense to other economic goods, are finite, and scarce in relation to the demand for them (McPake and Normand, 2008; Morris et al., 2007). Because of this scarcity, and because healthcare is seen as fundamental to people’s lives, choices around how to allocate healthcare resources are difficult. Economics, and the undertaking of economic analyses can potentially assist in making these difficult choices.

This chapter describes the theoretical and practical background to the use of decision-analytic models in healthcare decision-making. The chapter begins by setting the context of decision-analytic modelling in normative economics and economic evaluation, and outlines the different theoretical approaches that can underpin economic analysis. Next the different practical applications of economic evaluation are presented, and it is ascertained that economic evaluation and decision-analytic modelling is typically undertaken within an extra-welfarist framework, using cost-effectiveness analysis. Finally, a decision-analytic model is defined, and key aspects of model development are described.

1.2 Positive and normative economics

Within health economics, and the economics field generally, distinctions are made between positive and normative approaches to economic study (Zweifel and Breyer, 1997). Positive economics is concerned with ‘investigating the relationship between economic variables’, and providing empirical insight into societal behaviour, such as the factors that determine the demand for healthcare (Morris et al. 2007, p.15). Normative economics on the other hand involves addressing decisions about how economic systems should work (Black et al., 2009), such as how the demand for healthcare should be addressed. The difference in the approach of these two branches of economics, then, is the distinction between looking at how things are done (positive economics) and how things ought to be done (normative economics).

Normative economics requires value judgements to be made about the outcomes that are most desirable within a particular context, and how these can best be achieved. Morris et al. (2007, p.16) define a value judgement as ‘a weighing up of evidence based on the ethical and ideological values held by an individual or society’. In the context of health, normative economics is needed to make choices about how scarce healthcare resources should be produced and distributed, with these judgements being informed by efficiency and equity considerations (Black et al., 2009). Economic evaluation provides a normative approach to resource allocation within the health sector, allowing alternative health technologies, treatments and interventions to be compared according to both their costs and consequences (Drummond et al., 2005). The normative elements underpinning the analytical approaches used within economic evaluation are thus the decisions around how these costs and consequences are defined and measured. Although definitions of costs are fairly consistent across the different approaches to economic evaluation, definitions of consequences in terms of the outcomes of value, and the way in which benefit is measured, vary. The next sections of this chapter outline the theoretical frameworks that traditionally form the basis of the

different approaches to economic evaluation in healthcare. These are known as welfarism and extra-welfarism.

1.3. Welfarism

Welfarism or neo-classical welfare economics is the traditional theoretical basis for economic evaluation (McGuire, 2001). ‘Welfarism is generally defined as the systematic analysis of the social desirability of any set of arrangements, for example a state of the world or allocation of resources, solely in terms of the utility obtained by individuals’ (Morris et al. 2007, p.120).

This is utility defined as the representation of an individual’s preference for a particular good, in terms of the satisfaction and/or happiness that they gain from it (Brouwer et al., 2008).

Indeed, the only outcome of interest in resource allocation decisions undertaken within the welfare economics framework is the utility gained by individuals. This is reiterated in the four central tenets that form the basis of the welfarist approach (Hurley 1998, pp.376-77):

1. Utility maximisation – individuals are rational and consistent in their choices, and will seek to maximise their own utility.
2. Consumer sovereignty – individuals are considered to be the best judges of their own welfare, and thus individual assessments of well-being are what matter in decision-making. The opinions of any third party are irrelevant.
3. Consequentialism – based on consumer theory, welfare economics only considers the outcomes and resulting utilities for individuals in relation to their consumption of particular goods or services. The process involved in the individual ‘choosing or receiving those goods’ is not accounted for (Morris et al. 2007, p.211).
4. Welfarism – the desirability of any situation, for example, a resource allocation decision, can be judged solely on the basis of the subsequent utilities achieved by the individuals affected.

Within Paretian welfare economics, to allow particular states of the world to be compared and ranked, and thus decisions to be made around how healthcare resources should be allocated, trade-offs in terms of the utility attained by individuals need to be made. These trade-offs were originally undertaken on the basis of the criterion of Pareto optimality; for practical reasons, this was later replaced by the notion of the potential Pareto improvement (Hurley, 1998). Both concepts are outlined below.

1.3.1 Pareto optimality criterion

McGuire (2001) asserts that Pareto optimality encompasses two concepts: Pareto improvement and Pareto efficiency. The former concept is considered to be a weak value judgement, thus uncontroversial, as it suggests that all individuals should benefit from the reallocation of resources (Boadway and Bruce, 1984). Pareto efficiency however, is defined as a strong value judgement, in asserting that ‘a reallocation of resources is efficient if at least one individual in the economy is made better off and no other individual is made worse off’ (McGuire 2001, p.3). With both approaches, ‘a resource allocation is judged to be Pareto optimal if and only if it is impossible to increase one person’s utility without simultaneously decreasing another’s’ (Hurley 1998, p.377). Both concepts can theoretically be applied within welfare economics, and to policy decisions; however, there are problems with the practical application of these value judgements. First, Pareto optimality is not concerned with the distribution of utilities across individuals, and thus could result in a highly inequitable allocation of resources (Tsuchiya and Williams, 2001). Second, Pareto optimality cannot rank all states of the world, specifically those that are considered Pareto optimal, as to order these states would require some individuals to lose whilst others gain (Tsuchiya and Williams, 2001). As McPake and Normand (2008) note, there are very few policy decisions that would satisfy every individual and/or guarantee no losers. Therefore, use of the Pareto optimality

criterion eventually leads to ‘policy paralysis’, as a ‘single, best allocation’ of resources cannot be generated (Hurley 1998, p.377).

The practical drawbacks associated with the Pareto optimality criterion led to the adoption of the potential Pareto improvement concept. This concept is based on the independent works of Kaldor (1939) and Hicks (1939), and is also often referred to as the Kaldor-Hicks criterion, or the compensation principle. The notion behind the potential Pareto improvement is that if those who gain from a particular resource allocation decision can theoretically compensate those who lose and still be better off, in essence there are no losers, and thus a Pareto improvement is possible (McPake and Normand, 2008). Using the compensation principle, further trade-offs can be made between states, provided that the gains to those who gain from a particular reallocation of resources exceed the losses to those who lose (McPake and Normand, 2008). Although in reality compensation is rarely exchanged, the fact that the losers can potentially be compensated is considered an overall improvement to societal welfare.

The practical application of the potential Pareto improvement in economic evaluation is through cost-benefit analysis. The cost-benefit approach uses monetary weights as a proxy for utility, and asks consumers to value the benefits of competing resource allocations through their willingness-to-pay (Tsuchiya and Williams, 2001). A change in policy, for example, the introduction of a new healthcare service and/or intervention, is deemed to be justifiable if the monetary benefits outweigh the costs of implementation i.e. there is an overall net gain to society (Robinson, 1993a).

1.3.2 Criticisms of welfarism

Despite the theoretical basis that welfarism apparently offers to economic evaluation, related to its grounding in microeconomic theory, there are criticisms around the ability of welfare

economics to address concerns that are important to the allocation of healthcare resources (Morris et al., 2007). In terms of its theoretical groundings, Sen (1985) questioned the focus of welfare economics on utility, and the appropriateness of utility as a measure of individual well-being. Sen (1985) argues that individuals differ in terms of their capacity to derive utility and thus well-being from commodities, as people adapt their desires and levels of happiness to their expectations and real life situations. Individual assessments of well-being and utility are thus affected by personal characteristics (Morris et al., 2007); however, personal characteristics are ignored in resource allocation decisions based within the welfarist framework. Further, it has been argued that measuring social welfare only in terms of the utility that individuals gain from consumption is too restrictive, as health can only be valued insofar as it results in utility (Coast, 2009; Mooney and Russell, 2003). Arguably, additional factors, such as changes to an individual's health, may be important in decisions around the reallocation of healthcare resources. Such criticisms of utility raise issues around the suitability of using welfare economics in a healthcare decision-making context, given the reliance of the framework on consumer choice theory as a predictor of consumer behaviour (Morris et al., 2007). Indeed, Arrow (1963) highlights how the nature of people's demand for, and consumption of healthcare is irregular and unpredictable when compared to that for other economic goods, and how elements of traditional economic theory are largely not applicable within the healthcare market.

On a practical level, welfare economics is criticised because of its indifference to the distribution of resources, on the basis of the use of the potential Pareto improvement criterion. Monetary value is used as a proxy for utility within cost-benefit analysis, and thus benefit is measured on the basis of income gains or losses (McGuire, 2001). The problem with the criterion is that it assumes that individuals place the same amount of value on one unit of currency, when in reality this value is likely to be influenced by initial level of income

(Tsuchiya and Williams, 2001). The implication of this, as outlined by Coast et al. (2008), is that measures of willingness-to-pay are likely to be affected by an individual's ability to pay, and thus the allocation of healthcare resources could be 'skewed' towards the wealthy. Given that interpersonal comparisons are not permitted using the welfarist approach (i.e. everyone starts from the same point), and compensation under the potential Pareto improvement criterion is unlikely to be paid, decisions could result in the very inequitable distribution of healthcare resources (McGuire, 2001). In response to these criticisms, the extra-welfarist approach was developed.

1.4 Extra-welfarism

Brouwer et al. (2008) imply that the introduction and use of extra-welfarism has grown in part from 'seeds' of criticism sown by Sen (1970; 1980). In summary, this refers to the focus of welfare economics on utility as being too narrow a measure of well-being, and the argument that a broader perspective is needed to measure the wider capabilities that a person can acquire from a particular healthcare good. Further, welfarism was criticised for not explicitly taking into account equity concerns under the strict application of the Pareto principle. The work of Sen and also Culyer (1989) is largely credited for the development of the extra-welfarist approach as an alternative framework for economic evaluation. In terms of the contribution of Culyer (1989), this particularly relates to the maximisation and distribution of health, as opposed to utility, as the outcome of focus in resource allocation decisions. Culyer (1989) asserts that healthcare resources should be allocated on the basis of need, as the need for healthcare implies a deprivation in health, and thus healthcare can reduce this deprivation. The extra-welfarist approach therefore takes account of personal characteristics within its decision-making, given that these characteristics describe and provide a comparison of people's ill-health, incorporating factors into evaluations such as whether or not people are happy, free of pain, and/or physically mobile (Brouwer et al., 2008; Culyer, 1989). Equity

issues can then be addressed through adding distributional weights to particular characteristics of people, on the basis of societal values outlining who needs to benefit from health production (Culyer, 1989; Hurley, 1998). The very definition of extra-welfarism implies that the approach adds something ‘extra’ to welfarism, and indeed, Brouwer et al. (2008) suggest that extra-welfarism transcends welfarism in four general ways. These are that ‘it permits the use of outcomes other than utility, it permits the use of sources of valuation other than the affected individuals, it permits the weighting of outcomes (whether utility or not) according to principles that need not be preference-based, and it permits interpersonal comparisons of well-being in a variety of dimensions’ (Brouwer et al. 2008, p.330).

Culyer (1989) also advocates the use of the quality adjusted life year (QALY) as an approach to health measurement. ‘The QALY is a measure of health outcome which assigns to each period of time a weight, ranging from 0 to 1, corresponding to the quality of life during that period, where a weight of 1 corresponds to perfect health and a weight of 0 corresponds to a health state judged equivalent to death’ (Garber et al. 1996, p.29). The benefit of a health intervention is thus measured as ‘the difference over a period of time between expected QALYs with a particular procedure and without it (or with an alternative)’ (Culyer 1989, p.52).

However, extra-welfarism has been met with some criticism, particularly in terms of its practical application, as it has been argued that the focus on health as the single outcome of importance within economic evaluation is too narrow and restrictive (Coast, 2004; Hurley, 1998). Hurley (1998) asserts that extra-welfarism and welfarism share the same limitation, in the fact that they are both consequentialist in nature, and thus the benefits of an intervention within the extra-welfarist framework can only be valued in terms of the intervention’s ability to produce health. It is possible that other, non-health related, benefits could be generated as a result of a particular healthcare programme (Coast et al., 2008); however, these cannot be

captured using health focussed measures, such as the QALY. A further criticism relates to the consideration of equity within the extra-welfarist framework, and the implicit assumption underlying the calculation of health effectiveness, that QALYs are of equal value to society no matter who gains them (Birch and Donaldson, 2003). Russell et al. (1996) argue that, as a society, we are likely to place different values on different people's need for healthcare, and although the extra-welfarist approach theoretically allows us to do this, in practice there is uncertainty in applying these distributional weights.

Despite such problems, extra-welfarism, including the focus on the QALY as a measure of health, is typically the basis on which economic evaluation is undertaken within the UK and internationally (Coast et al., 2008; Torrance et al., 1996; Tsuchiya and Williams, 2001). In the UK and Canada (the settings for empirical work undertaken within this thesis), the National Institute for Health and Care Excellence (NICE) and the Canadian Agency for Drugs and Technologies in Health (CADTH) respectively, are responsible, among other things, for advising via technology appraisal or health technology assessment (HTA), whether new health technologies should be adopted (CADTH, 2015; NICE, 2015a). For NICE this is in the form of recommendations made to the National Health Service (NHS), and for CADTH, advice to decision and policy makers responsible for funding decisions in the various Federal, Provincial and Territorial Ministries of Health across Canada. In the case of NICE, the NHS, at least in England, is legally obliged to fund and resource treatments and medicines that are recommended within their technology appraisals (NICE, 2015b). Both NICE and CADTH rely on information submitted from external sources (for example, academic groups and manufacturers and sponsors), and base their recommendations on economic and clinical evidence, specifically the economic evaluation of competing health interventions in terms of their relative cost-effectiveness (CADTH 2006; NICE 2013).

The purpose of economic evaluation, and the practical applications of the welfarist and extra-welfarist approaches to economic evaluation, are outlined next. An introduction to decision-analytic modelling is given in section 1.6.

1.5 Economic evaluation in healthcare

In economic evaluation, choices around the allocation of scarce healthcare resources are based on comparative evidence of the costs and health benefits associated with competing interventions (Drummond et al., 2005). Gray et al. (2011) suggest that economic evaluation has developed from the recognition that cost and effectiveness information, particularly related to the benefits forgone by choosing a particular course of action, is necessary for decision-making. Thus, underlying the choices made in the allocation of healthcare resources, is the notion of opportunity cost, that committing resources to the production of one good or service 'is the benefits forgone from those resources not being used in their next best alternative' (Morris et al. 2007, p.3). A criterion used to help decision-makers to decide across alternative options, is efficiency (Shiell et al., 2002). There are two key types of efficiency with which economics is concerned: technical efficiency and allocative efficiency. Technical efficiency relates to how best to allocate resources, either in terms of maximising a particular output relative to its input, or minimising costs to achieve a desired level of output (Shiell et al., 2002). Allocative efficiency is concerned with questions around whether to allocate resource to an intervention, and is said to have been met when no further benefit can be generated from reallocating resources between interventions within a healthcare system (Shiell et al., 2002). The different approaches to economic evaluation addresses these efficiency concerns to various degrees, in addition to using different methods to value and measure outcomes for decision-making. The next sections briefly describe the different types of economic evaluation undertaken, specifically cost-benefit analysis, cost-effectiveness analysis and cost-utility analysis.

1.5.1 Cost-benefit analysis

Cost-benefit analysis is theoretically grounded in welfarism, and is a method of economic evaluation where both the costs and benefits of a healthcare intervention are valued in monetary units (Wonderling et al., 2005). Monetary equivalents are assigned to prolongations of life and change of health status, derived from the use of a particular treatment (Zweifel and Breyer, 1997). Using units of money as a measure of (positive and negative) benefit allows the cost-benefit of interventions and services to be compared within the health sector and across different areas of health, but also between different areas of expenditure within society (Gray et al., 2011). To generate the results of a cost-benefit analysis, the cost of an intervention is compared with the monetary value of the outcomes achieved, producing a calculation of net benefit (Wonderling et al., 2005). Net benefit involves subtracting the resource cost of an intervention from its monetary benefit, asserting that if the benefit from an intervention is greater than the cost i.e. the net benefit is greater than zero, an intervention is cost-beneficial and should be implemented (Wonderling et al., 2005). Competing interventions can thus be compared according to the level of net benefit offered. Cost-benefit analysis enables both technical and allocative efficiency concerns to be addressed, given that the results can indicate which health intervention(s), if funded, maximise societal welfare, and how resources can optimally be allocated within a healthcare system, and across broader sectors of the economy.

However, the cost-benefit approach has had limited use in economic evaluation and policy decision-making, due to equity (see section 1.3.2) and ethical concerns around whether it is appropriate to value health benefit, and thus human life, in monetary units (Ryan et al., 2003; Zweifel and Breyer, 1997).

1.5.2 Cost-effectiveness analysis and cost-utility analysis

Both cost-effectiveness analysis and cost-utility analysis have their theoretical underpinnings in extra-welfarism, and thus both focus on the comparison of healthcare interventions in terms of the health outcomes generated. Cost-effectiveness analysis involves comparing the cost of the resources spent on an intervention, with the quantity of health gained as a result (Wonderling et al., 2005). Health outcomes are measured in natural units on a unidimensional scale i.e. the unit of outcome is the same between interventions, for example, different diagnostic tests might be compared according to the cost per case detected (Gray et al., 2011; Zweifel and Breyer, 1997). The results of a cost-effectiveness analysis can therefore address questions around technical efficiency, in indicating whether an intervention compares favourably to an alternative, specifically whether it can minimise the costs associated with achieving a certain level of health benefit (Donaldson, 1998). However, information on allocative efficiency from a cost-effectiveness analysis is limited, as interventions cannot directly be compared across different areas of health, or across different sectors of the economy.

When using cost-effectiveness analysis, the costs and effects of an intervention and its alternatives are established, and the differences in costs and the differences in effects between the interventions are calculated and presented in the form of a ratio (Gray et al., 2011). This is referred to as an incremental cost-effectiveness ratio (ICER), and is the difference in costs divided by the difference in effects between competing options. An ICER can thus be interpreted as the incremental price of obtaining an additional unit of health effect from a particular health intervention, when compared to an alternative (Garber et al., 1996). When an intervention is both less expensive and more effective than its alternative, it is considered to be dominant, and thus the most cost-effective option (Robinson, 1993b). However, an intervention may be more costly but also more effective than its alternative, and in this case

the ICER must be compared to a given cost-effectiveness threshold, which represents society's willingness-to-pay for a unit of health benefit (Morris et al., 2007). Although there is uncertainty around how the cost-effectiveness threshold should be determined (McCabe et al., 2008), current cost-effectiveness thresholds are valued in cost-per-QALY (Morris et al., 2007), thus requiring a cost-utility analysis to be undertaken.

Considered to be a subset of the cost-effectiveness analysis approach, cost-utility analysis typically involves comparing the benefits of interventions in terms of QALYs. The QALY combines into one measure the effectiveness of an intervention in relation to survival in terms of life-years and health-related quality of life (Gray et al., 2011). Methods to derive quality of life weights draw on utility theory, and involve asking people to value or indicate their preference for spending time in a particular diminished health state, or, in the case of those affected, asking them to assess their quality of life directly (McPake and Normand, 2008). Like cost-effectiveness analysis, the most cost-effective intervention, from those compared, is considered to be that which produces the most QALYs for the least cost, subject to a budget constraint (Tsuchiya and Williams, 2001). A cost-utility analysis allows both technical and allocative efficiency questions to be addressed, as the results can give an indication as to which interventions offer the most health gain for the least financial investment, and interventions can be compared across different clinical areas. Both NICE and CADTH advocate the use of cost-utility analysis as the appropriate method for economic evaluation (NICE 2013; CADTH 2006).

'For the reference case, cost-effectiveness (specifically cost-utility) analysis is the preferred form of economic evaluation. This seeks to establish whether differences in expected costs between options can be justified in terms of changes in expected health effects. Health effects should be expressed in terms of QALYs' (NICE 2013, p.37)

'Use a cost-utility analysis (CUA) as the Reference Case where meaningful differences in health-related quality of life (HRQL) between the intervention and comparators have been demonstrated' (CADTH 2006, p.4)

1.6 Decision-analytic modelling for economic evaluation

To inform economic evaluation and cost-effectiveness analyses, data on the costs and effectiveness of interventions are needed. To gather these data, a primary research design can be used, in which economic and additional health outcomes are ‘piggybacked’ onto an existing randomised controlled trial (RCT) (Torrance et al., 1996). However, reliance on a single RCT to inform economic evaluation has been criticised because of the inability of a clinical trial to compare all relevant interventions for the treatment of a disease, to incorporate all (alternative) evidence that is important to a resource allocation decision, and to compare the long-term costs and outcomes associated with competing interventions (Mandelblatt et al., 1996; Sculpher et al., 2006). Decision-analytic models are thus recommended as a vehicle for economic evaluation, due to their ability to provide a platform for the comparison of all relevant treatment options, and to synthesise all relevant data over an extended time horizon (Buxton et al., 1997; Sculpher et al., 2006). NICE and CADTH advocate the use of decision-analytic models in economic evaluation and decision-making, with NICE describing modelling as ‘an important framework for synthesising available evidence and generating estimates of clinical and cost effectiveness’, and both organisations including guidance on model building in their respective reference case of methods (NICE 2013, p.49; CADTH 2006).

A decision-analytic model is essentially a mathematical structure that can represent the health and economic outcomes for patient populations receiving particular medical interventions (Kuntz and Weinstein, 2001). A model uses mathematical relationships to express the likelihood of particular health consequences occurring within each of the interventions compared. The nature of these consequences informs the structure of a model. Each consequence has a related cost and outcome, which, when weighted with the probability of its occurrence, enables a calculation of the expected costs and expected outcomes associated with

each intervention under evaluation (Briggs et al., 2006; Drummond et al., 2005). Although guidance on how to develop decision-analytic models varies (this will be explored further in chapter 2), there are broadly considered to be three key aspects to model construction. After a decision problem has been defined (for example, the setting, perspective and disease area of focus), these are: structuring the model, populating the model (identifying and collecting appropriate data), and assessing uncertainty within the model and its results.

1.6.1 Structure

In terms of structural development, there are a number of different types of model that can be used to address a decision problem. Those most commonly referred to are the decision tree, the Markov model and individual sampling models. Of these, the decision tree model is considered the simplest.

In a decision tree, an initial decision node represents the strategies being compared, and downstream from this are a series of branches with accompanying chance nodes, which denote the probability of a patient following one particular clinical pathway as opposed to another. At the end of a decision tree structure are a number of outcome nodes, which indicate both the costs and health effects associated with each pathway (Kuntz and Weinstein, 2001). The cost-effectiveness of each intervention is then determined by rollback calculation of the sum of the values in each pathway, weighted by the pathway probabilities (Briggs et al., 2006).

A Markov model structure is represented as a set of mutually exclusive and exhaustive health states that patients remain in for a fixed period of time (cycle length), and to which transition probabilities are assigned to indicate patients' movement from one health state to another. Each health state is assigned a utility (effectiveness) value, and the overall utility gained from each intervention being compared is dependent on the total length of time that patients spend

in each state, over the time horizon of the model (Sonnenberg and Beck, 1993). Costs are then similarly calculated according to the proportion of patients in a given health state during each cycle (Barton et al., 2004).

Individual sampling models require patients to be tracked individually through a model, following different sequences of health states, and allowing total expected costs and outcomes to be calculated for each patient (Barton et al., 2004). This type of model is aimed at representing the variability between the costs and outcomes of patients, for a more appropriate estimate of cost-effectiveness (Briggs et al., 2006).

Despite the mathematical nature of the relationships underlying a model structure, decisions around the structural inputs are rather more subjective in comparison. This includes decisions on the appropriate approach or type of model to use, how a particular structure represents the disease or condition in question (for example, the pathways or health states involved), and how the boundaries of a model are defined. These decisions are largely made in response to the context of the decision problem and the nature of the condition being modelled (Briggs et al., 2006; Drummond et al., 2005), but are also influenced by the choices and preferences of the modeller.

1.6.2 Data

After a model structure has been developed, data are required to populate the input parameters of a model. These include clinical data, specifically the probabilities associated with the transition of patients from one health state or pathway to another, and clinical effectiveness data, where the outcome of interventions are measured in natural units. Economic data are then needed for the costs and, in the case of cost-utility analysis, utility values (utilities) associated with patients' experiences of a particular pathway or health state within a model. Utilities generally refer to the quality of life weights used to value particular health outcomes,

and therefore inform the overall measure of the effectiveness of an intervention. Both clinical and economic data can be collected through a variety of methods, including the use of primary and secondary sources. Again, the choices around which data to use and how to identify them are somewhat subjective, as modellers have a number of different evidence sources to select from.

1.6.2.1 Discounting

Discounting is based on the principle that all future costs and health consequences within a cost-effectiveness analysis (and thus in a model) should be stated in terms of their “present value” to the decision-maker (Lipscomb et al., 1996). The process of discounting involves identifying the year in which costs and benefits occur, and adjusting these to represent their worth today. Discounting enables a comparison between interventions that incur costs and generate benefits over a number of years. The rationale for discounting encompasses the argument that individuals have a positive time preference for earlier consumption over later consumption, and thus favour having money and benefits now (hence why future costs and consequences are discounted) (Cairns, 2001). Despite widespread acceptance that discounting should be undertaken, there remains controversy over how discount rates should be determined, and what these discount rates should be (for example, whether rates should be the same for costs and benefits) (McPake and Normand, 2008). Currently, for both costs and outcomes, NICE promote a discount rate of 3.5% per year, and CADTH recommend a discount rate of 5% (CADTH 2006; NICE 2013).

1.6.3 Assessment of uncertainty

A requirement of modelling for economic evaluation is to acknowledge the potential uncertainty inherent in a particular cost-effectiveness result (Briggs et al., 2006). Andronis et al. (2009) distinguish between three types of uncertainty: methodological uncertainty,

parameter uncertainty and structural uncertainty. Methodological uncertainty is concerned with the analytical methods used to develop a model, such as how costs and outcomes are discounted, what outcome measures are used, and how costs are calculated and health outcomes valued (Briggs, 2000). Parameter uncertainty relates to the data that inform the input parameters to a model, and whether the estimates used are accurate and appropriate. Structural uncertainty asks whether an appropriate approach has been taken to develop a structure, in terms of pathways and health states used to combine the input parameters (Andronis et al., 2009).

Andronis et al. (2009) assert that all of these types of uncertainty can be addressed by using sensitivity analysis to determine the impact on the model results. Sensitivity analyses on the parameters in a model can be univariate or multivariate, with single or multiple data estimates being varied to note the effect on the results (Manning et al., 1996). With the latter, best/worst case scenario analysis can also be carried out, to check whether the results are robust to the most optimistic or pessimistic combination of input values. As an alternative to deterministic sensitivity analysis (in which uncertain parameters are assigned a point estimate value), probabilistic sensitivity analysis involves each parameter being assigned a distribution, allowing where appropriate for correlation between the uncertainty in different parameters. Values from these distributions are drawn randomly over the large number of times that a model is run, to produce a joint distribution of the differences in cost and effectiveness between the interventions modelled (Andronis et al., 2009; Doubilet et al., 1985). Similarly, methodological and structural uncertainty can be addressed by varying the analytic methods and structural assumptions used, and reporting the effect on the model results (Briggs, 2001; Manning et al., 1996). Model validation and other model checking activities focus on evaluating the base case model. This can involve checks to the inputs and internal workings of

the model, and the validation of the structure, data and results against external sources, such as other studies and expert opinion (Mandelblatt et al., 1996).

1.6.4 The submission of decision-analytic models to decision-making bodies

Although decision-analytic models are developed for different purposes, for example, to generate clinical knowledge in an academic context, NICE and CADTH are concerned with reviewing models that will potentially contribute to policy decision-making, and how healthcare technologies are funded. NICE have a well-established process for the submission of models for technology appraisal, following either the single technology appraisal process (STA) or multiple technology appraisal (MTA) process. Where the former refers to the appraisal of a single technology, through review of cost-effectiveness evidence submitted by a manufacturer/pharmaceutical company, the latter involves the appraisal of multiple technologies, with evidence submitted through a variety of sources and consultee organisations, including manufacturers (NICE, 2014). In the STA process, models are appraised by an independent evidence review group, who subsequently produce a report on the cost-effectiveness of a technology. The MTA appraisal process is similar, but it is an independent academic assessment group who prepare a review and report of the comparative cost-effectiveness of technologies, based on the evidence available. Submissions via both of these processes inform funding recommendations and clinical guidance within the NHS. CADTH uses a similar process to NICE, whereby manufacturers submit cost-effectiveness evidence (including a model) on a particular drug to the common drug review (CDR), where the Canadian Drug Expert Committee (CDEC) evaluate the evidence, and provide recommendations for the funding of a drug on its basis (CADTH, 2014).

1.7 Conclusion

This chapter has given an overview of the background to the use of decision-analytic models, including their theoretical and practical underpinnings in economic theory and economic evaluation. The importance given to use of decision-analytical models in economic evaluation has been outlined, particularly in relation to their ability to compare the cost-effectiveness of numerous healthcare interventions over an extended time horizon. The chapter has also demonstrated the reliance of organisations such as NICE and CADTH on the cost-effectiveness results generated by models, to inform healthcare funding and adoption decisions. However, there have been various reports of errors present in published and industry submission models (Chilcott et al., 2010), with the latter referring to those that are submitted by pharmaceutical companies to NICE, for recommended use within the NHS. The implication of the development of poor quality or ‘erroneous’ models is the potential for these to generate less than optimum cost-effectiveness recommendations and resource allocation decisions. The next chapter explores examples of models in which methodological errors have been identified.

CHAPTER 2: METHODOLOGICAL ERRORS IN MODELS AND A REVIEW OF THE MODEL-BUILDING GUIDANCE

2.1 Introduction

Chapter 1 has introduced the role of decision-analytic models in cost-effectiveness decisions within the UK and internationally. Clearly, and particularly in the context of NICE policymaking, these models play an important part in deciding how scarce healthcare resources are distributed. However, there is some suggestion that economic models are using inappropriate information and methods in their development (Chilcott et al., 2010; Karnon et al., 2007; Roberts et al., 2006). The first section of this chapter discusses the definition of an error, and focuses the thesis on the exploration of methodological ‘errors’ in modelling processes. The second section examines examples of literature that highlight these errors in industry submission and published models. Finally, the chapter finishes with a systematic review of modelling guidance, aimed at identifying the guidance that is currently available to modellers.

2.2 The definition of modelling error

Chilcott et al. (2010) distinguish between two types of model error: (1) those that are related to software implementation and the ‘technical’ aspects of model development, and (2) those that are methodological and concern the choice of approach used by modellers. Whilst Chilcott et al. (2010) argue that the former type of error is well-researched, but unavoidable

within models, the authors suggest that there is less known about methodological errors, and thus greater clarification required in terms of what constitutes poor practice in modelling methods. Decisions around model structure are inevitably subjective, dependent on both the context of the model and the choices of the modeller. However, these subjective decisions are likely to impact on model cost-effectiveness results and recommendations, and thus it seems right to argue that there should be basic standards of good practice in place for modellers to follow. Going forward, this thesis is focused on the context of methodological errors, with the next section exploring examples of existing models in which this type of problem has been identified.

2.3 Methodological errors in economic models

A number of papers have highlighted problems with existing economic models. These include industry submission models, and those that have been published for academic purposes, which can also inform decision-making. The majority of the papers cited refer to industry submission models, and were collected via a search of STAs published on the NICE website. This search generated a number of examples of models that had been deemed to be poor quality by NICE evidence review groups (ERGs).

A commonly highlighted problem in submitted economic models was the validity of clinical pathways or states in the structure. For example, concerning the submission for ranibizumab for the treatment of diabetic macular oedema, the NICE summary of the ERG's comments suggested that the manufacturer's model did not accurately represent clinical practice, as all patients were receiving treatment on only one eye, when in reality a significant proportion receive treatment on two (NICE, 2011). When the cost of this additional treatment was incorporated into the ICER calculations by the ERG, the ICER increased by 50%. Similarly, in the submissions for erubulin and fludarabine monotherapy for the treatment of breast

cancer and chronic lymphocytic leukaemia respectively (NICE, 2007, 2012), the ERG commented that manufacturers had not included important adverse events associated with the drugs, which if incorporated would increase the ICER due to associated disutility. In a submission for Eltrombopag, for the treatment of chronic immune (idiopathic) thrombocytopenic purpura, the ERG considered the model structure to be based on an inappropriate clinical assumption, specifically that patients received treatment solely according to bleeding events experienced, rather than additional outcomes, such as adverse events (NICE, 2010). The ERGs also continually cited problems with the data used to populate models, including the lack of a systematic review for clinical parameters, and highlighted instances of manufacturers using outdated and inappropriate costs, and utility values that were not representative of the patient population.

The above findings are supported by a more rigorous analysis of ERG's critiques of manufacturers' submissions to NICE (Kaltenthaler et al. 2012). Kaltenthaler et al. (2012) found that the majority of ERG reports expressed uncertainty about the quality of submitted models, implying that manufacturers were exaggerating the effect of their technologies, and were not fully representing risk to patients. This relates to the findings of the technology appraisals cited earlier, as manufacturers were criticised for the clinical representativeness of their models, and excluding important adverse events, generating more favourable model outcomes for their interventions.

The review of manufacturer's submissions and related ERG comments did not indicate why these 'errors' were present within the models, although there are a few possible explanations. It may be that manufacturers are intentionally submitting models that provide an optimistic representation of the cost-effectiveness of their interventions. Miners et al. (2004), for example, demonstrated in the case of MTAs, that ICERs submitted by manufacturers for their own interventions were significantly lower than those calculated by independent academic

groups. However, it is also possible that modellers are using inappropriate methods, for example, in not referring to clinical literature or seeking clinician opinion to determine the clinical aspects of a structure. Either way, the suggestion is that modellers could improve their approaches to model development. Although poor practice in these model submissions has been detected (and thus revised, or the particular technology has not been recommended), it is possible that despite thorough consideration from an ERG, some errors go unnoticed. More certain is that the submission of poor quality models to NICE for appraisal is an inefficient use of NICE resources.

There have been a number of studies that have identified errors in published models. Arguably, errors at this level have greater consequence, as the results produced by these models are widely available and could potentially be referred to by the clinical community in decision-making. The studies highlighting errors in published models were either systematic reviews, or case studies of economic evaluations within a specific disease area (Chen et al., 2006; Karnon et al., 2007; Roberts et al., 2006). In their systematic review of economic evaluations of screening for chlamydia trachomatis, Roberts et al. (2006) found that the majority of models published were using inappropriate structures, based on static rather than dynamic approaches (decision trees and Markov models). Only three of forty-two models reviewed were considered to be using a suitable structure, specifically a transmission dynamic model, which could account for interaction between individuals, and represent factors such as the risk of reinfection to a cured individual, as is possible within some infectious diseases. Roberts et al. (2006) concluded that the majority of models were methodologically flawed, which would impact on the validity of their results. Indeed, a study undertaken in the Netherlands that compared the results of a static and dynamic model for chlamydia screening, found that the static structure was likely to underestimate the cost-effectiveness of chlamydia screening, as it could not account for the 'indirect protection effects' associated with an

individual being screened and cured (Welte et al. 2005, p.478). This was in terms of the negative health outcomes and costs avoided in curing the individual and their current partner, and preventing future partners from contracting chlamydia.

A similar issue was identified in a systematic review of cancer screening models undertaken by Karnon et al. (2007). The authors noted that the use of simple decision tree structures within a number of the studies did not reflect the natural history of the disease, and thus did not represent the true costs and consequences associated with the cancer being detected at various stages of its progression i.e. in terms of the aggression of treatment required, and treatment effectiveness in eradicating cancerous cells. Chen et al. (2006) found in reviewing economic evaluations for the treatment of rheumatoid arthritis that published models had used inappropriate structural assumptions, for example, in assuming that patients could not swap between treatments, and had used irrelevant treatment comparators, and inappropriate cycle lengths within Markov models. The industry submission on which two of the published papers were based was deemed particularly problematic, as model health states, costs and utility data were not representative of a recent UK population, but instead were from a decade-old North American database. The authors noted that the two models published that had used these data had reported ICERs within NICE's 'acceptable range' (Chen et al., 2006).

The review of this modelling literature has highlighted methodological errors in industry submission and published models. Problems were identified at various stages of structural development, particularly the use of inappropriate information to inform model structures. These errors were seemingly related to modellers' decisions around model structure, which were demonstrated to have an influential impact on the quality of models, but also on cost-effectiveness results. Arguably more needs to be done to encourage modellers to use appropriate methods; however, it was unclear from the studies why methodological errors were occurring. A possibility is that modellers were not aware of what constitutes good

practice in model development, particularly in terms of structure. The next section looks at the modelling guidance currently available for the model-building process.

2.4 A systematic review of current model-building guidance

A systematic literature review was undertaken to explore current model-building guidance. The aim was to identify guidance available to modellers about modelling methods and processes. The review was motivated by the presence of methodological errors in models, and the question of whether modelling guidance could, if referred to, help to prevent these. This review was originally undertaken in September 2012 as part of an MSc dissertation, but was updated and extended in September 2015 to take into account any recent and relevant studies that had been published. The methods used within the systematic review are described next.

2.4.1 Review question

What guidance currently exists for the process of decision-analytic model building?

The review question was broken down into a number of different aims:

1. To search for guidance on the entire process of model building
2. To search for guidance that offers advice on particular stages of model development
3. To look at what guidance is currently available, for example, who is it aimed at? How detailed is it?
4. To establish which aspect(s) of model development are least focused on within modelling guidance (and potentially require improvement).

2.4.2 The search strategy

In preparation for a full review of the literature and in line with guidance from the Centre for Reviews and Dissemination (CRD 2008), the first stage of the search involved a check for similar reviews. This check was undertaken through a search of the Cochrane Database of

Systematic Reviews. Given that the objective of this phase was to justify the need for a new review in the proposed research area (CRD 2008), the search was designed to be as broad as possible, so as to demonstrate with confidence that there were no current reviews that specifically addressed the research question. The original search undertaken in August 2012 generated 1443 articles, none of which were deemed to be directly relevant to the systematic review proposed. One review was identified on model-building guidance, but its aim was to synthesise existing guidelines to provide advice on good practice, rather than comment on the type of guidance available (Philips et al., 2006). This paper was therefore later captured in the results of this systematic review. However, on updating this review, a recent study was identified that directly addressed the research question, and similarly aimed to explore the focus and quality of existing modelling guidance (Penaloza Ramos et al., 2015). However, the objective of this work was different to that of the current review, as the authors ultimately aimed to compare the advice available in current guidelines for various stages of model development (Penaloza Ramos et al. 2015), as opposed to highlighting stages for which guidance was less detailed, or missing completely. Additionally, the final search date for the current review was 24th September 2015, meaning that the search undertaken here extended beyond that of Penaloza Ramos et al. (2015), which was completed in March 2014. The findings of the Penaloza Ramos et al., (2015) review are compared to the results of the current review in section 2.4.6.4.

Following this stage, a full systematic review was undertaken. This review aimed to be both thorough and exhaustive in identifying papers that offered model-building guidance, to ensure that research conclusions were as informed and unbiased as possible (Petticrew and Roberts, 2006). The search strategy involved a number of stages:

- The searching of major electronic bibliographic databases for potentially relevant articles

- Forward citation searching on key papers within electronic databases
- The systematic scanning of reference lists of relevant key papers
- The searching of published textbooks
- The searching of relevant online sources.

2.4.2.1 Pilot search

Pilot searches were undertaken to test the effectiveness of different search strategies within electronic databases. Perhaps not surprisingly, it was established that searches that included further variations on a particular search term retrieved a larger number of relevant papers. This related particularly to the use of terms that broadened searches to include guidance on individual stages of model development, in addition to papers focused on the entire model-building process. The pilot searches also demonstrated the sensitivity of the databases to the use of particular search terms. For example, using MEDLINE®, additional papers were retrieved through including the term ‘Economics, Medical’ as an alternative to ‘health economics’.

2.4.2.2 Electronic database searching

As recommended within the CRD guidance (CRD, 2008), the search terms used in the electronic databases were determined by the research question. Given that the focus was on finding methodological rather than effectiveness papers, the widely used ‘PICO’ strategy was not considered appropriate. The research question was therefore broken down into three components: context (health economics), subject (modelling process) and area of interest (guidance). The first component related to the context of the search, i.e. the circumstances in which decision-analytic models are used, whilst the second referred to the phenomenon being studied, and the third, the particular aspect of the subject with which the research was concerned. To increase the scope of review findings, a full set of search terms were generated

as expressions, synonyms or spelling variants of the three original search components. The search strategy was implemented in the following major bibliographic databases: Ovid MEDLINE®, EMBASE, and Health Management Information Consortium (HMIC). Search terms were modified and expanded as searching progressed, relative to the contents of each individual electronic database. The searches were not restricted by date.

The ‘find citing articles’ feature of the electronic databases was used to retrieve papers that had referenced the articles found through database searching. A number of ‘key papers’ were selected for this, based on their perceived relevance to the research question. This particular element of the search was considered important, given that authors may have cited previous guidance in the development of their own guidelines. Details of the full search strategies undertaken within each database are given in Appendix 1.

2.4.2.3 Reference list searching

Reference list searching was used to identify additional papers relevant to the research question (Lefebvre et al., 2008). The process of reference list searching began with two key papers, whose reference lists were scanned for additional articles that met the inclusion criteria. Next, the reference lists of the papers identified via the key papers were reviewed, and this continued until the reference list of every article at each stage was searched, and no further relevant papers were retrieved. The key papers were recognised as such because they provided in-depth guidance on structural development (Chilcott et al., 2010; Roberts et al., 2012).

2.4.2.4 Searches in textbooks

The University of Birmingham’s library search engine was used to search for textbooks that contained guidance on decision-analytic modelling. This involved searching specifically for

modelling textbooks, but also for health economics textbooks that contained sections on modelling.

2.4.2.5 Online sources

Given the role of models in healthcare funding decisions within the UK and Canada, and the focus on these settings as the context for this research, both the NICE and CADTH websites were reviewed for their guidelines on the model-building process.

2.4.3 The selection of literature

The papers returned through the various methods of literature searching were all screened using predefined inclusion and exclusion criteria. The aim was to separate papers relevant to the research question from those that were not. The selection of literature for inclusion in the review was iterative, in the sense that the inclusion criteria were applied to the final list of papers a number of times. Initially, at the stage at which the papers were identified, their titles and abstracts were checked against the inclusion criteria with the aim of avoiding the exclusion of any literature that could potentially be relevant to the research question. However, the papers whose titles and abstracts were clearly not relevant were discarded at this point. The papers that made it through this stage were then downloaded or considered in their full-text format, for a more in-depth check against the criteria. Those included at this stage informed the results of the review.

2.4.3.1 Inclusion and exclusion criteria

The inclusion criteria were intentionally broad, accepting any article or form of literature that focused on offering some level of advice, for example, guidance, guidelines, methods, critical appraisal, or a checklist, on at least one aspect or stage of the model-building process. The exclusion criteria applied to papers that simply compared different methods for model building (without direction as to which were optimal) and those that focused only on the mathematical

construction of a decision-analytic model. For practical reasons, the exclusion criteria also extended to articles that were published in languages other than English.

2.4.3.2 Quality of the papers

Due to the economic context of the papers retrieved, it was difficult to assess their quality against the (mostly) clinical criteria outlined in the CRD (2008) guidance. However, the quality principles outlined in the CRD publication, and those recommended by GRADE (Balslem et al., 2011) were used broadly to make inferences about the standard of the studies collected, and the guidance offered within.

2.4.4 Data extraction

To facilitate data synthesis, a data extraction form was designed to retrieve and organise relevant information from each paper. The development of this form was iterative, in that the content of the papers informed the ‘themes’ that the data were extracted under. These themes referred to the subject and type of guidance represented within a particular paper which, when all of the completed forms were combined, gave an overall indication of the level and focus of the modelling guidance available. The data extraction form is available in Appendix 2.

2.4.5 Narrative synthesis

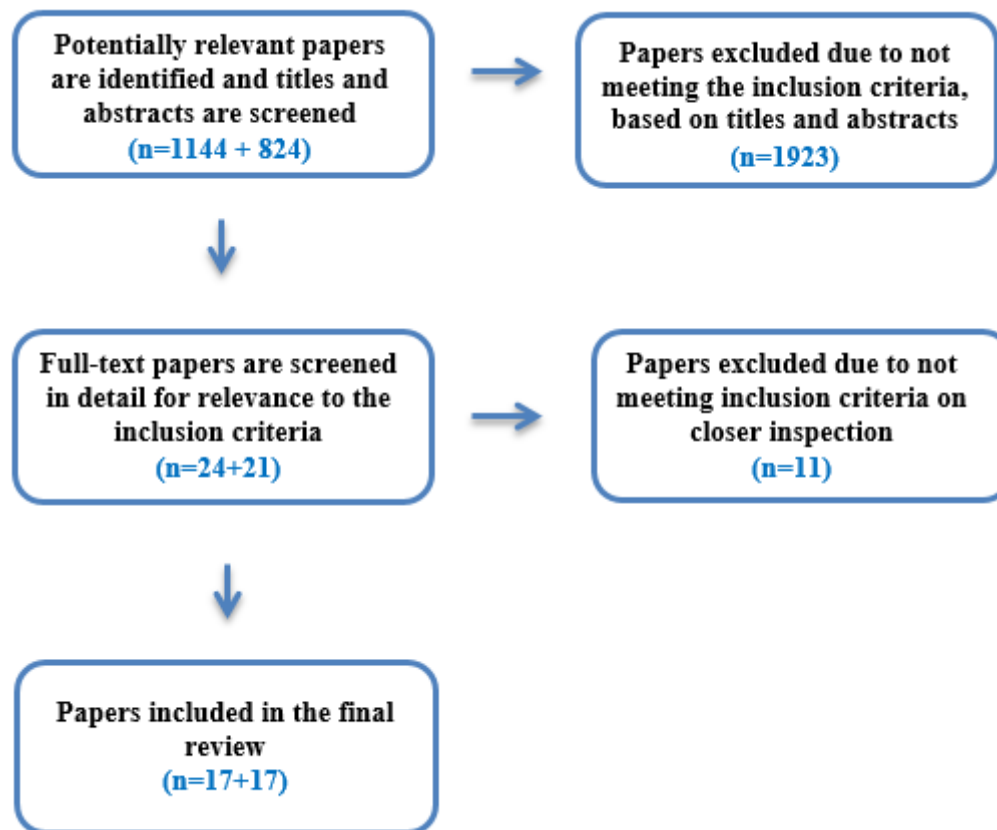
A narrative approach to the synthesis of the data was undertaken. This involved comparing and analysing text within and between papers to develop a general theory about the phenomena under study (CRD 2008). All papers retrieved were examined fully to generate an impression of the content and level of modelling guidance being offered, and then were compared to one another in terms of their similarities and differences. An iterative and qualitative approach was undertaken for the process of synthesis, as comprehensive themes of guidance were drawn and developed from the retrieved papers. The themes generated referred to five perceived stages of the model-building process. These are explored within the results section.

A research protocol detailing the methods used throughout the systematic review is available in Appendix 3.

2.4.6 Results of the review

The updated search of the modelling literature returned a total of 1968 papers for potential inclusion in the review. After paper titles and abstracts were compared against the inclusion criteria, 1923 papers were excluded. A further eleven papers were excluded on review of the full-text versions, leaving a final total of thirty-four articles for review. Of these, seventeen were recovered through searching in electronic databases, eight were retrieved through reference list searching, three were textbook excerpts identified through searches in the library database, and six were found through internet searches. Of the papers identified through online searches, four of these were specific searches for the remaining papers belonging to the ISPOR-SMDM Taskforce series, which were not identified through the electronic databases. The search process is documented in Figure 1. The figures in the QUOROM flow diagram represent the papers retrieved through electronic database searching, plus (+) those that were found via other strategies i.e. reference list searching, searches in textbooks, and online searches.

Figure 1: QUOROM flow diagram of study selection (updated review). Numbers represent the papers retrieved via database searching + those retrieved using other search strategies (Moher et al., 1999)



2.4.6.1 Characteristics of the papers

The papers collected in the review varied in terms of the date that the guidance was published, the setting in which the guidance was developed, the type of publication, and importantly, the aspect of the modelling process focused on. The earliest guidance captured was published in 1996, and the latest in 2013, with papers being produced at various times between these dates, suggesting that modelling guidance has been released and updated in line with the progression of modelling techniques. The majority of the papers were from the UK, although some were developed in international settings, and the ISPOR-SMDM taskforce guidelines combined the input of modellers from numerous countries. In terms of publication type, the majority of papers were journal articles, although one was a HTA report, three were textbook excerpts, and two were guidelines from NICE and CADTH. Two of the papers were based on

qualitative studies, and offered guidance and/or good practice recommendations based on the findings of primary research with modellers (Chilcott et al., 2010; Kaltenthaler et al., 2013).

Of the thirty-four papers retrieved, most provided guidance on more than one stage of model development, and ten of the articles were focused on the entire modelling process (Briggs et al., 2006; Caro et al., 2012; Chilcott et al., 2010; Drummond et al., 2005; Halpern et al., 1998; Petrou and Gray, 2011; Philips et al., 2006; Sculpher et al., 2000; Soto, 2002; Sun and Faunce, 2008), theoretically allowing a modeller to use the resource as an accompaniment to each stage of model building. This number increased when the ISPOR-SMDM guidance was considered as a whole, as each individual paper corresponded to a stage within the modelling process. However, the depth of the guidance provided in these papers and the others collected varied greatly, with some giving in-depth, methodological advice, but most providing less detailed recommendations.

The themes generated from the synthesis of all papers related to five perceived stages of the model-building process. These were: understanding the modelling context, structuring the model, populating the model (data), model implementation, and assessing uncertainty/model checking activities. The type of guidance offered in the papers in relation to each of these themes is examined next. Table 1 demonstrates the characteristics of the papers retrieved.

Table 1: Summary of paper characteristics

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of Paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Barton et al.	2004	UK	Journal article	Good	Individual stages	When to use modelling (context), model structure, handling uncertainty
Brennan et al.	2006	UK	Journal article	Good	Individual stages	Model structure, implementation
Briggs	2000	UK	Journal article	Good	Individual stage	Handling uncertainty
Briggs et al.	2006	UK	Textbook excerpt	Good	Process	Specifying the decision problem (context), structuring a decision model, identifying and synthesising evidence, dealing with uncertainty and heterogeneity
Briggs et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stages	Data and populating the model, dealing with uncertainty and sensitivity analysis

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of Paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Buxton et al.	1997	UK	Journal article	Good	Individual stages	Knowing how and when to model (context), data and populating the model, model validation
CADTH	2006	Canada	National guidelines – online resource	Good	Individual stages	When to use decision-analytic models (context), model structure, data and populating the model, addressing uncertainty and validation
Caro et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Process	Understanding the decision problem (context), model conceptualisation, data and populating the model, sensitivity analysis and validation
Chilcott et al.	2010	UK	Health technology assessment (HTA) report	High (qualitative)	Process	Understanding the decision problem (context), conceptual modelling, model implementation, model checking, engaging with decision

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Cooper et al.	2004	UK	Journal article	Good	Individual stages	Data and populating the model, evaluation of the model/sensitivity analysis
Cooper et al.	2007	UK	Journal article	Good	Individual stages	Developing the structure of a model
Drummond et al.	2005	UK	Textbook excerpt	Good	Process	The need for decision models (context), defining the decision problem (context), structuring a decision model, identifying and synthesising evidence, dealing with uncertainty
Eddy et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stage	Model validation

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Garrison	2003	USA	Journal article	Good	Individual stages	When to use modelling (context), handling uncertainty
Gold et al.	1996	UK	Textbook (series of chapters)	Good	Individual stages	When to use modelling (context), model structure, model validation
Halpern et al.	1998	France	Journal article	Good	Process	The need for modelling (context), understanding the study question (context), model structure, handling uncertainty, model validation
Jain et al.	2011	USA	Journal article	Good	Individual stage	Sensitivity analysis
Kaltenthaler et al.	2013	UK	Journal article	High (qualitative)	Individual stage	Data and populating the model

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Karnon and Brown	1998	UK	Journal article	Good	Individual stage	Model structure
Karnon et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stages	Model implementation
McCabe and Dixon	2000	UK	Journal article	High (review article)	Individual stages	Structure of the model, model validation
NICE	2013	UK	National Guidelines-Online Resource	Good	Individual stages	Model structure, populating the model, handling uncertainty

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Petrou and Gray	2011	UK	Journal article	Good	Process	Defining the question, structure of the model, populating the model, handling variability/uncertainty, model evaluation, model reporting
Philips et al.	2006	UK	Journal article	High (review article)	Process	Scoping the decision problem, structuring a decision model, data and populating the model, handling uncertainty
Pitman et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stages	Model structure, handling uncertainty
Roberts et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stages	Context of the model, model conceptualisation

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Sculpher et al.	2000	UK	Journal article	Good	Process	The purpose of decision models, structuring a decision model, handling uncertainty
Sendi et al.	1999	Switzerland	Journal article	Good	Individual stage	Model validity
Siebert et al.	2012	UK, Canada, USA, Australia, Austria, Sweden, Spain	Journal article	High (developed using workshops with modellers)	Individual stages	Model structure, data sources, validation and handling heterogeneity
Soto	2002	USA	Journal article	Good	Process	Understanding the decision problem (context), model structure, data and populating the model, sensitivity analysis and model validation

Author(s)	Year of publication	Country / setting	Type of publication	Quality	Scope of paper (Process/ individual stage(s))	Stage(s) of the modelling process referred to
Stahl	2008	USA	Journal article	Good	Individual stages	Framing the decision question (context), model validation/uncertainty, structuring the model
Sun and Faunce	2006	Canada	Journal article	Good	Process	The purpose of decision models (context), understanding the decision problem (context), data and populating the model, implementation, handling uncertainty
Tappenden et al.	2006	UK	Journal article	Good	Individual stages	Scoping the decision problem (context), developing a model structure, handling uncertainty
Weinstein et al.	2003	USA	Journal article	High (developed using workshops with modellers)	Individual stages	Model evaluation/validation, structure of the model, data and populating the model

Understanding the modelling context

For ten of the thirty-four papers, the initial requirement for model building was to understand whether modelling was actually needed to answer a particular research question. For example, Buxton et al. (1997, p225) argued that modelling should be used only as a ‘last resort when there is no more reliable way to provide appropriate information for decision makers’. The authors recommended the use of models only in the absence of suitable RCTs, and specifically for:

- *‘Extrapolating beyond the data observed in a trial...*
- *...Linking intermediate clinical endpoints to final outcomes...*
- *...Generalizing [findings] to other [clinical] settings...*
- *...Synthesizing head-to-head comparisons where relevant trials do not exist...*
- *...Informing decisions in the absence of hard data...’* (Buxton et al. 1997, pp. 218-222)

The guidance offered by Buxton et al., (1997) was similar to the perspective held by Gold et al., (1996) and Halpern et al. (1998, p.132), who gave the impression that ‘modelling is most appropriate for certain specified, select purposes’. A further seven papers cited one or more of the above circumstances as a reason to model; however, these were instead presented as situations in which modelling is required and valuable. The date of publications linked to their content demonstrates an apparent shift in the opinion of the literature towards the acceptance of models, with earlier papers tending to debate the use of modelling as a fundamental aspect of their content. More recent papers have tended to either advocate the advantages of modelling (Soto 2002; Garrison 2003; CADTH 2006; NICE 2013), or perceived it unnecessary to justify a model’s role in decision-making.

Very few papers offered advice on how a modeller could initially gain an understanding of the clinical and economic issues informing a model. Six of the papers gave advice similar to that of Sculpher et al. (2000, p.466), who argued that there should be ‘a clear statement of the decision problem prompting the analysis’ within a model write-up. On similar lines, both Soto (2002) and Sun and Faunce (2008) specified that an initial requirement was to define model objectives, and make clear the reasons for model development. Despite the emphasis within these papers on the importance of reporting the context for model development, none offered methods for gaining this contextual knowledge.

However, two papers did represent the modellers’ understanding of a decision problem as a distinct stage within the modelling process. Chilcott et al. (2010) and Roberts et al. (2012) emphasised a step prior to model development, which involved the modelling team learning about the clinical condition being modelled, and specifically why a cost-effectiveness analysis was needed. The papers suggested that modellers should talk to stakeholders and clinical experts, and review relevant clinical literature, and existing economic models.

‘Existing models addressing related problems should be reviewed. The clinical and policy literature describing the problem should be understood by the modeling team.’
(Roberts et al., 2012)

‘...this phase of model development may draw on evidence, including published research and clinical judgement...’ (Chilcott et al., 2010)

Structuring the model

Advice on model structure was an important focus of the guidance, with twenty-six of the thirty-four papers commenting on structure in some respect. The guidance was divided as to whether it offered methodological support or less detailed advice. A number of the papers came under the latter category and were seemingly aimed at ensuring the quality of structures and apparent robustness of structural development. Examples of advice were to make sure that models are ‘structured so that its inputs and outputs are relative to the decision-making

perspective of the economic evaluation' (Weinstein et al. 2003, p.11), 'possible pathways described by the model are feasible and sensible' (Chris McCabe and Dixon, 2000, p.508), and that model assumptions are 'clear, transparent and justified' (Philips et al. 2006, p.359). This type of guidance seems to be aimed at a model reviewer, or alternatively a modeller at the end of their model development, looking at what information to include in the model write-up.

In terms of content, nine papers referred to defining the boundaries of a model as an initial requirement of model building, involving 'making clear' a number of issues, (not limited to) 'the disease(s) or condition(s) being modelled, interventions in question, specific study populations, and study perspective' (Halpern et al. 1998, p.132). Five of the nine papers offered guidance on identifying boundaries; however, most guidance was fairly minimal, with three of the articles stating simply that the decision problem and boundaries of the model should be determined by the decision-maker/commissioner's requirements and/or the focus of the research question (CADTH, 2006; Petrou and Gray, 2011; Tappenden et al., 2006).

Drummond et al. (2005, p.290) expanded on this advice slightly by suggesting that '*decisions about the boundaries in decision models...should mainly be driven by the extent to which extending the boundaries is considered likely to impact on the cost-effectiveness of the options being compared*'. Both sets of statements left it unclear as to how making these decisions worked in practice. Roberts et al. (2012) was the only paper to offer methods for determining model boundaries, based on a process of model conceptualisation, in which all factors potentially relevant to a decision problem are first considered, before these are reduced to the elements most appropriate to the research question being asked. The authors suggested that this simplification involved consultation with experts, and diagrams to facilitate problem conceptualisation. However, the paper gave little detail on how communication should work

between the various parties, in terms of deciding on key elements of importance to inform model structure.

‘An explicit process (expert consultations, influence diagrams, concept mapping, or similar method) should be used to convert the problem conceptualization into an appropriate model structure...’ (Roberts et al., 2012).

There was also a lack of any detailed guidance on how to translate clinical and economic information into model pathways or states. Of the thirteen papers that commented on the required complexity of model structure, eleven gave guidance no more elaborate or explanatory than that offered by Soto (2002):

‘To know the level of complexity, the analyst must consider whether the model is able to capture the key issues necessary to fully describe the efficiency of all options under evaluation’ (Soto 2002, p.100)

Similarly, of the nine papers that offered advice on deciding what information should inform the basis of a structure, seven gave a similar level of guidance to that of CADTH (2006) and Sun and Faunce (2008). This guidance was vague both in terms of how modellers could learn about (the aspects important to) a clinical condition, and the practical steps involved in transferring clinical knowledge into model structure:

‘The model should incorporate all facets of the condition of interest that are important, and the potential impact of the interventions considered’ (CADTH, 2006, p.24)

‘The structure of the model is developed on the basis of an understanding of the nature of the disease progression’ (Sun and Faunce, 2008, p.314)

Only two papers discussed methods for identifying and/or selecting pathways or states for a model structure, with Roberts et al. (2012) implying that clinical experts should be consulted for their knowledge on disease progression, and Chilcott et al. (2010) suggesting that decisions around the pathways and states to include was an iterative process between modellers and clinicians. Again however, there was a lack of detail on how to involve

clinicians in structural development, for example, how many should be asked, and how to get the required information from them.

Fifteen of the papers, and most of the guidance on structure, focused on the appropriate type of model to use for a decision problem. The majority of papers commented on three main types of model, with the guidance being repetitive across papers, and authors agreeing that 'if the timeframe is short...a simple decision tree is appropriate' (Barton et al. 2004, p.111), Markov models 'are particularly effective in clinical situations which involve continuous risk over an extended time horizon' (Karnon and Brown, 1998, p.137), and discrete event simulation (DES) models should be used 'if the problem requires examining behaviour at an individual level' (Brennan et al. 2006, p.1306). This guidance varied in terms of the level of detail given, although most provided a descriptive summary of the situations in which particular structures were most suitable. Some of the papers were more methodological, with seven offering process style guidance, either on the use of specific types of structures (Karnon et al., 2012; Pitman et al., 2012; Siebert et al., 2012), or again on how to decide on a suitable structure for a particular decision scenario (Barton et al., 2004; Brennan et al., 2006; Cooper et al., 2007; Stahl, 2008). The latter group of papers all provided schematic methods for modellers to follow in selecting an appropriate structure, with a series of questions typically being posed to the reader regarding the clinical nature of their decision problem, to determine the most suitable model type.

Although collectively the guidance on choosing an appropriate model type was quite thorough, the use of the schematic tools and interpretation of the descriptive advice given relied somewhat on the modeller having a knowledge of the clinical condition and the intervention being modelled. Unfortunately, guidance on how to seek information on the clinical context of the decision problem has been demonstrated to be fairly limited.

Populating the model (data)

Twenty-one papers included guidance related to the use of data in models, specifically the data used to populate model parameters. For fifteen of these papers the main focus was on ensuring that the data sources that modellers used were transparent and of good quality. Statements included seven similar to the one given by NICE (2013, p.50), that ‘data inputs should be clearly documented and justified in the context of a valid review of alternatives’, and eleven related to the assertion by Halpern et al. (1998, p.137), that parameter values ‘should be based on the highest quality data available’. Guidance mostly focused on clinical data sources, with only four of the papers commenting on the use of economic data (NICE 2013; Halpern et al. 1998; Kaltenthaler et al. 2013; Soto 2002). Indeed, Petrou and Gray (2011, p.2) only discussed economic resources insofar as mentioning that there is not currently a clear strategy to ‘identify and synthesise evidence on other [non-clinical] variables, such as costs and health utilities’. Methodological guidance therefore applied mostly to data for clinical parameters, although some of the papers referred to principles for identifying evidence generally. For those papers that offered any methods for data collection, most were very brief, with the emphasis being on the importance of conducting systematic reviews to inform data parameters. More detailed guidance offered suggestions of data sources to use, and methods on how specific data could be retrieved (Halpern et al. 1998; Siebert et al. 2012; Soto 2002; Kaltenthaler et al. 2013; NICE 2013). Most of these papers were quite directive in their advice, with NICE (2013) specifying the data sources that they expected modellers who were submitting models for technology appraisal to include. Soto (2002) and Siebert et al. (2012) similarly made it clear that data for clinical parameters should be informed by epidemiological studies and/or data from RCTs. The study by Halpern et al. (1998) was alone in providing detailed methods for identifying appropriate economic data,

suggesting a range of sources for costs depending on the country of origin, and the standard gamble technique for deriving utility values during primary data collection.

Kaltenthaler et al. (2013) was the only paper focused solely on identifying evidence to populate models, and as such it gave the most detailed guidance on this topic. The authors gave an overview of the factors considered most important when using evidence to populate models, including an outline of review methods suitable for modelling, suggestions for minimising bias in data collection, and how to select the most appropriate data to populate parameters. However, although the guidance was methodological in content, some aspects were lacking in detail in terms of the level of advice given. For example, the authors emphasised that it was ‘important to prioritise parameters and focus reviewing resources on those most likely to have an impact on model outputs’ (Kaltenthaler et al. 2013, p.832). However, there was no clear indication within the paper as to how these important parameters could be recognised.

An additional focus of the guidance on populating a model was on the use of expert opinion to inform data parameters. Of the twenty-one papers collected, eight mentioned using clinicians to retrieve data values where evidence was lacking. Again this advice varied in terms of the level of detail given, as five of the articles offered fairly brief guidance on the involvement of experts in this context. Typically these papers implied that expert opinion should be used as a last resort (Buxton et al., 1997; CADTH, 2006; Petrou and Gray, 2011), or emphasised the importance of testing values from clinicians using sensitivity analysis (Philips et al., 2006; Weinstein et al., 2003). Of the three papers offering more in-depth guidance, Karnon et al. (2012) gave suggestions for how expert opinion used in DES models could be validated, and the remaining two articles offered relatively thorough and detailed methods for eliciting expert opinion (Soto 2002; Halpern et al. 1998), with both advising on possible ways to conduct meetings, synthesise values, and recruit appropriate clinicians. The

papers similarly recommended using a DELPHI style approach for gaining clinician estimates (allowing experts to modify values after they have been commented on by additional relevant experts), and involving experts from a range of clinical backgrounds and geographical locations (Halpern et al., 1998; Soto, 2002). Further, Halpern et al. (1998) commented on the optimum number of clinicians to consult, stating that the dynamic of involving five to eight experts would avoid discussions being dominated by a single clinician, but make it possible for all individuals to fully express their views. Interestingly, this was the type of guidance largely missing from the literature on model structure.

Implementation

Few papers recognised model implementation as a distinct stage within the modelling process. For the majority of papers offering guidance on structural development, the programming of a structure into a software platform was implied. Most focused their advice on the design of a structure, particularly on identifying an appropriate model type, without mentioning when and how the pathways or states of a model should be transferred into a computer program. One paper also discussed structural development only in the context of implementation, outlining how particular model types work and are represented within modelling software (Brennan et al., 2006). This leaves it unclear as to whether it is considered good practice to draft the structure prior to implementation (on paper for example) and then program it, or whether the structure should be drafted immediately within the software.

Only three of the twenty-six papers on structure mention (and separate) a design and implementation phase. Sun and Faunce (2008, p.314), for example, discussed ‘developing the model structure’ as a separate stage to ‘the calculation of...incremental analyses’, with the latter section containing advice such as how total costs and health outcomes are calculated within a decision tree. Similarly, Karnon et al. (2012, p.824) distinguished between guidance on designing a DES structure, particularly deciding on the downstream decisions/clinical

events to represent, and more technical advice specific to implementation, referred to as ‘transferring a defined structure into a computer program’. Chilcott et al. (2010, p.13) defined implementation as involving ‘the transposition of the pre-specified set conceptual model framework into a ‘hard’ quantitative model’, implying also that some level of design was recommended prior to implementation. However, given that this guidance was based on a qualitative study, Chilcott et al. (2010) reported that the modeller informants varied in terms of how far they drafted a structure before programming.

Arguably, the guidance as a whole would benefit from a clearer message regarding good practice in terms of the relationship between the design and implementation of a model structure. This clearly also relates to the lack of methodological advice currently available on the planning/conceptual stages of structural development, highlighted in previous sections.

Assessing uncertainty/model checking activities

Assessing uncertainty in models and/or undertaking model checking activities was the most discussed aspect of model development within the guidance, referred to in twenty-eight of the thirty-four papers. For five papers, model checking was the sole focus. Guidance was concentrated both on assessing uncertainty in model results (i.e. sensitivity analysis), and on validating the base case model.

The majority of papers, seventeen in total, focused on uncertainty in data/parameter values, and testing the effect that changes to these had on model results. These papers varied in terms of the level of detail provided, with most outlining definitions of deterministic and probabilistic sensitivity analysis. The primary message of ten of these papers was similar to Philips et al. (2006, p.361), that ‘probabilistic sensitivity analysis is the most appropriate method for handling parameter uncertainty because it facilitates the assessment of the joint effect of uncertainty over all parameters’.

Six of the papers provided more in-depth and methodological guidance on sensitivity analysis, with four of these articles being dedicated entirely to dealing with uncertainty. This guidance varied in terms of its specific focus and the level of technicality offered, with Briggs et al. (2012) for example, concentrating on how to choose parameter values for sensitivity analysis, Cooper et al. (2004) discussing how sensitivity analysis is undertaken within a Bayesian framework, and Jain et al. (2011) offering fairly technical guidance on the use of sensitivity analysis in the development of different types of models.

Only four of the papers mentioned undertaking structural sensitivity analysis, or accounting for structural uncertainty, which largely encompassed changing the structural assumptions within a model to gauge the impact on the results. For two of the articles the guidance given was rather brief, with the apparent focus being on highlighting the importance of testing for uncertainty in a structure (NICE, 2013; Petrou and Gray, 2011). The guidance offered by Pitman et al. (2012) was a little more detailed, and suggested carrying out uncertainty analyses on the structural assumptions likely to impact on the cost-effectiveness result; however, no methods for undertaking this were proposed. The paper by Briggs et al. (2012) was the only paper to provide methodological guidance on structural uncertainty, suggesting approaches that modellers could undertake.

Another lesser focus of the guidance was on model validation, and checking the base case model. Nine papers discussed the validation of a structure, data and/or the model results to some extent. Most of the guidance covered the same types of validation, although different terminology was sometimes used. The papers referred to internal validity (or technical validity), external validity, cross-validation, predictive validity, and face validity (or structural and content validity (Halpern et al., 1998)). For most articles, definitions of internal validation were similar, focused on ‘debugging’ models within computer software and checking for programming errors. The same was to be said for external validation, although the extent of

the model checking activities discussed varied between the papers. While some simply recommended that modellers compare the results of their models to those of similar models and/or clinical studies (Buxton et al., 1997; Halpern et al., 1998; Petrou and Gray, 2011; Soto, 2002; Weinstein et al., 2003), Eddy et al. (2012) provided methods on how to simulate external data sources to produce outcomes that could be used to verify those generated by a model.

In the same way as the guidance on sensitivity analysis, the validation papers varied in terms of the level of detail provided. For most who mentioned it (Chilcott et al., 2010; Eddy et al., 2012; Petrou and Gray, 2011), face validity involved the verification of the model, particularly structure, according to the feedback of those with relevant clinical expertise. Two papers suggested methods for undertaking these checks with clinicians, including emphasis on the importance of the clinical experts being external to the project (Eddy et al., 2012), and the use of modelling simulation software to allow a clinical advisor to understand how a model works, and facilitate feedback (Chilcott et al., 2010). However, there was no discussion regarding the number of clinical experts that should be consulted to ensure face validity. On the whole, and similar to the papers on sensitivity analysis, there was less guidance available on assessing uncertainty around structure, than on data inputs and results.

2.4.6.2 Principal findings of the review

This systematic review of the modelling literature has demonstrated that there is currently a lack of detailed and methodological guidance on particular stages of the model-building process. This lack of guidance relates particularly to steps involved in developing model structure, specifically, how modellers should define model boundaries and plan a structure, involve and communicate with clinicians regarding structural development, and decide on the information to inform model pathways or states. Other areas where guidance was relatively minimal included advice on how a modeller could understand and gain knowledge on the

clinical and economic context of a model, how to identify data to populate economic parameters, and how model structures should be checked/validated.

Despite the fairly large number of papers that contained guidance on the process of model building, papers were generally lacking a procedural element. Guidance tended to focus on distinct stages within the model-building process, rather than offering continuous and methodological advice. Many of the papers on structure, for example, were concentrated around recommendations on how to choose an appropriate model type, with choices being based on the modeller's knowledge of the clinical condition being modelled. However, no methods were offered within these papers as to how modellers could seek information on the clinical context of the decision problem.

The modelling guidance also generally lacked depth and detail, and was not reflective of all of the elements involved in the development of a model. Only one of the ten papers (Chilcott et al., 2010) offering guidance on the modelling process included iterations between modelling stages, which should be an inevitable part of model development, particularly if the structure is being checked at various intervals by clinical experts, and/or the structure is being updated in light of available data. Other perhaps more implicit features of the modelling process were also missing from the guidance, such as how communication between modellers and clinicians should be managed during structural development.

2.4.6.3 Quality of the papers

Although there were no relevant quality criteria available to directly assess the standard of the papers collected within this review, the guidance contained was broadly compared to the quality principles outlined in the CRD (2008) and GRADE (Balslem et al., 2011). These resources are aimed at assessing the quality of clinical research collected within a systematic review; however, general quality recommendations cited within these papers are potentially applicable.

Both favoured studies that incorporated evidence that had the least risk of bias, specifically findings based on robust primary investigation, as opposed to inferences and expert opinion. In the context of the development of model-building guidance, it may therefore be legitimate to argue that guidance based on first-hand research into the modelling process, if formulated as a result of the expertise of credible and experienced modellers, may be more valid and robust than that developed on the opinion of one or even a few authors. Indeed, the guidance by Chilcott et al. (2010) offered a more ‘truthful’ portrayal of model development, with the inclusion of iterations between stages in the modelling process. Adhering to this particular quality consideration suggests that the qualitative papers by Chilcott et al. (2010) and Kaltenthaler et al. (2013), and the guidance by Weinstein et al. (2003) and the seven ISPOR-SMDM papers, which were developed using working groups comprised of academic and industry modelling experts, should be considered of higher quality than others. Additionally, and based on frameworks for assessing hierarchies of evidence (Evans, 2003), one could also argue that systematic reviews of guidelines (McCabe and Dixon, 2000; Philips et al., 2006) are more robust than those created by only one set of authors.

The drawback of using this quality criteria is that it does not necessarily account for the usefulness of guidance (i.e. the level of depth and detail provided). One also cannot be sure of the methods used to develop guidance within the remaining papers. However, given that judgement around the quality of research relies somewhat on the quality of reporting (CRD 2008), perhaps authors of modelling papers should be encouraged to be more explicit about the methods that they are using to make suggestions around good practice in model development.

2.4.6.4 Other systematic reviews in the research area

In updating the original review undertaken in 2012, a further relevant and recent systematic review was identified (Penaloza Ramos et al., 2015). This review had similar objectives to those outlined within this research, although the focus of this alternative study was largely on

comparing the content and recommendations of the guidance for similarities and differences. The authors focused on highlighting any conflicting advice that was given across the guidance, implying that these aspects would require further clarification. The areas in which the guidance was apparently inconsistent referred to the methods used to check for structural uncertainty, the methods for performing predictive validity checks, and whether model structure should be affected by the availability of data. Although Penaloza Ramos et al. (2015) have a valid point in arguing that discrepancy within guidance can lead to uncertainty around what constitutes good practice, the review undertaken within this chapter has largely concentrated on identifying aspects of model development where there was little or no guidance available for modellers. Arguably, errors and poor practice are more likely to occur within stages of the modelling process for which there is no guidance available, providing that modellers are using modelling guidance. Developing recommendations in underrepresented areas may thus be a priority. However, both studies identified similar areas of the guidance to be 'problematic', specifically uncertainty (lack of guidance) around the information that should be used to inform model structure, and how model structure should be checked.

2.4.7 Conclusion

This chapter has highlighted the nature of errors present in industry submission and published models, and has presented a systematic review of the model-building guidance available. The first half of the chapter found that methodological errors identified across the models explored were similar, based around structure and the decisions made by modellers on what to include. Interestingly, and potentially relevant, were the findings of the systematic review, which revealed that very little in-depth and detailed guidance existed for particular aspects of structural development. This included guidance on planning model structure, the involvement of clinical experts in structural development, and the translation of economic and clinical information into model health states and pathways. The guidance was also lacking in advice

on how modellers could understand the context of a model, how to identify economic parameter data, and how to check model structure. This suggests that future research and guidance might be required in these areas.

However, one cannot be sure from the findings of the systematic review that modellers are experiencing problems with these aspects of model development, although the methodological errors highlighted within the reviewed models might be a good indication. Perhaps more importantly, it was not possible to identify whether there are any other aspects of model building that modellers are having difficulty undertaking, or where less robust practices are being used. Also, it is unclear why methodological errors are present within models, and whether all modellers are paying attention to the guidance available. Further research is needed to explore these issues.

The systematic review captured a small number of papers that had used qualitative research to develop model-building guidance. This guidance was deemed to be of greater quality than that offered in other papers, as advice was generated on the basis of first-hand research into the methods used by multiple modellers, rather than on the opinions of a few individuals. It is possible therefore, that qualitative research could be used to investigate the current practices and opinions of modellers, and to understand where and why problems are occurring, and potentially how these could be addressed. The next chapter looks at other examples of how qualitative methods have been used to understand and potentially suggest improvements to methodological issues present within health research processes, to see whether qualitative research might be a good option to undertake further exploration of model building.

CHAPTER 3: THE USE OF QUALITATIVE METHODS TO UNDERSTAND METHODOLOGICAL ISSUES IN HEALTH RESEARCH

3.1 Introduction

The previous chapter reported on methodological errors identified in industry submission and published models. In addition, the systematic review undertaken within the chapter found a lack of in-depth guidance on various aspects of the modelling process, particularly model structure. These findings suggest that modellers are encountering problems in their model building that could, without awareness and appropriate guidance for the modeller, potentially lead to sub-optimal recommendations on cost-effectiveness. Arguably, therefore, more in-depth research is required to explore the processes used by modellers, to identify good practice and to understand where and why problems are occurring. Qualitative research could potentially be used to explore these issues, and generate recommendations for model development. The current chapter therefore seeks to establish whether qualitative methods can provide a suitable and fruitful basis on which to investigate methodological problems in the modelling process.

The first part of this chapter focuses on exploring the methodological basis for qualitative research, including how qualitative research is used, the theoretical assumptions associated, and how the natures of qualitative and quantitative inquiry differ. Next, the chapter explores how qualitative research has been undertaken to understand, and potentially provide solutions to, methodological issues in other areas of health research, in order to learn from the approaches used. Finally, the chapter concludes with a systematic review of the papers that use qualitative methods to explore decision-analytic model development.

3.2 Qualitative research

Owing to its use across multiple disciplines and traditions, fields and subject matter, there is no single definition of qualitative research (Denzin and Lincoln, 1998; Snape and Spencer, 2003). Denzin and Lincoln (1998, p.2) argue that ‘qualitative research is a field of inquiry in its own right’, and thus encompasses a wide range of approaches to data collection and analysis. However, most qualitative inquiry appears to have a common interest, in seeking to understand and interpret phenomena in their natural context, particularly in terms of the meanings that people assign to them (Denzin and Lincoln, 1998; Malterud, 2001; Miller and Crabtree, 1999). Flick et al. (2004, p.3) suggest that qualitative research is concerned with gaining a deeper ‘understanding of social realities’ from the perspective of those ‘who participate...to draw attention to processes, meaning patterns and structural features’.

Therefore the emphasis of qualitative methods is often on ‘capturing the individual’s point of view’ and ‘seeking rich descriptions of [their] social world’ (Denzin and Lincoln, 2000, pp.9-10). Furthermore, qualitative analysis, whilst aiming to provide explanations for phenomena, also seeks to preserve the context and detail of the social world being explored (Bryman, 2008).

Qualitative inquiry typically takes place within a particular interpretive paradigm (Crotty, 1998; Denzin and Lincoln, 2000; Guba, 1990; Silverman, 2000). Guba (1990, p.17) defines a paradigm as a ‘basic set of beliefs that guide action’ and indeed Crotty (1998) argues that a researcher’s theoretical perspective directs their approach to research, and the methods that they use. Each paradigm has its own ontological, epistemological and methodological position, which in turn dictates how ‘reality’ is defined, the relationship that the researcher has with informants, and the strategy used to collect and analyse data. Denzin and Lincoln (2000) cite the four major interpretive paradigms underlying qualitative research as positivist and postpositivist, constructionist-interpretive, critical, and feminist-poststructural. Regardless of the belief systems and methods typically adopted within these paradigms however, Denzin and Lincoln (2000, p.8) argue that all qualitative research ‘implies an emphasis on the qualities of entities...that are not...measured in terms of quantity, amount, intensity, or frequency’.

Qualitative research is often understood and discussed in relation to the objectives and methods associated with quantitative approaches. Green and Thorogood (2009, p.5) argue that ‘qualitative studies seek answers to questions about the ‘what’, ‘how’ or ‘why’ of a phenomenon, rather than questions about ‘how many’ or ‘how much’’. Therefore, methods such as detailed interviewing and observation, rather than surveys, structured observation and experiments, are used within qualitative research to capture naturally occurring data, and the depth and nuances associated with social ‘reality’ (Silverman, 2000).

The benefits and drawbacks of quantitative and qualitative research are also frequently considered in relation to one another. For example, the reductionist and deductive nature of quantitative research is often seen as advantageous, as conclusions are drawn from general theories occurring in data, and findings are thought of as broad and generalizable (Patton, 2002). This is in relation to the inductive nature of qualitative research, which involves ‘the

formation of theories, concepts and types' from what is 'observed' in the field (Flick et al., 2004). Polit and Beck (2010, p.1452) state that qualitative findings are often critiqued for seemingly being specific to only one particular 'human experience', and biased towards the pre-existing views and experiences of the researcher (Dancy, 2001). However, qualitative researchers dispute the above claims, and Guba and Lincoln (1998, p.197) argue that rather than being objective and generalisable, quantitative research is 'context stripping' and 'exclusionary', as research questions typically focus on identifying patterns between pre-selected 'subsets of variables', not exploring 'other variables...that might...greatly alter findings'. Not only do qualitative researchers consider detailed investigation and rich description to be valuable, it is suggested that qualitative research should not be concerned with achieving statistical generalisability (Mays and Pope, 2000).

Polit and Beck (2010, p.1454) recommend that researchers implement strategies to enhance the generalisability and transferability of qualitative research conclusions. Such strategies might include techniques such as purposive sampling, which allows a researcher to capture a variety of perspectives on a phenomenon, and explore whether conclusions are replicated across informants. Transparency, in relation to the researcher providing sufficient detail on informants and phenomenon, can also allow the reader to make judgements about how valid research conclusions appear in relation to the data, as well as how transferable findings are to similar settings (BMJ, 2015; Mays and Pope, 2000). Other quality considerations involve triangulation, the use of multiple methods to strengthen research conclusions (Mays and Pope, 2000), and the importance of reflexivity during the qualitative process (Saukko, 2003), in terms of researchers recognising and acknowledging how their 'cultural baggage', such as their personal and intellectual background, might influence findings. Denzin and Lincoln (1998) argue that 'value-free' qualitative inquiry is not possible, meaning that researchers

must be open and reflective about their contributions to research conclusions, to ensure credible practice (Malterud, 2001).

3.3 The use of qualitative methods in health research

Qualitative methods are increasingly being used to explore phenomena within a health research, or health services research context (Mays and Pope, 2000). Health services research is a multidisciplinary field concerned with access to, and delivery of, healthcare (Phillips, 2006), via the study of ‘the organisation and culture of those who provide [it]’ (Pope and Mays, 1995, p.1). Gagliardi and Dobrow (2011, p.1) suggest that the ability of qualitative research to ‘produce rich data on perceptions, beliefs, experiences and behaviour to create a thorough understanding of a problem, and how it could be resolved’ has led to its application to ‘improve health service delivery [in] a variety of clinical... settings’. The emphasis of qualitative research on process (Bryman 2008) seems to fit well with the objective of health services research, which is to identify where healthcare processes can be improved, to make them more ‘effective, equitable, and efficient’ (Phillips, 2006, p.2).

Qualitative research has the potential to be as useful in the field of health economics.

Regarded as a component of health services research, health economics aims to improve healthcare processes, in terms of analysing ‘decision making by individuals, health care providers and governments with respect to [the allocation of] health and health care’ (Morris et al. 2007, p.2). Qualitative research has, for example, been used to explore the role of economic analysis in NICE’s technology appraisal process (Williams et al., 2007). However, some authors have reported on the dearth of qualitative research being undertaken within this field (Coast, 1999; Coast et al., 2004; Obermann et al., 2013), with Coast (1999) suggesting that this is due to tension between the philosophical positions of the discipline of economics, and the qualitative approach. Whilst mainstream economics is situated within a positivist

paradigm and advocates a single reality that can be studied objectively for universal truths, qualitative research typically adopts a constructivist position, which assumes multiple context-bound realities that can only be explored through interaction between ‘the investigator and investigated’ (Coast, 1999, p.347). A number of authors however, have noted the increasing use of qualitative research in health economics, albeit typically alongside the use of a quantitative approach (Coast et al., 2004; Obermann et al., 2013).

The next sections of this chapter explore how qualitative methods have been used in different areas of health research to understand methodological issues. This is in line with the objective of this research, which is to explore modelling processes and identify how and why problems and errors are occurring. The two areas of research selected for investigation were studies using qualitative research to understand how people complete valuation tasks to inform economic evaluation, and studies using qualitative methods alongside randomised controlled trials (RCTs) to identify problems with recruitment. These bodies of literature were chosen because of their similarity to the research proposed within this thesis, in that both areas of research are based on pre-existing concerns regarding the appropriateness of current methods, and aim to explore processes that contribute to decision-making within healthcare.

The searches of the literature undertaken for these papers were not systematic, but were broad, and aimed to identify a range of papers within each research area that had used different qualitative methods to address similar aims. The objective of the review of these papers in relation to the research proposed within this thesis was twofold: (1) to establish whether qualitative research would be useful and suitable for the exploration of model-building processes, and (2) whether any particular qualitative methods or approaches appeared optimum compared to others.

3.3.1 The use of qualitative research to understand how people complete preference-based valuation tasks

Although qualitative methods have been used more extensively in the development of attributes for preference-based measures (see Coast and Horrocks, 2007; Fegert et al., 2011; Ke et al., 2013; Miners et al., 2012), the focus of this review is on the smaller number of studies using qualitative methods to explore how people complete preference-based valuation exercises. Briefly, these exercises are designed to elicit people's preferences for particular healthcare goods or states, to use in economic evaluations to assist resource allocation decisions. These exercises or techniques include the visual analogue scale (VAS) standard gamble (SG), person trade-off (PTO), time trade-off (TTO), willingness to pay (WTP) and discrete choice experiments (DCEs). A total of eight papers were reviewed.

The studies captured had different aims but similar overall objectives, which concerned establishing whether current methods for eliciting preferences were appropriate, and whether the ways in which people value healthcare were in line with the economic theory underlying the measures. These studies were mostly undertaken out of concern that people's valuations of healthcare were not based upon their true and underlying preferences. Inevitably, the qualitative research was undertaken alongside quantitative research, as preference-based measures typically use quantitative techniques to calculate the values that individuals assign to healthcare goods (Morris et al., 2007). However, in line with the comments of Coast et al. (2004) and Obermann et al. (2013), almost half of these studies used qualitative research only to explore or explain quantitative findings. Studies also ranged from those undertaking deductive qualitative analysis, examining existing or pre-specified issues, to those exploring valuation processes inductively. Despite this, all of the papers collected had the same objective, namely to gain an understanding of how people value health states and/or goods.

Given the focus on learning about the type of findings that different qualitative methods can generate, the papers are compared according to the methods used.

The qualitative methods undertaken across the studies comprised ‘open-ended’ survey questions, semi-structured interviews, think-aloud, and the triangulation of these approaches. The qualitative survey was only used in one study, which aimed to investigate whether informants gave rational and valid responses within valuation exercises (Miguel et al., 2005). Qualitative analysis was undertaken alongside quantitative rationality tests of informants’ answers to a DCE questionnaire. Where informants failed the quantitative rationality tests they were asked to explain their answers on paper. The qualitative findings indicated that the informants often had rational reasons for their ‘irrational’ preferences, with the implication of the research being that those who would typically be ‘dropped’ from stated preference studies were actually offering meaningful answers. In terms of methods, however, the limited space that the authors allowed informants to explain their answers in the survey meant that around half could not be analysed and/or understood due to the lack of depth provided. This raised the question of how valid the authors’ interpretations of the informants’ reasoning was, given the lack of detail in their responses.

Semi-structured interviews were undertaken in two of the studies (Geneau et al., 2008; Robinson and Bryan, 2013). These studies aimed to ascertain whether informants’ preferences satisfied the axioms of expected utility theory on which WTP and PTO measures are based i.e. that people always act to maximise utility for themselves and thus have rational and complete preferences for particular goods. In both studies the authors reflected on their use of semi-structured interviews, and concluded that the data gathered were lacking in terms of what was required to answer the research question. Geneau et al. (2008) found that the limited length of interviews affected an informant’s ability to give a nuanced answer about the factors contributing to their WTP values. The authors implied that being unable to engage with and

allow informants time to consider how they valued a healthcare good, led informants to value goods according to their ability to pay.

Robinson and Bryan (2013) aimed to look at whether deliberation during the valuation process using PTO elicited more ‘considered’ preferences from informants. The authors stated that deliberation could help informants to fully reflect on and construct their preferences during the valuation task, but could also possibly lead to their valuations being influenced by the opinions of others, meaning that the PTO exercise was not capturing valid choices. The authors recorded examples of informants changing their valuations after deliberation, but their comments in interviews suggested that individuals considered discussion with others useful, but not influential, in realising their own preferences. However, because no observational analysis was undertaken on the deliberation process, the authors could not definitively conclude that informants’ valuations were not influenced by other members of the group, despite informants suggesting otherwise. This highlights the importance of using observational methods where communication is an important aspect of study, given that informants might not be aware of, or articulate, the effect that particular phenomena have on their behaviour. Both of these papers generated theoretical conclusions, in finding that preferences did not satisfy utility theory, however, each also had practical implications, with Robinson and Bryan (2013) stating that future research was needed into how researchers can use deliberation more optimally in the elicitation of people’s preferences.

Think-aloud was the most commonly used method among the studies (Ryan et al., 2009; Smith, 2007; Van Osch and Stiggelbout, 2008), with a further two papers using think-aloud alongside interviews (Baker and Robinson, 2004; Robinson et al., 1997). Within these studies informants were asked to verbalise their thought processes as they valued healthcare goods, generating rich and explanatory findings about how people make decisions during valuation tasks. Van Osch and Stiggelbout (2008) reflected on their use of think-aloud, stating that the

method had enabled them to explore bias in informants' responses to SG exercises, generating an in-depth understanding of informants' thought processes as they valued health states, and allowing the authors to identify the reference point from which people focused their SG valuations, and thus control for the impact of bias in SG results.

Despite the benefits associated with think-aloud, two studies reported that the use of this method may have biased study results, by encouraging informants to think more thoroughly about their preferences than they would have done otherwise. Smith (2007) found that responses at higher levels of WTP had higher rates of reliability, suggesting that informants gave more considered and valid responses where decisions impacted significantly on their income. However, the authors could not be sure whether more reliable responses were recorded because informants had been asked to verbalise their thought processes during the WTP exercise, facilitating their recall of previous answers during tests of reliability.

Similarly, Ryan et al. (2009, p.333) suggested that results indicating that informants held complete and well-defined preferences, may have been as a result of informants being 'more careful about, and conscious of, their thinking' and decisions whilst 'thinking aloud'.

The remaining two studies used a triangulation of think-aloud and interview methods (Baker and Robinson, 2004; Robinson et al., 1997). The authors used these in combination to different ends, as Robinson et al. (1997) investigated pre-specified differences between people's valuations using VAS and TTO techniques, and Baker and Robinson (2004) used an inductive approach, aimed at understanding more broadly how individuals completed SG valuation tasks. Both Robinson et al. (1997) and Baker and Robinson (2004) were able to fulfil the aims of their research via the approaches selected, as the former authors asked specific questions of informants in relation to differences in their TTO and VAS valuations, and the latter authors generated numerous themes around the factors that people considered when making their valuations.

Reflecting on their methods, Baker and Robinson (2004) reported that their use of semi-structured interviews as a follow-up to think-aloud was essential in collecting in-depth data, as informants sometimes found it difficult to verbalise all of their thoughts while completing valuation tasks. In this sense the shortcomings of one method were supported by the use of another. Robinson et al. (1997) commented that using structured interviews alongside think-aloud enabled them to generate data on issues of interest more efficiently. However, the impression was that think-aloud and questioning were undertaken simultaneously, with informants being asked specific questions about their responses at certain points in the exercise, potentially disrupting the flow of their verbalisations. Further, although informants were asked open questions about their responses to the valuation exercise, informants' answers were not then probed for detail. This deductive style approach may provide an explanation for why the authors struggled to explain certain findings, such as why older informants gave lower TTO valuations than younger informants. It is possible that the reasons underlying people's valuation of health states were complex and personal, suggesting that data collection may have benefited from further time and engagement between the researcher and informant.

The review of this literature has demonstrated how qualitative methods have been used to understand and explore preference-based valuation processes. Qualitative analysis was largely undertaken to identify issues related to the way in which people constructed and delivered their preferences, such as whether preferences adhered to the axioms of utility theory. This has emphasised the potential for qualitative research to be used in a similar capacity to investigate model-building processes. The underlying context of the papers reviewed is similar to that of the modelling research proposed, as qualitative methods would be used to understand and explore the processes that modellers are using to develop decision-analytic models, particularly any issues with the methods undertaken. Collectively, the preference-

based valuation studies generated a range of theoretical but also practical findings, including suggestions for enhancing valuation processes. These practical implications are indicative of those that this modelling research is aiming to generate.

Although all studies addressed the objectives of their research using different approaches, lessons were learnt about the strengths and limitations of the methods used. The authors of studies using surveys and semi-structured interviews were unable to capture enough detail on relevant issues because of the less exploratory nature of these methods. Robinson and Bryan (2013) noted that interviews were less useful where communication between informants was important, suggesting that the observation of informants' discussions and deliberations would have allowed for a deeper understanding of their research conclusions. Think-aloud was reflected on as a useful method for gaining a detailed understanding of informant thought processes, and thus the way in which informants carried out tasks. However, it was also suggested that informants' verbalisations of their thought processes may have directly influenced research conclusions, by encouraging informants to think more deeply about the exercises that they were undertaking. This emphasised the importance of choosing an appropriate research design when using qualitative methods. Finally, the use of the triangulation of qualitative methods appeared valuable, as using multiple methods allowed the researchers to capture additional detail on important issues. It seemed apparent therefore, that an inductive and exploratory approach to qualitative study was that which was most beneficial.

3.3.2 The use of qualitative research to improve recruitment to randomised controlled trials

Although qualitative research has been used extensively in health services research, for example in the development and evaluation of healthcare interventions (see Hart et al., 2005;

Schulze and Angermeyer, 2003), this review focuses specifically on how qualitative methods have been used to understand RCT recruitment. There is a growing body of health services research that focuses on using qualitative approaches to identify potential barriers to RCT uptake, in the context of patient recruitment to trials being challenging, and patient numbers being lower than required. A search of the literature was undertaken to retrieve examples of these papers, with a total of ten being discussed.

Most studies captured through the search were undertaken as part of an ongoing RCT. The studies used qualitative research to explore recruitment processes, with the aim of identifying factors that facilitated or prevented uptake to RCTs, and generating suggestions for improvements to recruitment processes generally. Although these recruitment papers had similar objectives to one another, different qualitative methods were used, including semi-structured interviews, in-depth interviews, focus groups, observation, and the triangulation of these methods. As with the previous review, the studies are compared according to the methods undertaken.

Two studies used semi-structured interviews (Eborall et al., 2011; Gopinath et al., 2013) to identify reasons for non-participation in RCTs from the perspective of patients who decided not to participate. Both studies generated findings on barriers to recruitment and suggestions to improve uptake. However, the deductive approach adopted by Gopinath et al. (2013) limited research conclusions, as the paper focused on exploring pre-existing barriers to recruitment, producing findings on patient treatment preferences that only reaffirmed those of previous studies. The focus on confirming rather than generating theory also inhibited exploration of patients' feelings around the issue of treatment preferences, as reasoning was explored in limited depth, and preferences were assumed to be 'non-modifiable'. Eborall et al. (2011) generated additional themes on barriers to recruitment through a more inductive approach, finding too that patient treatment preferences were responsible for non-participation

in RCTs. However, the authors asked informants closed-ended questions, questioning informants directly about their thoughts on potentially participating in the trial and not knowing which treatment they would receive. It is possible therefore that the style of questioning adopted may have influenced the themes and outcomes of this research.

Studies using in-depth interviews benefited from informants' detailed and lengthy discussions on recruitment. Das Nair et al. (2014) used focus groups to capture informants' views on the recruitment process planned for an upcoming RCT. The study found altruism to be a motivator for trial participation and thus the findings of informants' detailed discussions were used to update patient information resources to include the benefits to others of medical research. Two papers based on the same qualitative studies used in-depth interviews to understand recruiters' involvement in, and experiences of, recruitment (Donovan, Paramasivan, et al. 2014; Donovan, de Salis, et al. 2014). Both studies found that encouraging recruiters to speak in-depth and personally about their roles led to discussions over feelings of discomfort around recruiting patients. The benefits of an exploratory approach to research were demonstrated, as informants did not immediately recognise these feelings as potential barriers to recruitment. When asked directly about issues around recruitment, informants only identified organisational difficulties and patient treatment preferences as problematic. The studies determined that recruiters required further training and support to improve recruitment, aimed at helping recruiters to engage more comfortably with potential participants, and discuss trial participation more confidently.

Studies by Wade et al. (2009) and Mills et al. (2011) used observation of recruitment appointments (recordings) to gain insight into interaction between recruiters and patients. Wade et al. (2009) used conversation analysis to investigate whether patients appeared to understand recruitment information. In comparing conversational exchanges, the authors found the most 'successful' appointments in terms of uptake were those that allowed patients

to express their views and address participation concerns. The researchers observed patients moving from having set preferences for a particular treatment, to consenting to a trial and randomisation after this style of discussion. Mills et al. (2011) aimed to explore how patients communicated treatment preferences, finding that patients expressed strengths of preference, in that whilst some individuals were sure of the treatment that they wanted, others were less convinced, and changed their minds after detailed discussions with recruitment staff. The findings of the above studies were in contrast to those of Gopinath et al. (2013), who concluded that treatment preferences could not be changed, but who also did not capture the nuances associated with patients' opinions on recruitment. Both observational studies suggested that RCT uptake could be enhanced through improvements to communication with patients.

The remaining papers used a triangulation of methods. Through the use of in-depth interviews and observation of recruitment appointments, Paramasivan et al. (2011) identified a number of recruitment challenges, including recruiter difficulty with explaining trial design to patients, due to the complexity of the information involved. The recruiters interviewed reported such problems themselves, and were also observed struggling to convey recruitment information to patients concisely. These collective findings provided increased justification for clearer patient information within the trial. Hamilton et al. (2013) used semi-structured interviews with recruiters and observation of recruitment appointments. In contrast to the previous study, the triangulation of methods generated conflicting findings, in terms of what was observed as problematic during observation of appointments, and the reasons that recruiters gave for poor recruitment in their interviews. The authors found that recruiters were reinforcing their own perceptions of why patient uptake was low onto patients during recruitment consultations. The recruiters expressed the view that patients considered one treatment option to be less acceptable than the other, but were actually emphasising this

stance to patients, thus not presenting treatment options in an equivalent way. These findings highlight the potential for alternative qualitative methods to capture different interpretations of the same phenomenon.

Donovan et al. (2002) used semi-structured interviews with patients, and observation of recruitment appointments to explore how patients interpreted study information. Patients were asked to discuss their experiences of recruitment, and the authors ‘matched’ patient comments to the information delivered during RCT consultations. The data generated gave the authors some indication of why patients held particular views on recruitment, such as patient concern with the ‘conservative monitoring’ arm of the trial, which was presented as an inactive option compared to others, and thus perceived by patients to mean ‘no treatment’. The use of interviews and observation in conjunction allowed the authors to explore the findings of one method, using the other. All of the above papers and those using observational methods have highlighted the additional insight that the study of communication in the RCT recruitment process was able to provide.

The review of this literature has demonstrated how qualitative methods have been used to understand recruitment issues in RCTs, and suggest ways to improve uptake. Again this provides support for the use of qualitative methods to understand and identify improvements to modelling processes, by speaking to and observing those involved. A number of the recruitment studies emphasised the importance of using observation to capture communication between patients and recruiters, and in terms of the implication of this research for the modelling work proposed, it is probable that communication between those involved in model development could have a similarly important effect on the modelling process and final model.

Arguably, and similar to the conclusions of the preference-based valuation studies, the RCT papers using methods that allowed for deeper exploration of informants’ feelings and the

recruitment context, provided the opportunity to investigate the more nuanced but important issues associated with RCT uptake. Further, the studies using triangulation highlighted the benefits of using multiple qualitative methods, in generating support for research findings, but also different insights into a particular phenomenon.

3.3.3 The use of qualitative research to explore decision-analytic model development

A systematic review was undertaken to explore how qualitative methods have been used to understand decision-analytic model development. This review was important in identifying similar work to that proposed within this thesis, and therefore highlighting whether further research is required and justified. The next sections describe the methods used within the systematic review.

3.3.3.1 Review question

How have qualitative methods been used to understand and/or improve health economic modelling?

3.3.3.2 The search strategy

The systematic review was undertaken in line with guidelines from the CRD (2008). Prior to undertaking the systematic review, a search was carried out in the Cochrane Database of Systematic Reviews, to check that there were no existing reviews on the use of qualitative methods in model building. Despite searches of the database using a number of different search terms, no relevant reviews were retrieved. Next, a full systematic review was undertaken, involving a number of stages and approaches to identify relevant material and ensure thoroughness. Wider searches are advised to identify different types of relevant literature, and to avoid publication bias (CRD, 2008). The search involved the following stages:

- The searching of major electronic bibliographic databases for potentially relevant papers
- Forward citation searching within electronic databases
- The systematic scanning of the reference lists of key papers
- Online searches using Google Scholar.

The electronic bibliographic databases searched were MEDLINE®, EMBASE and HMIC. A pilot search was undertaken within these databases to test the sensitivity of search terms in capturing papers relevant to the research question. The initial search indicated that terms had to be broad, but specific in relation to the context of the papers being searched for as, when ‘qualitative’ and ‘modelling’ were searched for in combination, many irrelevant papers were returned. Therefore, the format followed in the main search focused on dividing the research question into three components: context (health economics), subject (modelling) and methodology (qualitative research). These components were expanded on during the review process to include synonyms and variations of the original search terms. The search strategies tested and undertaken as part of the review process are recorded in Appendix 4.

Forward citation searching was used to find papers that had referenced those identified through electronic database searching. This was undertaken using the ‘find citing articles’ feature available in MEDLINE®. This approach to identifying relevant papers was important, given that authors may have cited others’ work to justify the methodology that they were using in their own. The reference lists of all papers retrieved through electronic database or forward citation searching were also thoroughly searched for other relevant research. These approaches were used in an iterative manner, in that at the point at which a key paper was identified, forward citation and reference list searching was undertaken on the publication.

Finally, online searches were undertaken in Google Scholar, to identify any further and relevant work by key authors and by using key search terms.

3.3.3.3 The selection of literature

All relevant material was considered for inclusion in the review, including full papers, reports, conference abstracts and theses. Where papers or material were not accessible through the electronic databases or online searches, authors were contacted for a copy. All material returned through the various search strategies was screened using pre-defined inclusion and exclusion criteria. These criteria were intentionally broad so as to attract as many relevant papers as possible. The exclusion criterion was implemented for practical reasons.

- **Inclusion criteria:** Any form of literature that has used qualitative methods to investigate the elements or processes involved in the development of health economic models.
- **Exclusion criterion:** Non-English language papers.

The selection of papers for the final review was iterative, in the sense that the collected material was reviewed a number of times against the inclusion criteria. At the point at which papers were initially identified, their title and abstracts were checked for eligibility, with only material that was clearly not relevant being excluded. Material that appeared to be related to the research question then progressed to the second stage of the screening process, where it was reviewed in full-text format. Papers that were not excluded at this point were included in the final review.

3.3.3.3.1 *Quality of papers*

The CRD (2008) state that papers collected in a review should be assessed for quality. As this review was focused on methodology, studies could be critically assessed according to how

appropriate and robust their uses of qualitative methods appeared. Papers were compared broadly against the relevant criteria outlined in the British Medical Journal (BMJ) editor's qualitative research checklist (BMJ, 2015).

3.3.3.4 Data extraction and synthesis

To facilitate data analysis and synthesis, a data extraction form was developed to record the details of the papers retrieved. The information taken from the papers was that which was considered most important to the research question, namely the aims of the study, the qualitative methods used, and the nature of the findings that the author(s) had generated. A copy of the data extraction form is given in Appendix 5. Narrative synthesis was used to compare the key components of the papers. The research protocol in Appendix 6 details all of the methods used within the review.

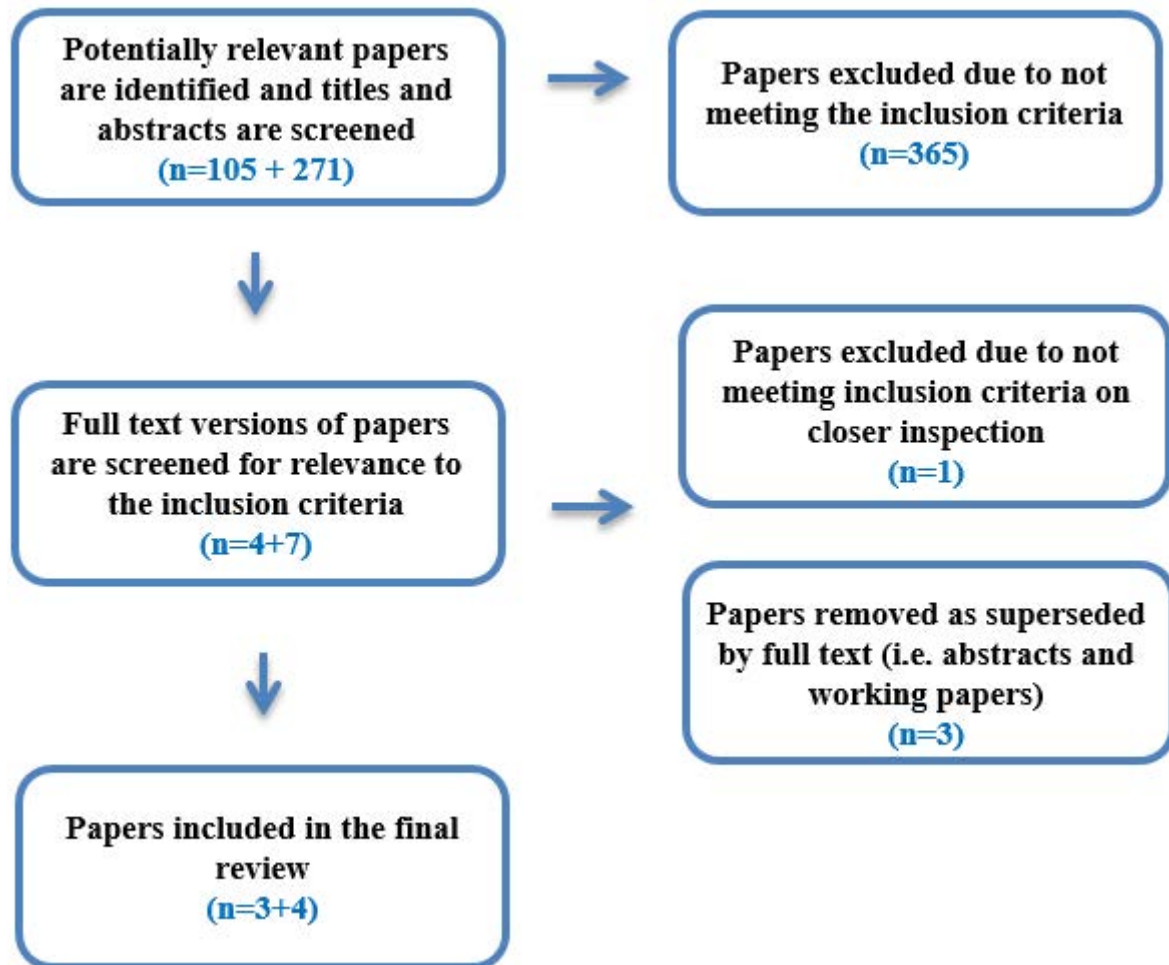
3.3.3.5 Results of the review

In total, 376 papers were identified for potential inclusion in the review. After screening of the titles and abstracts of these papers, 365 were excluded for not meeting the inclusion criteria. A large number of papers were removed at this stage, given that it was easy to see from the methods section of the abstracts whether, or in what context, the authors had used qualitative research. Eleven papers reached the second stage of the screening process, which involved evaluating the text of the papers to check that their content met the inclusion criteria. One paper was excluded at this point, because, although the paper was focused on identifying errors in model building, the authors had not used qualitative methods, but had only cited others' uses of them. Three further papers were removed for being earlier versions of full-text papers also captured in the review (i.e. abstracts and unpublished discussion papers). Given that these papers were referring to the same research study as the published full-text versions,

and did not contribute any additional information, they were excluded for ease and readability.

The systematic review therefore retrieved seven relevant papers. Of these, three were found through searching the electronic databases, one was retrieved through online searches using Google Scholar, and two were identified using reference list searching. The remaining paper (Squires, 2014) was retrieved through contacting the author directly, as a conference abstract of her PhD thesis was identified in a Google Scholar search, which then prompted a request for her full work. Although identified via a reference list search, Paisley (2012) also had to be contacted for her PhD thesis. The search process is demonstrated in the QUOROM flow diagram represented in Figure 2.

Figure 2: QUOROM flow diagram of study selection (Moher et al., 1999). Numbers represent the papers retrieved via database searching + those retrieved using other search strategies



The numbers in each of the boxes of the flow diagram represent the papers retrieved through electronic database searching, plus (+) those that were found via other strategies i.e. reference list searching, searches in textbooks, and online searches. A full version of the systematic review is given in Appendix 4.

3.3.3.5.1 Characteristics of the papers

Of the seven papers retrieved through the review, five were full-text papers and two were PhD theses. The papers included the work of only four lead authors, with four of the papers being written by the same author, and all papers originating from the same UK University. Further, Kaltenthaler et al., (2014, 2013) were papers derived from the same qualitative study. This all

suggests that there has been minimal work using qualitative methods to explore modelling processes. In the next sections, the papers are compared according to their aims, methods and the nature of their findings. The characteristics of these papers are summarised in Table 2.

Aims of the qualitative research

All of the papers retrieved had similar aims, seeking to explore model development processes, mostly from the perspective of those involved. For the majority, model development referred to models that compared clinical technologies, although one paper focused on model building within a public health context (Squires, 2014). Similar to the research proposed within this thesis, all of the empirical studies had an objective to explore model development and/or highlight issues with current methods and processes. Further, the majority of these studies had been justified on the basis of poor practice in the finished product or written processes of existing decision-analytic models, or a lack of guidance on good practice. All of the papers appeared to aim to generate findings that would be of assistance to modellers.

Table 2: Summary of paper characteristics

Author(s)	Year of publication	Type of publication	Setting/ country	Aims of the paper	Methods	Findings	Summary of quality concerns
Chilcott et al.	2010	HTA report	ScHARR, Sheffield, UK	To explore modeller perceptions of model development: to understand the similarities and differences in practice, and how errors might be introduced during the process.	Qualitative, in-depth interviews	Description of the modelling process, including methods used in each stage and advice on how to avoid common errors.	Sample size fairly small (12 modellers), but modellers were from different professional backgrounds.
Kaltenthaler et al.	2011	NICE Decision Support Unit (DSU) technical support document	ScHARR, Sheffield, UK	To report on suggested methods for structuring a model, and methods for the identification of evidence to inform model parameters.	Focus group	Paper provided detailed methods on how to conceptualise a structure, and for identifying evidence to populate data parameters, on the basis that current methods require improvement.	Limited information available on research methods. Only one focus group undertaken. Unclear how many people were involved and what their backgrounds were.

Author(s)	Year of publication	Type of publication	Setting/ country	Aims of the paper	Methods	Findings	Summary of quality concerns
Kaltenthaler et al.	2012	Journal article	ScHARR, Sheffield, UK	To highlight common issues identified by ERGs during analysis of manufacturer's submissions to NICE, including problems with submitted models.	Documentary analysis	A list of research recommendations for manufacturers, aimed at improving the quality of manufacturers' submissions.	Sampled only the first thirty ERG reports of those available. Therefore, questions were raised around the representativeness of this sample.
Kaltenthaler et al.	2013	Journal article	ScHARR, Sheffield, UK	To report on findings from the focus group on what might constitute a systematic and transparent approach to reviewing information to populate model parameters.	Focus group	Findings presented as issues of importance when using evidence to populate models, and paper provides advice on each of these.	Findings based on one focus group with 18 experts. Experts had different areas of expertise related to the research context. The authors reported reaching saturation regarding the comments of the informants on the themes generated.

Author(s)	Year of publication	Type of publication	Setting/ country	Aims of the paper	Methods	Findings	Summary of quality concerns
Kaltenthaler et al.	2014	Journal article	ScHARR, Sheffield, UK	To explore the issues concerned with the identification and use of information for model development.	Focus group	The paper reported on key issues of importance in the identification of evidence for models, and included guidance on good practice, and areas that require future research.	Findings based on one focus group with 13 experts. The authors reported reaching saturation of the themes generated.
Paisley	2012	PhD thesis	ScHARR, Sheffield, UK	To explore information requirements for the development of a model.	Case study research: interviews, focus groups and observation, within an action research framework	Recommendations on suitable approaches to the retrieval of evidence for models. Recommendations were based on methods used and problems reported by informants, but also the author's implementation of different search strategies.	The research involved case studies of the development of two models. However, it was unclear how many observations and interviews were undertaken within each case study, and how many informants were involved overall.

Author(s)	Year of publication	Type of publication	Setting/ country	Aims of the paper	Methods	Findings	Summary of quality concerns
Squires	2014	PhD thesis	ScHARR, Sheffield, UK	To explore the methods used and issues encountered in the development of public health models, specifically to understand how modellers make decisions around model structure.	Interviews, focus groups and observation	Research generated a conceptual modelling framework for the development of public health models.	Research followed the development of only one model in a single organisation. Three observations of meetings and two interviews were undertaken with the modellers involved. Focus groups were with only 5 informants, although these informants were from different organisations.

Five of the seven papers focused on investigating particular stages or aspects of model building (Kaltenthaler et al., 2014, 2013, 2011, 2012; Paisley, 2012), whilst the remaining two papers undertook research on processes of model development (Chilcott et al., 2010; Squires, 2014). The majority of the former group of papers used qualitative methods to explore how evidence was identified and/or retrieved to inform decision-analytic models. For Paisley (2012) and Kaltenthaler et al., (2011) this included evidence for the structure of the model, in addition to data parameters. The paper by Kaltenthaler et al. (2012) focused on the end stage of model development, specifically the submission of a model to the NICE STA process. The authors' aim was to identify common concerns with manufacturers' evidence submissions. Chilcott et al. (2010) and Squires (2014) both aimed to capture data on the process of model building, with the former authors concentrating on the entire process of modelling, and the latter particularly on structural development.

The qualitative methods used

A variety of qualitative methods were used to explore model development. Of the seven papers retrieved, one used documentary analysis, three used focus groups, one used in-depth interviews, and two used a combination of observation, focus groups and interviews. The papers differed in terms of how much they reported about the methods used, which was seemingly due to whether the publication was focused on the methods and findings of the qualitative study, or the recommendations made as a result. Of the papers that discussed the methods in any depth, most involved conducting primary research with the individuals involved in model development. The exception was the work of Kaltenthaler et al. (2012), which used documentary analysis to categorise 'key elements' of ERG reports regarding the aspects of manufacturer's submissions that the ERG identified as problematic.

The remaining studies differed in terms of whether they aimed to capture the opinions of those involved in model development, their behaviour, or both. Focus groups were used

within three papers (Kaltenthaler et al., 2014, 2013, 2011), as Kaltenthaler et al. (2014) suggested that this method was appropriate because of an objective to identify consensus between informants, and synthesise data according to the themes that individuals agreed were most important to modelling practice. Chilcott et al. (2010) were the only authors to use in-depth interviews alone, as the aim was to gain a detailed and ‘personalised view’ of the processes that modellers used to develop models. Chilcott et al. (2010) were interested in understanding how model development differed between modellers, and where informants considered errors were likely to occur in the process.

Squires (2014) and Paisley (2012) used interviews in combination with focus groups and observation. The authors both aimed to capture the behaviour of modellers, in addition to their opinions, to explore how particular aspects of model development were undertaken. Squires (2014) and Paisley (2012) also included their own modelling experiences and reflections as methods of investigation. Squires (2014) analysed notes from her involvement in a previous public health modelling project, to identify what she had deemed to be important to the development of a model structure, and Paisley (2012) undertook her study within an action research framework, implementing different approaches to identifying evidence within her role as an Information Scientist, and reflecting on the usefulness of the alternative searches used.

The findings of the qualitative studies

The findings of studies differed according to the objectives and contexts within which they were undertaken. However, common themes could be identified in terms of the nature and type of findings that the methods were able to generate. This was valuable in evaluating the usefulness of the different qualitative approaches, but also in determining the overall attitude within this literature towards modelling practices. The qualitative studies generated two

themes: (1) that current modelling practices require improvement, and (2) that further research and guidance is needed on model development.

Current practice requires improvement

All seven of the papers retrieved found problems with current methods of model development. Kaltenthaler et al. (2012) concluded that the general quality of manufacturers' submissions to the NICE STA process required improvement. The authors found through analysis of ERG reports that problems with submitted models were common, including the lack of detail available in submissions on modelling decisions, and the appropriateness of the data used to populate model parameters. Paisley (2012) concluded that the Cochrane style approach to evidence retrieval was not suitable for the requirements of decision-analytic modelling. These findings were based on the opinions and comments of informants, and Paisley's (2012) observations of informants' information seeking behaviours. The three papers using qualitative focus groups (Kaltenthaler et al., 2014, 2013, 2011) also suggested that methods for identifying evidence to develop and populate decision-analytic models required improvement. For example, Kaltenthaler et al. (2014) reported that focus group informants were uncertain about how they should search for data to populate certain parameters, particularly probabilities for rare adverse events.

Kaltenthaler et al., (2011), Chilcott et al. (2010) and Squires (2014) argued that improvement was required to model development on the basis of inconsistency in modelling methods. Chilcott et al., (2010) and Squires, (2014) found that the modeller informants interviewed and/or observed were undertaking different stages to one another in structural development. This disparity in informants' methods was viewed as problematic, as Squires (2014) interpreted these differences as uncertainty among modellers regarding good practice, and Chilcott et al. (2010) suggested that subjectivity in structural decision-making invited criticism around whether credible methods were being used across all modelling processes.

Guidance and further research is needed

In light of findings relating to problems with current modelling methods, the papers suggested that guidance and/or further research on model development was needed. As a result, all papers provided some advice to modellers, ranging from highlighting the issues that appeared most important to model development (Kaltenthaler et al., 2014, 2013), to a list of research recommendations (Kaltenthaler et al., 2012; Paisley, 2012), to good practice modelling processes (Chilcott et al., 2010; Kaltenthaler et al., 2011; Squires, 2014). Arguably, the style of advice offered by the latter set of authors provided the most in-depth and detailed guidance for modellers. Whilst Squires (2014) offered a conceptual modelling framework to guide a modeller stage by stage through the development of a public health model structure, Chilcott et al. (2010) presented a synthesised version of the modelling process as given by the interview informants, with advice on how to avoid common errors at various stages. Similarly, Kaltenthaler et al., (2011) gave detailed methods on how to conceptualise a structure, and identify appropriate evidence to populate model parameters. Although Kaltenthaler et al. (2013) and Kaltenthaler et al. (2014) provided numerous suggestions on good practice and issues to be aware of when identifying evidence for cost-effectiveness models, detailed methods and guidance on the processes involved with this aspect of model development were not given.

For Kaltenthaler et al., (2012) the level of detail and, arguably, the usefulness of the research output was restricted by the methodology chosen. Kaltenthaler et al. (2012) limited the scope of their research to identifying issues with manufacturers' submissions from ERG reports, rather than involving manufacturers and/or ERG members to collect data on how or why the problems recorded were occurring. The consequence of this was uncertainty and a lack of guidance on how modellers should change current modelling methods to ensure that they meet the standards expected by the ERGs.

Quality of the papers

The quality of the papers was assessed against the BMJ editor's qualitative research checklist (BMJ, 2015). When compared against the quality checklist criteria, the sampling and recruitment strategies used in the studies invited most concern. The BMJ checklist states that the sampling undertaken in qualitative research must be justified and 'theoretically comprehensive to ensure the generalisability of the conceptual analysis' (BMJ, 2015).

However, the majority of the studies had relatively small sample sizes, with Kaltenthaler et al. (2013) and Kaltenthaler et al. (2014) undertaking one focus group with eighteen and thirteen informants respectively, and Chilcott et al. (2010) undertaking interviews with only twelve modellers. Squires (2014) followed the development of only one model, carrying out three observations of meetings and two interviews with the modellers involved. Paisley (2012) observed practice and undertook interviews on the development of two models, however, it was unclear how many observations and interviews were undertaken throughout the two case studies. Only two of the studies reported reaching saturation in terms of the themes generated (Kaltenthaler et al., 2014, 2013). Kaltenthaler et al., (2011) reported their methods in very limited detail, stating only that a focus group with participants from one academic modelling organisation was undertaken, and thus making it difficult to assess the quality of the study.

Although qualitative research is not typically concerned with the number of informants sampled, the importance of capturing a range of perspectives and generating findings that are generalisable is emphasised (BMJ, 2015; Mays and Pope, 2000; Merkens, 2004). While some of the studies offered purposive and reasonably justified sampling strategies, such as the focus group studies (Kaltenthaler et al. 2013; Kaltenthaler et al. 2014) and Chilcott et al. (2010) who each aimed to sample a diversity of modelling experts, others did not. Kaltenthaler et al. (2012) sampled the 'first thirty ERG reports' with no justification as to why. In terms of achieving a typical set of findings, it would have been more robust to sample letters over a

period of time, given that issues identified within manufacturer's submissions are likely to evolve with changing modelling approaches. Further, both Squires (2014) and Paisley (2012) based their research on model development within one academic institution, specifically their own, which raises questions around the generalisability of findings, given that recommendations are based on the practice of a single team of modellers. Squires (2014) did however, sample modellers outside of her institution for the focus group, although this only involved five informants. None of the studies, to the author's knowledge, sampled the clinicians involved in model development, or modellers or other experts based outside the UK.

3.3.3.6 Conclusion of review

The review of the literature into the use of qualitative methods to explore model development has demonstrated that only very limited research has been undertaken within this area. However, the studies have indicated a need for changes to current modelling practices, and further guidance and research on modelling processes. The papers appeared to consider different aspects of model development important as the focus of research, although the majority concentrated on evidence retrieval and/or structural development. Only a few studies researched the entire modelling process. Most papers were concerned with exploring modelling from the perspective of those involved, although none of the studies included clinical experts. All of the papers reviewed were able to generate output to help modellers in their model development, and thus potentially improve the general standard of models. However, those using less exploratory qualitative approaches did not appear as able to produce detailed forms of guidance and recommendations as output. Concerns around the quality of existing studies have been raised, as the majority of the papers had relatively small samples, and lacked a clear and purposive sampling strategy.

3.4 Conclusion

The review of the health services research and health economics literature within this chapter has demonstrated the ability of qualitative methods to explore healthcare related processes, and to generate data that can help researchers to understand methodological issues. The findings of these qualitative studies facilitated the development of recommendations around the methodological issues present in preference-based valuation tasks and RCT recruitment, to potentially improve processes and limit future problems. The review of these areas of research suggested that qualitative methods could also be useful in exploring modelling processes, and in understanding how and why problems occur within model development.

The systematic review undertaken within this chapter identified studies that have used qualitative research to explore model development, and generate recommendations around good practice in model building. The findings of all of the studies indicated that future research and guidance were needed to improve modelling processes. However, the review also demonstrated that only a limited number of qualitative modelling studies had been carried out, and that those that had generally lacked quality in terms of the sampling strategies employed. Further, a potentially important omission from existing research was the involvement and role of clinicians in the modelling process. These findings suggest that qualitative research can be used, and is needed, to understand and address methodological issues in model building, but also that broader sampling and more robust qualitative approaches should be undertaken in future studies.

A consistent theme running through all of the literature reviewed was the strength associated with the use of more exploratory qualitative approaches, and the triangulation of qualitative methods. The general impression from the papers was that those using more inductive and in-depth approaches were able to generate a more detailed and nuanced understanding of the phenomena under study. The papers also demonstrated that using a combination of qualitative

methods was beneficial in providing potentially different insights into a phenomena, and emphasised the importance of choosing an appropriate research design, for example, using observational methods to capture communication where this might contribute to the conduct and results of a particular process. Drawing on the lessons learnt through the review of the qualitative studies within this chapter, this thesis aims to use a combination of in-depth qualitative methods to explore and document current processes of model development, using the findings to generate suggestions for improvements to modelling practices. The next chapter outlines the methods used in detail.

CHAPTER 4: METHODS

4.1 Introduction

This chapter reports on the qualitative methods used in each phase of the empirical research undertaken for this thesis. The research was divided into two phases: qualitative interviews with modellers and modelling case studies with teams of modellers and clinicians. The purpose of both stages of the research was to explore the processes used to develop decision-analytic models, including the methods followed, examples of good practice, opinions on poor practice and any problems encountered. The overall objective of the research was to generate findings highlighting the current processes used by modellers, and to make recommendations for aspects of model building that could be improved through further research or guidance.

Ethical approval was sought and granted for both phases of the research from the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham (reference: ERN_12-1533). Approval was also obtained from the University of British Columbia Behavioural Research Ethics Board for the Canadian phase of the research (reference: H13-01796). Evidence for the ethical approvals for the UK and Canadian phases of the research is given in Appendices 7-9.

4.2 Overview of research design

The theoretical framework applied to the research was one of constructivism-interpretivism, which assumes a relativist ontology and subjectivist epistemology, specifically that there are multiple realities, and that the 'reality' captured through research is one created through the combined understandings of the researcher and informant (Denzin and Lincoln, 2000). The

methodological position was one of naturalistic inquiry, which typically uses strategies such as ethnography and case studies to explore the empirical worlds of informants (Denzin and Lincoln, 2000).

The research undertaken within this thesis used qualitative methods to explore decision-analytic model development from the perspective of those involved. Based on the findings of the previous chapter, in-depth and exploratory methods were selected due to their ability to produce detailed and nuanced data on the experiences, opinions and behaviour of informants, in relation to model building. The research used in-depth interviews in its first phase, and case studies using non-participant observation, think-aloud and qualitative interviews in its second phase. The research sampled those currently working in the area of model building, specifically modellers and clinical experts. The sampling strategy undertaken in the research was to recruit informants who could provide a broad range of perspectives. Therefore, the modelling experts interviewed in phase one were from a variety of backgrounds, including academics and non-academics, and individuals from the UK and Canada. The modelling teams sampled for the case studies were also intentionally diverse, one being academic, and the other policy-based. The research used constant-comparison and framework methods for analysis.

4.3 Phase one: Qualitative interviews with modellers

4.3.1 Research design

Qualitative interviewing is the most commonly used qualitative method within health research and the social sciences (Green and Thorogood, 2009; Legard et al., 2003; O'Reilly, 2009).

Interviews are generally used to explore phenomena that cannot be observed, such as a person's thoughts, opinions and past behaviours (Patton, 2002). The aim of qualitative interviewing, also referred to as in-depth interviewing, is to produce 'rich, detailed accounts'

of the research topic under study ‘from the perspective of the interviewees’ (Green and Thorogood, 2009, p.95). Legard et al (2003) suggest that the key to generating personal and detailed informant accounts is to be flexible in allowing informants to freely discuss important issues, and to encourage informants to give in-depth responses to questions. As the objective of the research was to capture in detail the model development processes used by modellers, in addition to their individual perspectives on current modelling methods, qualitative interviews were considered optimal.

All qualitative interviews were conducted face-to-face, to facilitate the interactivity required to generate in-depth interview data (Legard et al., 2003). O’Reilly (2009) argues that establishing good rapport and trust with an informant is essential to the richness and quality of the data collected. Legard et al. (2003, p.143) suggest that rapport can be developed through the researcher ‘demonstrating a real desire to understand from the perspective of the interviewee’, while Coffey (1999) similarly argues that good data relies on an interviewer’s willingness to invest in an informant. Therefore Rubin and Rubin's (2005) approach of responsive interviewing was used. This involved the researcher formulating future questions based on informants’ previous answers, demonstrating awareness and interest in what the informant was discussing, but also gaining more depth and detail on matters that they considered important.

4.3.2 The sampling and selection of informants for interview

To enhance the credibility of research, Rubin and Rubin (2005) suggest selecting informants based on their knowledge and experience of a topic, choosing those who have a first-hand and in-depth understanding of the phenomenon under study. Informants were therefore selected to take part in the qualitative interviews on the basis of their experience of building decision-analytic models, and their perceived ability to offer opinions on the model-building process. This is referred to as purposeful sampling, which focuses on choosing ‘information-rich’

cases ‘from which one can learn a great deal about issues of central importance to the purpose of the inquiry’ (Patton 2002, p.230). Merkens (2004, p.167) also draws attention to the importance of having a sample that represents as many ‘facets’ of a particular group of informants as possible. Mays and Pope (2000, p.51) argue that the validity and credibility of qualitative research depends on the researcher ensuring ‘that the research design explicitly incorporates a wide range of different perspectives so that the viewpoint of one group is never presented as if it represents the sole truth about any situation’. Firestone (1993, p.17) also contends that sampling for different opinions on the same research issue can promote generalisability, stating that ‘similar results under different conditions illustrate the robustness of the finding’. Therefore, modellers were also approached and sampled according to the different experiences of modelling that they possessed, their seniority and the context in which their modelling work was conducted.

Sampling was continuous and purposive in the sense that the background and outlook of previous informants influenced the selection of informants for the next round of interviews. This is also referred to as gradual selection, and involves assessing who to approach next based on the nature of the data previously collected, and the desire to gain a range of perspectives (Flick, 2009). The informants eventually sampled were modellers working in either an academic or non-academic role. The academic modellers sampled worked for universities either in the UK or Canada, and the non-academic modellers were working for a modelling consultancy, pharmaceutical company, or policy institute in the UK. Snowball sampling was used to gain access to most of the non-academic and Canadian academic informants with whom the author or her supervisors did not have existing links. Snowball sampling involves asking existing informants to recommend individuals whose experiences are related to the research (Kuper et al., 2008). The non-academic and Canadian academic informants were in each case mostly identified through one key informant.

4.3.2.1 Sample size

In line with the ongoing sampling approach (O'Reilly, 2009), informants were recruited to the interviews in waves, to allow future data collection to be informed by the themes emerging in previous findings. For example, it was suggested in the first wave of interviews with UK academics that the modelling processes and methods used by those in industry might differ from their own. Therefore, going forward, it was considered important to recruit non-academic modellers to ensure a representative view of model development.

Sample size was not decided in advance of the research, with sampling instead continuing to the 'point of diminishing return, where increasing the sample size no longer contributes new evidence' (Richie et al. 2003, p.83). This was determined as the point at which informants were continually offering similar accounts and opinions of the model-building process, for example, repetition of the stages, issues and methods used. It was also considered important that enough data were collected for each of the different groups of informants, and that saturation was occurring across all interview accounts. This view of saturation of data categories is aligned to that defined by Bowen (2008, p.148), who argued that sampling should continue until the interviewer 'hears the same comments from different participants in different places'.

4.3.2.2 Recruitment

Methods of recruitment differed depending on the type of informant sampled. Searching staff profiles on various university websites was the strategy used to identify UK academics. The first step was to compile a list of UK universities that undertook modelling work, identifying potential informants within these universities whose role involved decision-analytic modelling. From this list, a shortlist of informants to approach was decided on in conjunction with the research supervisors. This decision was based on the information given within an informant's staff profile, with the aim of sampling those working within different size teams,

in different disease areas and at different levels of seniority, and to obtain a geographical spread of informants.

Rubin and Rubin (2005) suggest that informants are more likely to agree to participate in an interview if they have a shared background with the interviewer, for example, a mutual contact. Therefore, all UK academic informants on the initial shortlist were initially emailed by one of the research supervisors who had existing connections with the individual. Potential informants were sent an introductory email outlining the research, introducing the author, and drawing attention to email attachments. The attachments included an official letter inviting the informant to take part in the research, and an information sheet that outlined research objectives, data collection and storage and ethical considerations. Copies of a sample invitation letter and the informant information sheet are provided in Appendix 10 and Appendix 11. Potential informants were asked to reply to the author directly if they were interested in participating, and were sent a follow-up email after seven days if they had not responded.

Non-academic UK informants were almost all identified through a key informant, who was an existing contact of one of the research supervisors. The key informant was approached for interview because he worked for a modelling consultancy, and was able to suggest other modellers working for consultancies and pharmaceutical companies who could be sampled. Policy informants were recruited directly by the author through an established contact, who took part in the interview himself but also recommended a further policy informant. Potential non-academic informants were contacted by either the author or key informant with an introductory email, official invitation letter and information sheet.

The interviews with academic modellers in Canada were facilitated by funding gained through the Universitas 21 Doctoral Scholarship Programme. This programme allowed the author to travel to a university outside the UK, where research was similarly being undertaken

using decision-analytic models. Canada was specifically chosen because it shares a similar regulatory structure and health technology assessment process to the UK, meaning that the data generated between the informants in Canada and the UK were comparable. It was anticipated that interviewing Canadian academic modellers on their model development processes would provide an international perspective, and could highlight differences to the methods used by UK modellers. The author was assigned a research supervisor at the host university who assisted with the recruitment of modellers to interviews. This research supervisor identified and recruited a key informant, and the key informant then identified other informants to be contacted to take part. Canadian informants were selected on the same basis as those in the UK, on their knowledge and experience of decision-analytic modelling. The key informant was also asked to identify a diverse group of modellers in terms of the context of their work and their seniority. Academics from a number of different institutions affiliated to the host university were therefore considered as potential informants. In the same way as with the UK academic informants, all Canadian academic informants were initially contacted through the research supervisor in Canada, and sent an official invitation letter and information sheet.

4.3.3 Data collection

For those interested in participating in an interview, a convenient date and time was arranged via email. Most interviews took place at the informant's place of work, as this was considered a setting where an informant would feel comfortable and relaxed, facilitating the development of rapport (Green and Thorogood, 2009). However, one interview was conducted in a café, as this was identified as the most convenient place by that particular informant. All of the interviews, with the permission of the informant, were recorded using a digital audio-recorder, to allow the interviewer to focus full attention on the interview exchange (Legard et al., 2003), and provide a rich source for data analysis. Written notes were taken to remind the

author of issues to follow-up with informants later in an interview. All of the interviews were undertaken by the author.

The questions asked in the interviews followed the framework proposed by Rubin and Rubin (2005, p.152), namely 'a balance of main questions, follow-up questions, and probes'. The main questions were prepared prior to the interview and were designed to explore the issues perceived to be important to the research topic, including the processes used to develop decision-analytic models, and informant opinions on the methods used. Main questions were phrased in an open-ended way in order not to influence responses, and to encourage informants to discuss issues most salient to them (McNamara, 2009). Follow-up questions were devised during the interviews and in response to what informants had said. The purpose of these questions was to gain a deeper understanding of issues that informants raised when answering main questions (Creswell, 2013). Probes were used to encourage the informant to expand on specific points, for clarification or to facilitate the flow of the interview. Many of the probes used were non-verbal and were to reassure the informant that they were discussing issues that were important and interesting to the research, for example, 'leaning forward to express interest, and...taking notes in ways that signal the interviewee to keep talking' (Rubin and Rubin, 2005, p.164).

An interview guide was developed to remind the author of important questions to ask. Questions were ordered according to the funnelling technique, which involved moving from 'general enquires to more specific and focused questioning' to allow informants 'to relax into the interview' (Wilkinson and Birmingham, 2003, p.47). Questions were initially focused on asking for background information from the informants on their modelling experiences, and moved on to detailed issues, such as informant opinion of current model-building guidance. However, the use of the interview guide was flexible in the sense that questions, and their order, were influenced by topics that informants wanted to discuss. Informants were free to

introduce and discuss ideas that were not in the interview guide, and these were followed-up for depth and detail in the same way as answers to planned questions. Further, questions at the end of the interview focused on asking for suggestions for future research, and on ensuring that the informants had nothing else that they wanted to contribute (Legard et al., 2003). The interview guide was modified a number of times over the course of the interviews to encompass additional issues that informants raised. A copy of the interview guide is provided in Appendix 12.

4.3.4 Analysis

All of the interviews were analysed by the author. Interviews were transcribed verbatim to ensure a reliable account of the informants' thoughts and opinions (Green and Thorogood, 2009). Analysis drew on grounded theory procedures (Strauss and Corbin, 1990) and the method of constant comparison (Glaser and Strauss, 1968; Glaser, 1965). Data collection and analysis occurred simultaneously, as analysis of the data collected influenced the sampling of further data (Strauss and Corbin, 1990), and continued analysis allowed a deeper understanding of the data categories emerging. Interview transcripts were coded individually in the first instance, with 'open coding' used to 'break down' and 'conceptualise' data into incidents that could be assigned a representative label (Strauss and Corbin, 1990).

Open coding of transcripts was initially undertaken using line-by-line analysis; however, as coding continued and categories were identified and became well-established, transcripts were coded by sentence or paragraph, as focus moved toward developing a deeper understanding of existing themes (Strauss and Corbin, 1990). As required by the constant comparative method, data coded with a particular label or within a particular category were continually 'compared with the previous incidents coded in the same category' (Glaser 1965, p.439). This allowed the author to begin to understand the properties of a category, and how emerging categories related to one another (Glaser, 1965; Strauss and Corbin, 1990). As analysis progressed, new

instances of data were compared to the properties of emerging categories, to further make sense of how concepts were connected. Comparisons were facilitated through continuous questioning of what each instance of data meant, and how it was similar or different to previous incidents, and concepts presented within categories as a whole (Strauss and Corbin, 1990). The use of questioning also improved the theoretical sensitivity of the analysis, as data were examined for additional meanings and relevance (Strauss and Corbin, 1990).

Coding was initially carried out on the first three interview transcripts (referred to as I1, I2 and I3). As comparisons continued within and between these transcripts, data were sorted and reduced into first and second level codes. These codes were noted and applied to each of the initial three interview transcripts in their entirety. Further coding and revision of the data, referred to as 'axial coding' (Strauss and Corbin, 1990), was carried out to establish hierarchical relationships between codes, forming categories and sub-categories of data. Written and diagrammatical coding structures were developed to record this hierarchy. A descriptive account was generated for the first batch of interviews, to facilitate analysis and comparison of informant responses and synthesise the data within each category.

Interviews continued to be undertaken and analysed in batches of three to seven, with informants being grouped according to their backgrounds, rather than order of interview. The coding structure developed through the full analysis of the initial batch of interviews was applied to the transcripts of the second batch, and a descriptive account was produced. This continued for all batches of interviews, as the analysis of previous transcripts updated and revised the categories and sub-categories that were applied to future batches. Collecting and analysing the data in this way allowed for previous analyses to influence the questions asked in future interviews, and facilitated the author's understanding of the important issues that were emerging from the data. For example, clinician involvement in the modelling process was perceived to be important after two interview batches; therefore, in future interviews

informants were probed to give greater depth and detail on this aspect. Comparison of the coding structure diagrams developed at the beginning and end of analysis demonstrate that as analysis progressed, some original codes were combined, whilst others were divided to form distinct categories. For example, 'integrity' was originally represented as a category in its own right, but was later understood to be an aspect of what informants considered as good practice in model development. Likewise 'clinician involvement' became a first level code, as it became apparent that there were many issues around their involvement in model development. The initial and final versions of the coding structure diagrams are available in Appendix 13 and Appendix 14, and the final version of the coding structure definitions is available in Appendix 15.

The final stage of analysis involved combining all descriptive accounts to display all informants' thoughts and opinions under each category and sub-category of data (Richie et al., 2003b). The comments of the informants were compared for similarities, and specific attention was paid to any circumstances indicative of differences in opinion or practice. Given that a focus of the research was to investigate how modellers typically undertake their modelling processes, matrices were developed to compare and contrast informant responses, specifically to see which stages of model development were most commonly followed, and where there were differences in methods used. An excerpt from these matrices is available in Appendix 16.

The coding of the interview transcripts was carried out both manually and electronically for thoroughness. Codes were noted in the margins of hard copies of transcripts, and then later in the comments section of electronic copies in Microsoft Word. To improve validity, the coding and analysis of the interview transcripts were challenged in various ways by the research supervisors. Each of the research supervisors independently coded the first interview transcript, and their insights were incorporated into the development of initial data categories.

The original coding structure was also developed in conjunction with research supervisors, who read the coded versions of the first three interview transcripts and offered their understandings of how the data were related and what appeared important. The research supervisors also read each descriptive account thoroughly and gave opinions on the appropriateness of the content of data categories, and feedback on analysis technique.

4.3.5 Ethical considerations

Ethical issues were considered prior to commencement of the interviews, and ethical conduct was maintained throughout the research. Much time was spent at the planning stages considering any potential risks to informants, and although there appeared to be little risk of harm due to the professional rather than personal nature of the research, precautions were still taken. Given that informants were being interviewed in a professional capacity, there was a possibility that discussion of their work could make them identifiable. Therefore, to safeguard informants' identities and guarantee anonymity and confidentiality, interview transcripts were anonymised, and all data were stored in a secure way. Any information that could potentially identify an informant, for example, the informant's name, details concerning their work history, modelling projects and modelling teams, were removed from the transcripts and replaced with a code. All data were stored on a secure, password protected computer network that could only be accessed by the author. The personal information of the informants was stored separately from their interview transcripts.

Gaining the informants' consent to take part in interviews was seen as a continuous process, rather than a single event, due to the right of the informants to reconsider their involvement in the research at any time (The Research Ethics Guidebook 2015). Informed written consent was taken from each informant prior to their interview, after they had been 'informed fully about the purpose, methods and intended possible uses of the research, [and] what their participation in the research' entailed (The Research Ethics Guidebook 2015). Informants

were encouraged to read the informant information sheet prior to consenting to participate in the interview, and to ask the author any questions that they had. Importantly, the informant information sheet (see Appendix 11) outlined the informant's right to withdraw from the research at any time, even after providing written consent. Karnieli-Miller et al. (2009) stressed the importance of the contributions and rights of informants being respected throughout the research process, and as a result informants were reminded just before their interview that they still had the right to withdraw during or after data collection. A copy of the written consent form is available in Appendix 17.

4.3.6 The interview informants

A total of twenty-four interviews were undertaken with health economic modellers from various backgrounds. Table 3 shows the basic characteristics of these informants, in addition to comments on their interviews. The informants were almost evenly distributed between academics working in the UK, academics working in Canada and non-academic UK informants. Sixteen of the informants were male and eight were female, however, the gender of the informants was not considered to be important to this research. The interviews aimed to sample a mixture of senior and more 'junior' informants. Senior informants were defined as those who worked in a managerial role and were largely overseeing the work of others. The ratio of senior to junior informants was nine to fifteen; however, there were varying degrees of experience among the junior informants. While some appeared to have been modelling for a number of years and worked on numerous models, others had worked on only one or two. Capturing this range of informant experience was seen as beneficial in gaining a breadth of perspective on the modelling process. The differences in the informants' experiences of modelling are examined in more detail in the next chapter. The average interview lasted for around one hour, and most took place in a quiet location. A few of the interviews suffered from interruptions such as background noise or telephone calls; however, these did not

prevent the interviews from continuing. Some of the informants were more difficult to engage than others, appearing nervous or tending to offer short answers to the questions asked. However, in most cases the informants became more relaxed as the interview progressed, and the open-ended style of questioning and prompting encouraged the informants to speak for longer and in more depth.

Table 3: Interview informant characteristics

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on interview
I1	Female	Academic	UK	Senior	Informant's place of work	Quiet office, no disruptions	1 hour	Informant was very talkative and spoke in-depth in response to most questions.
I2	Male	Academic	UK	Junior	Informant's place of work	Quiet office, no disruptions	35 minutes	Informant seemed nervous before the interview but was fine answering the questions. He remembered a lot of information toward the end of the interview that he had not said during.
I3	Female	Academic	UK	Junior	Informant's place of work	Quiet office, no disruptions	50 minutes	Interview flowed well, the informant spoke for long periods of time on each topic without being prompted.
I4	Male	Academic	UK	Junior	Meeting room, informant's place of work	A bit of background noise but no disruptions	1 hour	Interview went well.

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on interview
I5	Female	Academic	UK	Senior	Informant's place of work	Quiet office. Interview was paused once to allow informant to search for a reference	40 minutes	Some of this informant's answers to questions were fairly short and required further prompting.
I6	Male	Academic	UK	Junior	Informant's place of work	Quiet office, interview interrupted once when informant stopped to answer the telephone	1 hour, 20 minutes	Interview went well but felt long and the informant spent a long time discussing the context of his work. He spoke freely and at length in response to the questions asked.
I7	Male	Non-academic	UK	Junior	Meeting room, informant's place of work	Quiet room, no disruptions	55 minutes	The informant appeared to be nervous. He required prompting and further questioning to gain clarification on his points.

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on Interview
I18	Female	Non-academic	UK	Junior	Meeting room, informant's place of work	Quiet room, no disruptions	1 hour, 10 minutes	Good interview, very informative, although the informant spoke rather technically at times.
I19	Female	Academic	UK	Senior	Informant's place of work	Quiet office, no disruptions	50 minutes	Informant was very easy to chat to and rapport was quickly established.
I10	Male	Academic	UK	Senior	Informant's office at his place of work	The office overlooked a very busy main road, which at times was a little disruptive.	1 hour, 15 minutes	The interview seemed to last quite a long time but generally went well.
I11	Female	Non-academic	UK	Junior	Meeting room, informant's place of work	Quiet office, no disruptions	35 minutes	This informant gave fairly short answers to all of the questions asked.

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on interview
I12	Male	Non-academic	UK	Senior	Meeting room, informant's place of work	Quiet office, no disruptions	55 minutes	Interview went well.
I13	Female	Non-academic	UK	Junior	Meeting room, informant's place of work	Quiet office, no disruptions	1 hour	The informant was enthusiastic and talkative, giving detailed answers to most of the questions asked.
I14	Male	Non-academic	UK	Senior	Meeting room, informant's place of work	Quiet office, no disruptions	1 hour, 20 minutes	This informant was very enthusiastic and led a lot of the discussion and the direction of the interview.
I15	Male	Academic	UK	Junior	Informant's place of work	Interview was disrupted a few times to allow informant to answer the phone	1 hour, 45 minutes	The interview lasted for almost 2 hours and I felt very tired towards the end. The informant spoke freely and confidently about all topics, although rather technically at times.

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on interview
I16	Male	Non-academic	UK	Junior	In a cafe, near to the informant's place of work	A lot of background noise and people around, which felt distracting	1 hour, 15 minutes	Interview was quite long but informant often did not provide a lot of depth and explanation to the points that he was making.
I17	Male	Non-academic	UK	Senior	Informant's place of work	Quiet office, no disruptions	55 minutes	Informant was very informative and clear in all of his answers.
I18	Male	Academic	Canada	Junior	The author's office at the university	Quiet office, no disruptions	30 minutes	This informant struggled to recall some of his modelling practices, making the interview difficult at times.
I19	Male	Academic	Canada	Junior	The author's office at the university	Quiet office, no disruptions	1 hour	The informant gave a lot of background to his projects and the interview was slow to 'get started'.

Informant	Gender	Nature of work	Location of work	Level of modelling experience	Place of interview	Comments on setting	Approx. length of interview	Comments on interview
I20	Male	Academic	Canada	Junior	The author's office at the university	Quiet office, no disruptions	40 minutes	Interview went well.
I21	Male	Academic	Canada	Junior	The author's office at the university	Quiet office, no disruptions	45 minutes	The informant was enthusiastic and provided in-depth answers to most questions.
I22	Male	Academic	Canada	Senior	The author's office at the university	Quiet office, no disruptions	55 minutes	The interview went well, the informant was very talkative. Rapport was quickly established.
I23	Female	Academic	Canada	Junior	The author's office at the university	Quiet office, no disruptions	40 minutes	Interview went well.
I24	Male	Academic	Canada	Senior	The author's office at the university	Quiet office, no disruptions	50 minutes	Interview was slow at first with the informant giving fairly short answers to the questions asked. However, answers became more informative as the interview progressed.

4.4 Phase two: Model-building case studies with modellers and clinicians

4.4.1 Research design

The second phase of the research involved model-building case studies with teams of modellers and clinicians. The case studies and the methods used within them were designed to complement and broaden the findings of the first phase of the empirical research: the qualitative interviews with modellers. The case studies aimed to complement the findings of the qualitative interviews in two ways. Firstly, the findings of the case studies could either support or challenge the accounts of modelling processes given by the interview informants. Secondly, and due to the observational methods used, the case studies could generate findings on what people do, in addition to what they say they do (Green and Thorogood, 2009). Mays and Pope (2000) argue that comparing the findings of two or more different data collection methods can ensure comprehensiveness in the research conclusions. The case studies could also expand on the findings of the qualitative interviews, as they allow for the first-hand, close and detailed exploration of the 'real-life situation', i.e. the model-building process, under study (Flyvbjerg 2006, p.223).

The literature offers different definitions of a case study, with some defining it as the sample of people or phenomena being studied (Hammersley, 1992; Stake, 2003), and others identifying it as a methodological approach to research (Yin, 2014). From either perspective, a case study is concerned with gathering 'comprehensive, systematic, and in-depth information about [the] case of interest' (Patton 2002, p.447). For this research, the case of interest referred to the model-building process, and the stages and methods that the modellers and clinicians were undertaking. Considering the case study as a methodological approach to research, Yin (2014, p.4) argues that it is an important method when an answer to a 'how' or 'why' question is required, and when 'an extensive and in-depth description of some social

phenomenon' is needed. Indeed, the objective of this research was to explore how decision-analytic models were being developed in practice, and to capture any problems that occurred within the model-building process. If any problems were observed, the research aimed to explore why these occurred, in addition to how they were resolved. Therefore, the case study 'method' lent itself well to the questions being asked around the modelling process, and also addressed the need to generate a 'thick description' of the 'issues, contexts, and interpretations' involved with model development (Stake, 2003).

Yin (2014, p.17) identifies the use of 'multiple sources of evidence' as being fundamental to a case study inquiry, and advocates the combination of methods such as interviews, collection of documentary information, and observations. Although case studies are not essentially qualitative (Stake, 2003), all of the methods used within this research were. The case studies used non-participant observation, qualitative interviews and think-aloud interviews to gather data on the model-building process. These methods were selected with the aim of collecting data on as many of the stages undertaken in model development as possible, with reflections from those involved. Non-participant observation was used to observe the meetings that took place between those working on the model, whilst the qualitative interviews allowed the author to question those involved on aspects of the model's development and the process in general. Data collection also extended to emails that were sent between the modellers and clinicians, and any documents that were used or discussed within the meetings that were observed. The think-aloud interviews were used to elicit data on specific stages of the modelling process that typically involved the modeller working independently.

Using these methods in conjunction with each other was expected to give a breadth and depth of understanding of the way in which modellers and clinicians undertake stages of the modelling process, and their opinions on the methods used. Stake (2003, pp.446-448) identified the value of case studies as being an 'opportunity to learn' from the study of a

particular phenomenon and to suggest ‘complexities for further investigation’. The research aimed to provide full and analytical accounts of the model-building process that those working within the field could potentially take guidance from, and also make recommendations for future research that could enhance the methods of model development.

4.4.1.1 Case study selection

Two case studies were selected for analysis, with the focus being on the process used to develop a decision-analytic model. The decision to undertake two case studies appeared optimal given the desire to maximise the generalisability of the research conclusions within the time available. Schofield (2000, p.79) argues that ‘a finding emerging repeatedly in the study of numerous sites would appear to be more likely to be a good working hypothesis about some yet unstudied site than a finding emerging from just one’. Each case study aimed to capture the stages involved in the development of one health economic model from beginning to end, involving multiple exposures to model development to ensure that a representative perspective was being captured (Krefting, 1991).

The case studies were selected on the basis of the information that they could generate, and the potential implications of their findings more generally. Flyvbjerg (2006, p.230) suggests that sampling for ‘critical cases’ provides a more useful insight into a particular phenomenon, and enhances the generalisability of conclusions. Critical case sampling involves looking for cases that permit ‘logical deductions of the type’ i.e. ‘if this is valid for this case, then it applies to all cases’ (Flyvbjerg 2006, p.230). Therefore, the aim was to sample modelling teams which had a particularly good reputation for model building, on the assumption that any problems or issues that were encountered within their modelling processes were also likely to be encountered by modellers with less experience. The aim was also to sample a range of model development processes, on Polit and Beck's (2010, p.1454) recommendation that replication of findings between similar ‘cases that vary on attributes likely to affect

conceptualization can...strengthen generalization'. Therefore academic and policy-based case studies were sampled, referred to as Case Study A and Case Study B respectively. This was on the assumption that they might take alternative approaches to, and use different methods in, model development given the different contexts of their work. It was anticipated that the recommendations made from the case study findings, would focus on the issues that were most challenging in model development, and would be applicable to multiple model development settings.

4.4.1.2 Recruitment

The two case studies were selected and recruited using the suggestions of the interview informants, as they were asked towards the end of their interviews to suggest teams of modellers who would be an interesting focus for the research. As a result, many informants recommended undertaking case studies within institutions that had a good reputation for model development, and a number suggested sampling models that were being developed in different contexts to one another, in anticipation of capturing a broader perspective. Also, some informants offered modelling projects that they were due to work on as potential case studies. Based on these recommendations and ease of access, two of the modelling case studies offered by the interview informants were selected.

The two interview informants who set up the case studies were asked to distribute an information sheet to those involved in model development, to ensure that they were fully informed and happy to take part. The information sheet outlined the methods of the case studies, how informant data would be used and managed, and ethical aspects such as the informant's right to withdraw. The information sheet also stated that the informants might be asked to take part in an interview during the case study, to give their opinions on the modelling process. Written informed consent was obtained from those involved, with

additional written consent being taken later for those who agreed to be interviewed. The informant information sheet for the case study research is available in Appendix 18.

4.4.2 Methods and data collection

4.4.2.1 Non-participant observation

Patton (2002, p.262) argues that directly observing phenomena has the advantage of allowing the researcher to understand the context of an interaction, and observe aspects that an informant would not necessarily be aware of or discuss. The aim of using observations was to gain a deeper insight into the model development processes being used, including the nuances of how those involved contributed and communicated. Non-participant observation was selected as it was not deemed necessary for the author to have an active role in the research setting. Glesne (2006, p.50) refers to this position as ‘participant as observer’ on ‘the participant-observation continuum’; defined as such because ‘the researcher remains primarily an observer but has some interaction with the study participants’. The aim of the non-participant observation was, as far as possible, to capture the typical processes used by modellers and clinicians, without any influence from the author. In addition, it was not considered practical for the author to be involved in the model building given the professional nature of work undertaken.

The observation of the modelling processes involved observing all meetings between the modellers, and the modellers and clinicians. This included the observation of face-to-face meetings, documents referred to within discussion, and email communication between those involved. For Case Study A, the author was permitted to audio-record all meetings, which provided a rich and detailed source for analysis. Field notes were taken in addition to the meeting transcripts to record details of the setting and the author’s perceptions and thoughts. Data collection for Case Study B relied mostly on field notes, as the author was not permitted

to audio-record meetings and had limited access to the modellers' own recordings. The field notes taken therefore had to be very thorough to ensure that enough detail was captured on every interaction. On the recommendations of Patton (2002), both descriptive and analytical field notes were recorded, with the former involving a description of the setting and what was being observed, in addition to direct quotations from the informants, and the latter involving the author's interpretations of what was happening. Silverman (2000) refers to the latter as 'expanded field notes' and suggests that that these provide an important basis for data analysis. Expanded field notes were taken in the form of 'contact summary sheets' (Miles and Huberman, 1984, p.25) that were created for each observation of a meeting, and summarised the main issues that were emerging from the data, including any 'new hypotheses or speculation'. Miles and Huberman (1984, p.25) argue that data that are analysed as they are collected are 'radically more analysis-rich', and the analytical field notes generated allowed the author to develop a deeper understanding of important issues as the research progressed. The descriptive notes were recorded separately from the author's interpretations. Field notes were handwritten but typed up as soon as possible after each observation.

4.4.2.2 Qualitative interviews

Face-to-face or telephone interviews were conducted with the modellers and clinicians involved in both case studies. Most of the interviews were face-to-face and took place in the case study setting; however, the interviews with the clinicians in Case Study B were by telephone, due to their busy work schedules. Although there are concerns among qualitative researchers that it is more difficult to develop rapport with informants and collect rich data using telephone interviews, it is also argued that telephone conversation may put people at ease 'and allow [them] to disclose sensitive information more freely' (Novick, 2008, p.8). Therefore telephone interviews were seen as a fruitful and practical method of data collection. The interviews took place at various points in the models' development and were designed to

capture the reflections of the case study informants on aspects of the development processes that had been observed thus far.

The interview guide contained questions that were specific to the models being built, in addition to questions about the general processes. The intention was to gauge from the modellers or clinicians what they felt was good about the current model and processes, and what they felt was not as good. This was designed to generate findings on methods that might be considered as good practice, and aspects that might require improvement. The interview guide also asked the informants for their opinions on issues that had emerged from the findings of the qualitative interviews in the first phase, such as the suggestion of guidance for clinicians.

The qualitative interviews undertaken within the case studies were more structured than those in the first phase of the research. There was less flexibility in terms of the topics discussed, as much of the interview focused on asking informants about issues that had emerged during the observations. These questions were prepared in advance of the interviews and typically involved checking the author's interpretations of observations, and asking informants to expand on model development activities and aspects of their role. However, informants were still given the opportunity to introduce and discuss issues that they considered important, as follow-up questions were asked to gather further detail on comments and opinions, and all informants were asked at the end of their interviews whether there was anything else they wanted to contribute that they felt was relevant. A copy of the interview guide for the case study interviews is available in Appendix 19.

Informants were asked prior to their interviews to fully read through the interview information sheet, ask the author any questions that they had, and sign the additional interview consent form (copies of which are available in Appendix 20 and Appendix 21 respectively). All of the interviews for both case studies were audio-recorded with the permission of each informant.

4.4.2.3 Think-aloud

Think-aloud generates data on the ‘reasoning processes’ that individuals use to undertake particular problem solving tasks (Fonteyn et al., 1993). More specifically the method works to ‘identify the information that is concentrated on during problem solving and how that information is used to facilitate problem resolution’ (Fonteyn et al. 1993, p.430). Data are obtained by asking informants to ‘talk aloud’ while performing cognitive tasks, capturing verbalisations of their thought processes (Ericsson and Simon, 1984). The focus of the think-aloud used in this case study research was to explore how modellers completed stages of the modelling process that were undertaken individually and implicitly. Think-aloud therefore allowed the author access to processes within modelling that could not be explored through observation alone. Further, think-aloud was considered a more appropriate method than qualitative interviews, as rich and detailed data could be captured on the entire stage of model development under study, without the concern of recall bias (van den Haak et al., 2003).

Think-aloud was carried out within one of the two case studies and focused on following and exploring the methods that the modeller used to test the model results, specifically the sensitivity analysis. This aspect of the process was selected not only because the modeller was working independently, but also because in-depth data could be generated on the cognitive processes that the modeller used to interpret analyses (van Someran et al., 1994). The aim was to learn about the type of sensitivity checks carried out, but also about how the modeller understood and made sense of the test outputs. In line with the suggestions of Fonteyn et al (1993), think-aloud was undertaken in a quiet location, specifically in the informant’s work office, with as little interaction from the author as possible to avoid interruption to the informant’s thoughts. The informant was instructed prior to the think-aloud to ‘speak constantly without regard for coherency’ whilst completing the sensitivity analyses, and was only communicated with to be reminded to ‘keep talking’ after periods of silence (Boren and

Ramey, 2000). The think-aloud was audio-recorded with the permission of the informant, and the recording was transcribed verbatim. The author also took note of any questions or clarifications that seemed important during the think-aloud, so that these could be discussed with the informant after the session.

4.4.3 Analysis

Case study analysis followed the framework method outlined by Richie and Spencer (1994). Framework was selected because it facilitates a combined approach to deductive and inductive thematic analysis, allowing existing and specific issues to be explored, but also providing an opportunity for new experiences and interpretations to emerge (Gale et al., 2013). This was essential given that the aim of the case studies was to look for support for, but also expand on the findings of the qualitative interviews. The method is also orientated towards the management of large data sets, and promotes ‘within-case and between-case analysis’ (Srivastava and Thomson, 2009, p.77). This meant that informant experiences and opinions on particular modelling stages could be compared within the same case study, and similarities and differences in modelling practices across the two case studies could be explored. The framework method involved undertaking five key stages of qualitative data analysis:

- familiarisation,
- identifying a thematic framework,
- indexing,
- charting,
- mapping and interpretation (Richie and Spencer, 2002).

Familiarisation was the first step of analysis, and required the researcher to become 'immersed' in the data that were gathered from each case study (Richie and Spencer, 2002). Case studies were analysed individually and all collected material was revised. This specifically involved reviewing meeting and interview transcripts, re-reading observational field notes, and in some cases listening to original audio-recordings. This was whilst creating 'memos' on the ideas and themes that appeared important and were recurrent throughout the data. For example, as memos were generated for Case study A, it became apparent that miscommunication and misunderstanding between the modellers and clinician was an issue at a number of stages of the modelling process.

From the 'key issues, concepts and themes' identified through review of the case study material, the next stage of analysis involved the development of a 'thematic framework', used to sort and label the data (Richie and Spencer, 2002, p.313). This coding framework was also informed by themes raised in the first phase of the research, and by issues covered on the case study interview guide. This was important given that the aim of the case study research was to further explore issues that were identified as important in the in-depth interviews, whilst also seeing whether similar modelling methods were used across the interviews and case studies. The first version of the thematic framework therefore contained codes indicative of the overall stages of the modelling process demonstrated in the qualitative interviews and case study material, in addition to codes for issues that were important in the first phase, such as clinician involvement.

Consistent with the findings of Richie and Spencer (2002), the thematic framework was refined as it was applied to the case study data, as new codes, particularly those at the secondary level, emerged and were added. The fluidity of the thematic framework therefore allowed new and important issues to be explored, and provided an overview of how the case study data were related to the findings of the qualitative interviews with modellers. For

example, the framework initially contained and coded data according to the three broad elements of model development that were generated in the first phase of the research, namely structural development, populating the model and model checking. However, it became clear after reviewing the case study material that there were data available on the results stage of model development, indicating the need for an additional first level code. Similarly, as the data were coded, further lower level codes were added to the framework to account for the description of model development stages and richness of data that the case study methods were able to generate. The sub-categories of ‘where patients go and why’, ‘decisions on how to model pathways’, and ‘outcomes of pathways’, developed from the ‘drafting the structure’ code generated through the qualitative interviews, as observation allowed the author to learn in detail of the information that appeared most important in developing the pathways of the model.

The thematic framework generated through the review of all case study material, and the coding of Case Study A was then applied to Case Study B. Although no first level codes changed, indicating similarities in the overall model development stages undertaken between the case studies, some lower level codes were added. Where existing lower level codes could not be applied to Case Study B data, it was often indicative of a difference in methods and practice between the two case studies. The final thematic frameworks for both case studies are available in Appendices 22 and 23.

The next stage of analysis, undertaken alongside the development and refinement of the thematic framework, was indexing. Indexing referred to the application of the codes outlined in the framework to the case study data, including interview transcripts, observational transcripts and field notes, and documents such as presentations and emails between informants. Materials were coded by line or passage, with consideration of the meaning of each instance of data and how it related to the thematic framework and data coded previously.

Indexing was carried out through the use of NVivo software, which facilitated the organisation and coding of large amounts of data (Srivastava and Thomson, 2009) and easily allowed data to be compared within codes, to explore whether existing labels were appropriate and whether further lower level codes were required. This comparison of data also allowed the author to gain an early overview of the findings generated within each theme, particularly the methods used to complete each stage of the modelling process, and differences in informants' opinions on particular issues.

The penultimate stage of the framework analysis involved charting the data, specifically 'lifting' it from its original context and rearranging it with reference to the themes identified during indexing (Richie and Spencer, 2002). The charts in this research were organised according to the codes in the thematic framework, and entries were made for the relevant data generated via each method of data collection, i.e. observations, interviews, documents and 'other', including think-aloud and diagrams. This typically allowed what had been observed at each stage of model development to be directly compared to specific comments informants made in interviews. The author was therefore able to compare informants' reflections on particular aspects of model development, and also confirm that the practices observed had been understood and interpreted correctly. In each section of the chart appeared a summary of the data captured by each type of method in regard to a particular stage or issue. Data were typically summarised according to key information on practices used to complete stages of model development, problems that occurred, and informants' reflections on both. The charts also contained a summary section for each code, synthesising all important information gained. Separate charts were created for the two case studies, although the codes used were almost identical, so findings were easily comparable. An excerpt from the framework created for Case study A is demonstrated in Figure 3.

Figure 3: Excerpt of framework for Case Study A

	Pathway plotting	Where patients go and why
SUMMARY:	Pathway decided through discussion with C1. Modellers wanting to learn what happens to a typical patient in clinical practice. Asks for pathway for intervention and how comparators differ	Pathway plotting concerned with finding out the order of treatments that patients receive, and what causes patients to receive particular treatments instead of others.
OBSERVATIONS:	I1 asks C1 for patient pathway - <i>"as in the patient would have this and then this would happen and then they get treated"</i> . Both modellers ask C1 for the "general case" or "standard patient" pathway.	Modellers ask clinician what drives (and divides) patients' pathways. Modellers ask: whether treatments sequential or non-sequential, how pathways/treatments differ for patients with different symptoms/diagnoses.
INTERVIEWS:	I25: <i>"I wanted to understand what follows what and what if something doesn't happen, what actually drives the decisions made in the clinical pathway"</i> .	
DOCUMENTS:		
OTHER: e.g. think-aloud, diagrams	Diagram of structure referred to throughout. Appeared useful for clarification of where interventions occur in pathway and for directing discussion.	

The final stage of analysis was mapping and interpretation. Richie and Spencer (2002) referred to this as the process of 'detection', where the author interprets the data collected in relation to the original objectives of the research. This involves reviewing and comparing aspects of evidence, and structuring it according to underlying, important issues (Richie and Spencer, 2002). Although the objective of the case study research was to record and compare the processes and methods used by different modelling teams, it was also to explore why modellers and clinicians undertake particular practices and encounter certain problems.

4.4.4 Ethical considerations

Ethical considerations for the case study research followed similar processes to those outlined for the qualitative interviews. Again any potential risks to informants were assessed prior to commencement, with ethical issues being discussed with those involved. As with the first phase of the research, informants were not deemed to be at risk of any harm from participating, but steps had to be taken to protect the identities of the modellers and clinicians involved, and the modelling organisations. Due to the volume and richness of data collected, there was a risk that individual or combined aspects of evidence could reveal the identity of the modelling team under study. Hopf (2004) argues that particular care is needed when disseminating the findings of empirical research undertaken within an organisation, ensuring that workplace or regional contexts are not recognisable to the public. Therefore within all written reports of the data, including thesis chapters and publications, all characteristics that could potentially identify the context of the case studies were anonymised and replaced with a code. This included informants' names and job titles, the names of others involved in model building, and the details of the medical conditions that were the subject of the models. The organisations were referenced to only as academic (university) for Case study A, and policy-based for Case study B. To ensure further that the organisations remained anonymous, case study informants were asked to review thesis chapters for identifiable information that should be removed. This step was also important to the ethics around informant ownership of data, and the credibility of research, as Baxter and Jack (2008, p556) suggest that 'member checking' allows informants to discuss what has been written, and whether they feel that findings are a fair and accurate representation of what was observed.

Raw data, such as observational and interview transcripts, were anonymised as far as possible. This involved coding the identities of informants, locations and disease references. As a precaution, all data and documents were stored on a secure, password protected computer

network that could only be accessed by the author. Further, the personal information of the informants was stored separately from any transcripts. With the qualitative interviews, informed consent was viewed as a continuous process, whereby separate consent was taken for the observation and interview elements, and informants were reminded each time that they were free to withdraw from the research.

4.4.5 The case study informants

Two case studies were carried out with teams of modellers and clinicians. Case Study A involved the development of a model within a UK university. Those immediately involved in model development were two health economic modellers (I25 and I1) and one clinical expert (C1). The modellers held different roles in model development and were identified as having different levels of expertise. Whilst I25 undertook the ‘hands on’ model building, I1’s role was in a supervisory capacity. Although I1 did not get involved in the ‘technical’ development of the model, she took an active role in clinical meetings with C1 and modelling meetings with I25, and in the model’s structural development. I25 identified this as the first modelling project that he had worked on, whilst I1 had supervised the development of many models previously. C1 had no previous knowledge or experience of health economics, but was Principal Investigator (PI) on a clinical project that required modelling expertise. I25 and I1 were therefore recruited by C1 to develop a model, as the required economics element of the clinical study. Immediate model development involved an additional member of the team, referred to as Modeller 2, who was identified as an expert modeller. A wider team of clinical collaborators and statisticians were also involved in the RCT and model development, providing clinical input and statistical expertise where necessary. The informants referred to in Case Study A and Case Study B are presented in Table 4.

Table 4: Case study informants with their roles in model development

Case Study	Informant	Role in model development
Case Study A:	I1	Modeller (supervisory/senior modeller)
	I25	Modeller (main modeller)
	C1	Clinician
	Modeller 2	Modeller (ad-hoc involvement)
Case Study B:	I8	Modeller (supervisory/senior modeller)
	I26	Modeller (main modeller)
	R1	Researcher
	C2	Clinician (interviewed)
	C3	Clinician (interviewed)
	C4	Clinician (interviewed)
	C5	Clinician
	C6	Clinician
	C7	Clinician
	C8	Clinician
	C9	Clinician
	C10	Clinician

The focus of Case Study B was the development of a model in a UK policy context. The modelling team consisted of two health economic modellers, two project managers and three researchers, whose roles were focused on identifying data for the model. The two modellers held different roles in model development and had different levels of expertise. Whilst I26 reported this to be the first model that he had worked on since his Master's degree, I8 had developed many more and thus was supervising I26 who undertook the 'hands-on' modelling. The clinicians involved held various clinical roles within Medical field 2. The extent of a clinician's involvement in model development was subject to their ability to attend meetings

and contribute to discussions. Although model development involved twenty clinical experts, the research focused on the opinions and reflections of three clinicians (C2, C3 and C4) and two modellers (I8 and I26). The three clinicians interviewed each had some experience of health economics, with C3 having been involved in the development of other models, and C2 and C4 having worked on projects with a cost-effectiveness element.

4.4.6 Presentation of findings

The findings of the qualitative interviews and case studies are presented in the next two chapters of this thesis. Chapter 5 provides the analysis of responses of the interview informants on the model-building process, and issues around model development. Chapter 6 presents the findings of the two model-building case studies, including analysis of observational data on the stages undertaken in model development, and interviews with modellers and clinicians for their reflections on the practices used. In these chapters, quotations from informants or observational data are presented verbatim. Ellipsis (...) is used to indicate where any text has been omitted to shorten or simplify quotes. ‘Umms’, ‘errs’ and repeated words were omitted without the use of ellipsis.

CHAPTER 5: QUALITATIVE INTERVIEWS WITH MODELLERS ON THE MODEL- BUILDING PROCESS

5.1 Introduction

This chapter presents the analysis of the responses of the modeller informants on the model-building process. It covers the informants' experiences of modelling, accounts of the modelling process, clinician involvement in model development, reflections on model building, opinions on modelling guidance, and avenues for future research. Given that clinician involvement appeared to be such an important issue, it is discussed in relation to the relevant stages of the modelling process, but also considered in-depth as a separate theme. Suggestions for future research related to what the informants thought it important to explore within the model-building case studies, presented in chapter 6. The implications of the findings of these interviews in relation to the case study research are considered in the Discussion chapter (7).

5.2 Experience of modelling

Differences between informants' experiences of modelling, other than those that were purposively sampled for, became apparent through the analysis of their responses. These differences are demonstrated in Table 5, and displayed by number of informants.

Table 5: Informant's backgrounds and modelling work (by number of informants)

Informant modelling background		Informant's first experience of modelling		Nature of informant's role		Focus of informant's work		Type of models worked on		Number of models worked on		Disease area(s) modelled in	
<i>Economics/health economics</i>	13	<i>During postgraduate degree</i>	15	<i>Supervisory only</i>	8	<i>Modelling only</i>	15	<i>Markov models only</i>	6	<i>1-5 models</i>	8	<i>Specific disease area</i>	13
<i>Mathematics/modelling</i>	7	<i>At work</i>	9	<i>Hands-on modelling</i>	10	<i>Modelling and other areas</i>	9	<i>Individual sampling models only</i>	2	<i>6-15 models</i>	5	<i>Various disease areas</i>	11
<i>Clinical/Medical</i>	4			<i>Combination of hands-on and supervisory</i>	6			<i>Various model types</i>	16	<i>16+ models</i>	11		

In terms of informant background, most were economics or mathematics related; however, a small number originally studied a clinical discipline. Informants were divided between those who learnt to model whilst in employment and those who learnt on a postgraduate programme. The nature and purpose of the modelling work undertaken varied, as academic informants typically engaged in research-orientated activities, and the non-academic informants mostly worked for modelling consultancies. The aim of the work of the consultancy informants was to develop models for submission to decision-making bodies, on behalf of pharmaceutical companies. One informant, I16, worked directly for a pharmaceutical company, where his role concentrated on the design of models, and ensuring that structures were fit for submission to decision-makers. Two of the non-academic informants, I7 and I8, worked for UK policy institutes, and generated research to directly inform healthcare policy.

The strategy of sampling for senior and junior modellers generated a divide between informants typically modelling in a supervisory capacity, and those doing 'hands on' model development. Senior academic informants were mostly in a 'supervisory only' role, focused on overseeing the development of a model without being involved in its programming, whilst the role of the junior informants was typically concerned with the 'everyday' conceptual and technical development of a model. However, most senior non-academic informants appeared to still engage in some 'hands on' modelling, although their role mostly involved supervising and contributing to the work of others. Similarly, there were a few junior informants who undertook some supervision of others.

Informants' working environment and seniority seemed to affect how much their current role focused on modelling. The non-academic and junior informants were more likely to be working on model based projects only, whilst senior academics typically reported having to undertake projects in additional research areas. Collectively the informants had modelled in a

range of disease areas. The non-academic consultancy informants discussed developing models in a variety of conditions, whilst over half of academic informants' modelling work appeared to focus on a specific disease area. Unsurprisingly, those who had built models in a number of disease areas also tended to have more experience of working with different model types. Generally the non-academic informants had worked on higher numbers of models than those in academia.

5.3 The modelling process

Informants offered descriptions of the typical processes that they used to develop a decision-analytic model. Their processes appeared to follow the same three broad elements: developing the structure, populating the model and model checking. The first section looks at the informants' definitions of the structure of a model. Presented next are the stages and methods referred to by informants in structural development.

5.3.1 Defining model structure

Most informants offered a similar definition of the structure of a model. Almost all referred to the structure as the pathways or the health states and outcomes that were related to a patient's experience of a particular disease and its treatment.

I7: *“that particular intervention and its effects and if there are any adverse effects that come along...”*

I8: *“... a pathway of different events in the correct way for a given population or treatment”*

Most informants gave the impression that the pathways and health states used in a structure should reflect what happens to patients in reality, including the intermediate and long term outcomes that result from disease progression or treatment.

I19: *“...we are trying to mimic the human physiology”*

I10: “...a set of intermediate outcomes, so change in blood pressure, a set of final outcomes, death or heart attack”

However, a few of the non-academic, consultancy informants suggested that although based on clinical reality, their structures reflected the health states and outcomes considered most important. I14 and I11 both implied that this referred to the health and quality of life outcomes that were likely to have the greatest impact on patients, but also on the health economics output.

I14: “...what’s important clinically...or it can be what’s important because it clearly matters to the patient...”

I11: “a patient reported outcome study found the thing that made the biggest difference for the patient and their quality of life was actually whether they were in a wheelchair or not, so the structure ...[was based on] wheelchair use because that’s what’s relevant”

While most informants confined their definitions of structure to a representation of patient pathways and health states, a few informants, both academic and non-academic, suggested that the structure should incorporate the data used to populate it.

I6: “I see it as one big thing, you’ve got your data and you’ve got a mechanism for synthesising that data and turning lots of parameters into costs and QALYs...”

5.3.2 Developing model structure

Informants discussed using other model structures, clinical literature and clinical data to inform and develop their own model structures. However, there were some differences in opinion among the informants as to how much influence these sources do or should have. The next sections outline the methods used by informants to gather information that would form the basis of model structure.

5.3.2.1 Other model structures

Over half of informants discussed using existing model structures to inform structural development. Most informants mentioned using other structures as a source of inspiration for

their work, generating ideas about how other modellers had represented a particular disease or intervention, and what appeared to be good and poor practice.

I6: *“the easiest way to get a handle on a condition and how you’re going to model it is to see if there have been any other models developed...”*

I12: *“NICE criticise models and so you take those on board and you develop your model to take into account those criticisms”*

However, a smaller number of informants discussed using existing structures more prominently in their work, taking an existing structure and developing it further, rather than building an entirely new model. Both of the informants below suggested that using existing models was a more efficient way of working.

I10: *“sometimes there’s a model that’s already there, we have complete access to it... [We] don’t reinvent the wheel”*

I2: *“you can use other models, improve on them, develop them, and re-parameterise them...you don’t want to reinvent the wheel”*

This said I17 and a few others gave the impression that the need to use other structures depended on the complexity of the disease area being modelled, and the quality and standard of other models available.

I17: *“[if] they don’t look like they’ve been really well accepted or we don’t believe that they are robust then we would start with a blank piece of paper...”*

Just under half of informants who discussed using other structures, thought that the search to retrieve them should be systematic. I21, for example, argued that a robust search of the literature should be undertaken to ensure that the new structure is valid in light of the work that others have done.

I1: *“We do a [systematic] review of the papers out there, looking at different model structures, always...”*

I21: *“people ask you ‘how valid is your model?’ and you have to have some convincing answers, when you’ve gone through all the literature you will know what all the papers have done so far...”*

5.3.2.2 Clinical literature

Over half of informants referred to using the clinical literature as a resource to inform structural development. This referred to clinical information on the natural history and treatment of a disease. Sources included clinical papers, clinical patient information, and NICE clinical guidelines. Most of these informants gave the impression that reviewing such material gave them an initial and essential understanding of the disease that they were modelling.

I7: *“you have to research the clinical area first; you can’t just make a stab in the dark...”*

I19: *“...delved into reading about a disease itself...studying it to the point where I feel like I’m good enough, well-versed in the natural history of that disease”*

Other informants discussed using the clinical literature to directly inform structures, looking for and including information such as disease symptoms, health outcomes and adverse events to represent states or pathways.

I14: *“...looking for what are the symptoms of these things, what adverse events are important to a patient?”*

I10: *“what are the long-term outcomes of a child having a [particular infection type] ...they might get [serious health outcome], those are things that you want the model to be able to cope with”*

From a non-academic consultancy perspective, I13 discussed using the clinical literature to determine the model population, looking for the patient group in which an intervention is most clinically and cost-effective.

I13: *“you often find people are using the drug in other countries in a particular setting, that’s a good indication of what their clinicians think is their best subgroup”*

5.3.2.3 Use of clinical data

Just under half of informants used clinical data in the development of their model structures. This referred to the data on the treatment and intervention effectiveness, and the probabilities

associated with patients experiencing particular patient pathways, health states, outcomes and adverse events. These informants implied that they would build a structure using only the health states or pathways that they had clinical data for. These were almost all Canadian academic or UK non-academic, consultancy respondents.

I21: “...*you build your model based on your data*”

I11: “... *first step is sifting through the data, seeing what we can do [with the structure]*”

In contrast, the remaining half of informants appeared to develop their structures prior to knowing the data available to populate them. In reference to the earlier comments of these informants when defining model structure, the suggestion was that they believed that a structure should be based on the health events that could happen to a patient in real life, rather than those for which there were available data. Some of these informants then talked about having to make simplifications to their structures after later realising that they do not have the data needed to populate the original pathways or states (see section 5.3.2.7).

I10: “*then you begin looking around for the data to fill the model that you’ve created*”

I3: “*you build a model and then you start checking whether the data you have, to what extent it can fit the model that you have built*”

A few informants discussed whether they considered it appropriate to develop a structure based on the data used to populate it. While I6 appeared to be of the opinion that it was a standard and inevitable practice, I17 argued that it could bias the results of a model as the structure would include only selected health events and outcomes. I17 seemed particularly wary of developing the structure on the basis of data provided by the pharmaceutical company, as this would likely favour their intervention. This said, I17 also stated that structural development had to be pragmatic, and that a structure may have to be modified in light of the data available to populate it.

I6: *“people get really hung up on having a structure and having the data... I think our attitude is slightly different, the data informs certain things so you may be adapting the model, changing the structure to reflect that data...”*

I17: *“I’m never a fan of building models just around the data you’ve got because that can lead to all sorts of biases especially if the data is provided by the manufacturer of the product you’re evaluating...that said if you don’t have any data available then you have to be quite pragmatic”*

5.3.2.4 Drafting the structure

Most informants suggested that they drafted a structure before implementing it into a software platform. A number of informants described the methods used to develop an initial structure, indicating that they linked relevant clinical and quality of life outcomes to the pathways and health states that patients had to experience to reach them.

I8: *“I try and match up [the] key events that would get to [primary] outcomes...”*

Most informants indicated that they drafted a structure in diagrammatical format, to record how health events and outcomes linked together, and to facilitate making early changes to the structure. The majority of informants seemed to design their initial structures using pen and paper.

I13: *“map out the health states that you are interested in modelling and that are relevant to your question and then the relationships between them, this is a really rough bubble and arrow kind of diagram”*

I20: *“just sitting down with a bunch of sheets of scrap paper and drawing and crossing out and redesigning”*

Model specification

A quarter of informants discussed developing a model specification document as part of designing the initial structure. The specification document typically outlined how the modellers planned to develop a structure, including what information the structure was based on, and the data necessary to populate it. Only one informant who referred to this stage was an academic modeller.

I11: “...introduction and background and then a proposed model structure with a diagram explaining where we are going to get the data sources from, [and] how we have arrived at that structure...”

The model specification document appeared as a formal aspect of most non-academic processes, used to communicate the planned development of a structure to a client and other parties involved. I14 stressed the importance of clients agreeing to the document’s content, to prevent them from wanting to change aspects of the structure after the model had been developed.

I14: “... the client will sign it off, the reason is you’ve signed it off you can’t then come back to us and say, ‘I want you to do it all again’...”

The specification document appeared central to I16’s involvement in model development. I16 gave the impression that he used the document to communicate with model consultancies about the important aspects to capture in a structure, for example, where different groups of patients required different pathways or health states.

I16: “we think there might be a differential response rate between patients so [we say] ‘you’re going to have to build that in somehow’ ...so the spec is what we think needs to be captured...”

There were some contrasting opinions of the benefit of model specification documents among those who used them. Most informants seemed to perceive it as a necessary stage, including I12, who suggested that it facilitated understanding between those involved in model development. However, a few informants viewed the document as restrictive in the sense that it required a lot of work and information to be available at the beginning of the process. I17 indicated that “signing-off” the structure at such an early stage was impractical as it was likely to have to change in light of data available.

I12: “you’ll have difficulty explaining to the client what a model does if you’ve not got a document there that details the processes...”

I17: “...it’s no good finalising your protocol, signing it off and then six months later finding ‘oh right, we haven’t got a utility for that health state’”

5.3.2.5 Clinician involvement

Almost all informants involved clinical experts in their structural development. Informants suggested that the role of clinicians was to inform the modeller about the pathways that patients typically followed or health states that they experienced when they had a particular disease. This included providing information on the interventions and treatments that patients received, and the outcomes of these.

I18: “get some expert clinical input into the...different pathways that patients could go through”

I16: “find out about what happens in this particular condition so ‘how is pancreatic cancer, for example, managed currently? What’s a typical patient course from diagnosis to treatment to treatment failure to progression...?’”

A number of informants, including I9, additionally implied that it was the clinician’s role to converse with the modeller over how disease pathways or states could feasibly be represented within a model framework. Informants suggested that this involved the clinician informing the modeller of the clinically important events and outcomes to include.

I9: “simplify reality in a way that’s simple enough but not too simple, you need a clinician to help you out with that”

I5: “clinical input would guide you as to what might be the important factors to include in your model”

However, there were differences in practice between the informants, regarding when and how clinical input was sought. Just over half of informants gave the impression that they drafted a structure in conjunction with clinicians, whilst others suggested that an initial structure was developed independently, later receiving expert validation. Almost all of the Canadian academics used the latter practice.

I6: “it’s getting the clinicians into a room, drawing lots of diagrams...”

I15: *“I build a framework of the structure...and then I go to the clinician and I arrange a meeting and we go over the model...”*

Most informants gave the impression that they met clinicians face-to-face when developing the initial structure. The methods used by a number of them appeared to be rather formal, with discussion of setting up focus groups and workshops to determine aspects of a structure. Two informants, both UK academics, discussed on occasion using a Delphi style format, which involved structured methods for retrieving and combining clinical opinion on what a model structure should include. One informant suggested that Delphi was required where there was variation in treatment practice, whilst another implied that Delphi was necessary when developing structures in complex disease areas. I9 suggested that the Delphi methods that she had used had been very formal on the one hand, using recommended techniques and processes, and less formal on the other, using a Delphi type approach to contact a broad range of clinicians for ad-hoc input.

I5: *“getting small focus groups together of clinicians, toing and froing about structure...”*

I9: *“...some of the early economic models we’ve done, talking to the clinicians we’ve actually formalised that, we’ve done some Delphi processes because the current practice is so disparate that it’s actually quite difficult to actually get your comparator set... or surveys, we’ve used surveys as well just to get an idea of what we were looking at...so we just wanted a snapshot of how people use that particular drug....”*

A few informants, who discussed involving clinicians in initial structural development, also discussed situations in which they would not involve clinician input, namely if building a model in a disease area that had numerous previous structures. Of note, this mostly applied to the non-academic, consultancy informants.

I17: *“If it was a really well established disease area... we probably wouldn’t speak to a clinician at all at the model design stage”*

I14: *“if there’s nothing that has been done before or if what has been done before doesn’t make sense, it is then, ‘okay we’ll talk to the clinicians’...”*

Only two informants, both Canadian academics, appeared to never involve clinicians at the conceptualisation stage. Both informants implied that a clinician's first involvement was after a structure had been populated. I21 perceived this to be the most relevant time to ask for their feedback, whilst I19 attributed this practice to his clinical background. However, I19 implied that he did not consider it to be good practice generally to seek clinical expertise at such a late stage in the process.

I21: *“you don't speak to them because you first have to have something to show them. Their job is to validate your model or to suggest something but as long as you don't have any data it doesn't make sense to go [and get them involved]”*

I19: *“... I do model building with a pretty good knowledge of the clinical know how's that go into the health condition I am modelling...people expect of me and I expect of myself to be good in the natural history of the disease... [I involve clinicians] very late in the model development process, that's probably not good practice, but it has worked for me that way...”*

5.3.2.6 Model type

Most informants discussed selecting an appropriate model type for their decision problem based on the timeline associated with the disease being modelled. The general opinion was that decision trees were used to represent short-term conditions, whilst Markov structures were used in chronic conditions, where there were repeat events. The use of individual sampling models was discussed less by the informants but also referred to, with the implication that such models would be required in the modelling of complex diseases, to encompass a larger number of health states and events.

I10: *“the primary consideration is first of all the time span of the outcomes, so if it's a...simple treatment, no long-term harms to think about it might be best to be a simple decision tree... [if] you're talking about a chronic issue where someone might be somewhat better by three months, much better by six months and then the device needs replacing at twelve months... 'okay this ought to be a Markov model'”*

I11: *“I can remember we had one in [Condition 9] where we had so many states at that point we realised we needed a discrete event simulation...”*

The most commonly used model type was a Markov structure, with a third of informants giving the impression that due to personal preference or requirements of decision-making bodies, they would default to a Markov structure provided it was suitable.

I4: *“the usual approach would be if you can get away with doing a Markov model in Excel, get away with doing a Markov model in Excel, because it’s simple...”*

I8: *“because the NICE reference case specifies a lifetime approach, you generally start off with the assumption it will be a Markov model unless you can talk yourself out of it”*

Two informants discussed whether they considered choice of model type to be important in structural development. I1 was of the opinion that model selection was important, as it could affect the outcomes and results that a model generated. I24, however, argued that model type only impacts on the time and resources used, stating that different models were capable of representing all decision problems, but in a more or less efficient way.

I1: *“understanding the disease is important, and the fact that model type can make a difference...”*

I24: *“there isn’t a right or wrong, there’s a more appropriate. There’s an [infectious disease] model that I’ve been involved in, which is a crazy Markov structure and I think it’s got a million tunnel states, it’s ridiculous, they’d have been much better off doing an individual sampling model. The model works but it’s just not a very efficient way of doing it”*

5.3.2.7 Making structural assumptions

A third of informants discussed having to make structural assumptions within their models. The majority of these informants gave the impression that assumptions were made in collaboration with clinical experts. Informants cited different reasons for making structural assumptions including lack of data to populate a structure, a lack of knowledge of the disease being modelled, and a need to simplify patient pathways to make them feasible to model. The informants implied that making an assumption often involved removing a part of the original structure, or predicting unknown aspects.

I1: “...we realise that we haven’t got data for all points of the pathway; we try and make a shortcut, an assumption...”

I5: “some diseases are pretty well defined and you know what’s happening, whereas the natural history of some diseases is so unknown...so you’re making huge assumptions...”

I7: “we have to make them because otherwise the model becomes unwieldy and a little bit crazy...”

However, a number of these informants stated that they would only make an assumption on the basis that it would not affect the results and overall outcome of a model. This appeared to involve ensuring that any changes to the model structure would not lead to any important costs or clinical outcomes being missed.

I20: “we assumed that nobody had an occurrence of [specific disease outcome], we knew that was not reflective of reality but we also knew that would add so much complexity to the model. Because it’s a rare outcome and a low cost intervention, it was very unlikely to change the decision point...”

5.3.2.8 Implementation

The majority of informants referred to an implementation stage in structural development, which involved programming a structure into a software platform. Most informants gave the impression that they drafted their structures entirely on paper prior to implementation; however, a few suggested that they used software as a tool to develop their structures. The latter informants implied that the software facilitated their understanding of how the model was working, and the way in which patients were progressing through a structure.

I13: “I’m not one of those people who likes to put everything down on paper massively first before I start putting it into Excel because I find it a lot easier to try and link things and see how they work and develop the structure that way”

I19: “...I have used Excel to get a sense of what are the stages, what are the transition probabilities and ‘if I apply this transition probability to these disease states, what would happen down the road?’...”

The most commonly used software package among the informants was Excel, although TreeAge, R and WinBUGS were also mentioned. Informants tended to have preferred software that they consistently used, although a few implied that usage would depend on the type of model being built. A number of informants complained that modelling software tended to be difficult to run.

I20: *“the program that we used was really slow, it took twenty hours to run and it was a hellacious experience”*

5.3.2.9 Iterative

Almost all informants commented or gave the impression that structural development was an iterative process. The iterative aspect appeared to be in terms of the involvement of clinicians and the cyclical process of having them check a structure, and the modeller making required changes.

I15: *“... first meeting and then going back, change the information from the first meeting and then arrange another meeting until we are happy with the structure of the model”*

For the non-academic consultancy informants, structural development also appeared to be an iterative process, with this iterative process particularly formed on interaction with their clients.

I13: *“...All iterations of the model go to the client, for their feedback...”*

5.3.2.10 Summary: developing model structure

Informants appeared to follow similar processes in the development of their model structures. This included the use of other model structures, clinical literature and clinician opinion to inform their models. However, informants held contrasting opinions on how some of these stages should be undertaken, for example, whether existing structures should be used to directly inform a new model. Further, although most informants involved clinicians in structural development, there were different perspectives on when and how clinical opinion

should be sought. Two of the Canadian academic informants did not involve clinicians at all at the structural stage. Informants held different opinions on whether clinical data should be used to inform a model structure, and it was largely UK informants who suggested that a structure should initially be developed without prior knowledge of the clinical data available. In addition, informants used contrasting methods to plan their structures, as most discussed drawing it on paper prior to implementation, but others stated that they initially developed the structure in modelling software. In terms of the differences between informants' practices in relation to their backgrounds, the processes of the non-academic informants largely included an additional step, involving the development of a model specification document. Further, a few of the non-academic consultancy informants offered a different definition of model structure to the majority, implying that a structure should aim to incorporate only the health events that were of the greatest clinical and economic importance.

5.3.3 Populating the model

All informants discussed the typical processes used to populate a model. This included methods for identifying and collecting data inputs. The first sections look at the methods for identifying evidence for the clinical parameters of the model, and then report on the sources used to retrieve economic inputs, for example, costs and utility estimates. The remaining sections explore how data assumptions were made, and the typical involvement that clinicians had in data identification.

5.3.3.1 Identifying data

Informants provided a variety of different methods for identifying data to populate their models. These included literature searches, primary data collection and use of existing models. Sources of data appeared to be project and context dependent. Whilst non-academic consultancy informants were more likely to have access to primary evidence provided by

pharmaceutical clients, most Canadian informants were able to use government administrative data. The majority of informants though, particularly UK academics, used a combination of primary data and literature sources.

Clinical evidence

Just under half of informants suggested that they conducted a systematic review to inform the clinical parameters of their models.

I22: *“start with a systematic literature review of the clinical effectiveness literature”*

Informants who did not undertake a systematic review for clinical parameters still tended to carry out fairly thorough searches, or looked for existing meta-analyses. One informant stated that he did this because he considered the accuracy of the clinical parameters to be more important than others.

I24: *“treatment effects...we rely on meta-analysis and stuff like that for those parameters that we deem to be most important”*

For those who did not undertake a systematic review or carried out less thorough searches, the most commonly cited reason was a lack of time or resources.

I15: *“there's no time or money to do it in a systematic way...”*

I2: *“you don't have to do a systematic review because that does take some time...”*

A number of informants cited problems with identifying clinical data to populate the comparator (current practice) arm of the model structure. These informants suggested that data on the effectiveness of current clinical practice was difficult to obtain, because clinical trials tended to focus on the efficacy of new interventions, often against placebo or other new interventions.

I21: *“we didn't have transitioning rates for the status quo. Nobody gives you the transition rates for the status quo because all the clinical trials follow the patients who... they impose medication on...”*

I13: *“the comparator that’s used in real life, there’s no evidence for that, [it] happens frequently...”*

Economic evidence

Clinical evidence searches seemed to be more systematic than those used to identify cost and utility data. Searches for economic evidence appeared mostly to be unsystematic, as informants reported using ad-hoc searches to retrieve utility data, and specific sources to obtain costs and resource use. A number of informants discussed searching for relevant economic evaluation papers to inform utility values. However, only a few informants suggested that this search involved a systematic review of the literature, and these tended to be non-academic, consultancy or policy informants, who were following practice recommended by NICE.

I19: *“... start from previous published cost-effectiveness papers to see what values they have used...”*

I18: *“we do a systematic search to get an idea of quality of life”*

I17: *“we’re obliged to do a systematic review because we have to follow the NICE Methods Guide”*

Informants reported mostly identifying costs via government websites and databases, giving the impression that this aspect of data collection was straightforward.

I14: *“it is NHS reference costs, it is BNF or EMIT for costing, costs are usually fairly standard”*

However, a few informants reported having difficulty identifying resource use data, with informants implying that data were not specific enough, or were not generally available.

I4: *“for time that people spend in the clinic, actually going, because they don't publish much detail on it”*

5.3.3.2 Making data assumptions

Most informants mentioned having to make data assumptions to populate parameters for which there were no data available. Data assumptions typically involved the use of clinical

opinion; however, other methods were mentioned such as the use of proxy values for utility parameters.

I7: *“if you didn’t have quality of life data for a particular condition, whether a similar condition would be a good proxy for that”*

A number of informants stated that assumptions should be validated by clinicians and tested using a sensitivity analysis. Many informants also stressed the importance of assumptions being transparent.

I1: *“[We] tested the assumptions afterwards to say 'well look, if we'd assumed something else what difference would it make?'...”*

I5: *“if assumptions are made they are documented as assumptions and we tend to do a huge amount of sensitivity analysis around them”*

5.3.3.3 Clinician involvement

Around half of informants stated that they involved clinicians in identifying and/or collecting data to populate models. A number of informants gave the impression that it was the clinician’s role to provide data, or indicate where they could find values themselves.

I10: *“drilling into their expert knowledge of the clinical literature to find a particular bit of information”*

I23: *“they can provide data if it’s an area that they’ve worked in a lot before”*

Clinician opinion was most commonly used to inform data assumptions. The informants reported on a variety of different methods for gaining clinical input on assumptions, ranging from informal to very formal techniques. It was suggested that the method used could differ according to a particular project, and was dependent on factors such as time and budget.

Illustrative of the approach taken by the majority of informants, I19 and I24 described asking clinicians in an informal and ad-hoc manner for data values to populate parameters.

I19: *“it has been fairly unstructured...asking through email or phone call ‘what do you think the impact of this treatment is?’”*

I24: *“I would go to physicians and say... ‘what is the max it could be, what’s the min it could be?’ and fit some distributions around that”*

Of the informants who discussed having used formal expert elicitation methods, for most this appeared to be in the context of one-off projects. However, there were a few informants who gave the impression that they always take a formal approach. The formal aspect seemed to refer to how data were collected and validated, and the methods used to combine clinician opinion.

I15: *“I build a spreadsheet telling the experts what is asked of them ... once I’ve collected the data... I’ll show everyone ‘these are the anonymous results of your answers and these are the combined estimates’ ...and ask if there’s anything that they would like to change...I use a different process of combining the experts’ data together that captures the wide range of opinions...it doesn’t reduce the uncertainty across all the different experts...”*

However, a number of informants were sceptical of the use of clinical opinion to inform parameters, and suggested that it would only be used as a last resort. One informant seemed to question the validity of clinician responses, questioning how experts select values in these contexts.

I22: *“I am deeply sceptical about expert opinion, deeply sceptical, so because I know a little bit about behavioural economics, I know expert opinion can be wildly wrong...”*

5.3.3.4 Summary: populating the structure

Informants’ methods for identifying data seemed dependent on the nature of their work, and who the model was for. Searches for clinical evidence appeared to be more robust than those used to identify cost and utility data, as most informants reported using systematic methods or existing evidence reviews. Utility and cost data were mostly retrieved through ad-hoc searches, with few informants stating that they undertook a systematic review for economic evidence. Those who did were mostly non-academic informants, and were doing so to adhere to the NICE Methods Guide. Time and lack of resources appeared to be factors in the

modellers' decisions to use less thorough methods. Informants reported using specific sources to identify cost data, but implied that these did not include enough detailed information on resource use. Clinician involvement again appeared to be important in this aspect of model development; however, some informants criticised the use of clinical opinion to make data assumptions. This was in the context of the majority of informants currently using informal and unstructured methods to collect the clinicians' inputs.

5.3.4 Model checking

All informants discussed undertaking checking activities within their model development. Most described model checking as a process, involving a number of different stages, namely sensitivity analysis, internal validation, external validation and structural validation. These are reported on next, as are findings related to the involvement of clinicians and other experts in model checking.

5.3.4.1 Sensitivity analysis

Almost all informants discussed undertaking sensitivity analysis on the results of their models. The majority of informants implied that they applied sensitivity analysis to uncertain data parameters, including data assumptions.

I18: *"[we carried out] sensitivity analysis...because we had no great confidence in that initial estimate"*

A number of informants stated that they would vary all parameters to note the effect that it would have on model results. The majority of these informants indicated that this would involve a full probabilistic sensitivity analysis.

I12: *"sensitivity analysis throughout all of the parameters that you entered into the model"*

I3: *"...we made it probabilistic so all the variables that at the beginning were point estimates, we just included the corresponding distribution that was associated to each one of them..."*

5.3.4.2 Internal validation

Internal validation was the most talked about and, seemingly, the most undertaken check among the informants. Internal validation appeared to involve testing for errors within a model, to ensure that everything was working as it should be. Informants referred to this as ‘debugging’ or ‘internal consistency checking’.

I15: *“it's a matter of looking for bugs, doing internal validity...”*

Run the model

Many informants suggested that they ‘run’ models within a software platform as a validation check. This involved programming ‘extreme values’ into model parameters, and ensuring that the model results changed in an appropriate way.

I6: *“stick in massive numbers and small numbers and see what happens... see if the results go the way you expect”*

A few informants stated that they run their models to confirm that they are working, and that patients are progressing through the structure as intended.

I12: *“making sure it works as it's supposed to work... 'are patients going through the health states as they are supposed to or is something happening, are you losing patients?’”*

‘Eyeballing’

Just under half of informants indicated that they check their models by examining in detail the inputs and formulas that have been programmed. Most suggested that this was an exhaustive visual check of the data within the software platform.

I2: *“make sure you've typed the right things, the right numbers into the right places on your spreadsheets”*

I11: *“go through on a sheet by sheet basis... ‘does this formula make sense, does this one make sense?’”*

Face validity

A number of informants discussed looking for face validity within a model, apparently assessing whether clinical inputs were a valid reflection of what happened to patients in reality. This appeared to require knowledge of the nature of a clinical condition, and the prognosis of a typical patient.

I13: *“we will look for face validity, so we will look for whether utilities look sensible given the context of the disease”*

Reprogramming the model

A few informants stated that they will program a model twice, checking for inconsistencies between the two versions. Any discrepancies may highlight errors in the original model. For most this involved building a second model in alternative software.

I9: *“if I build it in Excel, if I can and if it’s feasible I will try and build it in something like TreeAge as well to check the simple maths is working right....”*

I21 was the only informant to state that he reprogrammed aspects of his model in the same software.

I21: *“I double up some codes and run my model simultaneously on different computers, and then I compare the results...”*

Most of these informants gave the impression that this is not an activity that they carry out on every model due to a lack of resources.

I24: *“I don’t do it all the time because of time and resource constraints”*

Plausible results

Around a third of informants included a stage of model checking that involved an explicit assessment of whether a model’s results seemed sensible. This was described as a phase of reflection in terms of whether informants felt that results were consistent with what they were expecting. Similar to face validity, this check appeared to require the informant’s knowledge

of the disease and interventions being modelled. It was implied by the majority of these informants that if the results were not as expected they would undertake further checks on their models.

I19: *“always make sure that I am getting some results that are sensible to me, it is very important in detecting bugs in your model...”*

I2: *“if you see a result that you don’t believe, the first thing to do is to check to make sure the model’s working right”*

5.3.4.3 External validation

Less than half of informants discussed external validation, which involved confirming a model’s results against relevant clinical and economic sources. Around half of these informants suggested that if differences were found between the model results and other clinical findings or economic studies, they would need to be justified.

I20: *“make sure that the findings that you get are in line with the findings from the medical literature, so if you find that people who have taken drug X vs. drug Y live an average of six months longer and the finding in the literature is that they live two years longer, you’ve done something wrong ...”*

I1: *“produce results that aren’t wildly different from someone else’s, and understand if they are different what was new that we put in our model that was likely to drive the difference”*

I3 was the only informant to refer to an alternative method of external validation, which involved presenting and comparing model results to the opinion of experts working on similar projects. She implied that this was to identify discrepancies and possible errors.

I3: *“in the last two years I attended at least ten conferences with the research of the project presented...I discussed the results of my model with people who are working in ...other areas in similar topics, so this is a way to compare the results we are gaining...it helps us if there is something that is not right, someone may say something or give us comments that will identify these things”*

Of the informants who did not undertake external validation, a few indicated that this was due to this type of check being too difficult and time consuming.

I20: *“there are some models that are built that are widely valued against external data and that’s what you should do, but that’s an incredibly laborious process especially when data are not available or accessible”*

5.3.4.4 Structural checks

Only one informant discussed undertaking analysis on the structure of a model as a means of addressing uncertainty. However, this informant gave the impression that this was a one-off because she was unsure about particular model findings. The nature of the specific check carried out was, however, unclear as it was conducted elsewhere.

I11: *“recently we sent one to [Modeller 23] at [University F] because they are quite heavily involved in structural uncertainty, to get her opinion on it”*

One further informant referred to structural sensitivity analysis but only to indicate that he did not undertake it. This was due to the apparent difficulty associated with doing so.

I17: *“always do lots of sensitivity analysis on the numbers within a model but we never do sensitivity analysis on the structure of a model because it’s too hard and it takes too long”*

5.3.4.5 Clinician involvement

All informants referred to clinician involvement in validating their models, and for most this was the only form of structural checking undertaken. For many informants this involvement was iterative in the sense that clinicians were validating models at the structural development, data population and results stages. In terms of checking model structure, the implication from a number of informants was that a clinician would feedback on various drafts until they reached a version that they were happy with.

I8: *“I’m taking notes on what I’m interpreting from that meeting and then I’ll follow it up with a more specific meeting, ‘I’ve understood this to be the case, do you agree?’”*

It was suggested by a few informants that receiving clinician ‘sign-off’ on a structure was an important stage of the process. These informants suggested that formal clinician agreement

was a method of structural validation, to ensure that the clinical aspects of the model were robust to external critique.

I24: *“I don’t do too much on the structure of a model if I’ve got sign off from the clinician about how it’s working...”*

I4: *“the main thing from our perspective is to be able to justify it; if you’ve got some justification for the clinical parts of your model...I’m covered...”*

In terms of checking the data used to populate a model, around half of informants indicated that clinical opinion was used to validate data inputs or assumptions. As with the development of structure, the impression from a few informants was that it was important to have formal ‘sign-off’ from clinicians on these values.

I12: *“...clinical opinion, you’d be foolish just to make an assumption and not have anything to back it up”*

I1: *“...I got her to put in an email ‘these are the assumptions we had to make’, and they’ve emailed back ‘yeah that assumptions fine, we can live with those assumptions’”*

Most informants indicated that clinicians were involved in checking the results of their models. The clinician’s role appeared to involve confirming the clinical findings, ensuring that results were intuitive to their knowledge and practice.

I23: *“we initially run the model and take the preliminary results back to the clinicians...”*

I20: *“it’s whether or not the result completely violates [their] clinical experiences...”*

As one of few informants to involve clinicians only at the results stage of model development, I19 reflected on this practice, explaining that at times, errors have not been identified until very late in the process.

I19: *“...it was pretty much at a production level that one of the [clinical] co-investigators said ‘I19, the -medications that you have put in your model as a treatment, the costs are way too low...’ so that was a big wake-up call at the last minute that we were using [the] wrong values from a wrong source...”*

In terms of methods used to gain clinical feedback, all of the non-academic consultancy informants discussed using advisory boards. For most of these informants, this stage occurred when a model structure had been developed and populated, although a few informants suggested that their advisory boards happened after results were generated. The typical format appeared to involve the presentation of a model to clinicians, clients and other experts, followed by discussion of the model's validity and any suggested changes.

I11: *“an advisory board would generally come at the stage when we've got a structure in place and we are just looking to firm things up”*

I14: *“we'll have the company present the clinical data...then we will present the model, 'this is what it is, these are the methods, this is where the data comes from, these are the results...’”*

5.3.4.6 Other involvement

Almost half of informants discussed the involvement of other experts, aside from clinicians, in model checking. For most informants this referred to other health economists, typically colleagues, who were external to model development. The involvement of these experts tended to be ad-hoc and informal, focused on checking the technical aspects of a model. A lot of these informants suggested that they involved a health economist more senior than themselves.

I4: *“...look at a model together and spot stuff because it's quite easy to make basic errors like using a standard deviation instead of a standard error when you're sampling from a distribution, just bits and bobs that you don't really notice because you're in the zone doing modelling...”*

I3: *“and then you speak to people who are experts, who have been working for a long time... just by having a glance they will see something immediately, identify 'this is not right'...”*

The non-academic, consultancy informants however, involved external health economists formally through advisory boards. Again the suggestion was that it was the health economist's role to validate and feedback on the economic and technical aspects of a model.

I14: *“then the economist will ask his questions and then there will be a big discussion as to whether the model is right...”*

I13: *“... they let us know whether they think what we’ve done is sensible and if there’s any additional sensitivity [analysis] we should do”*

A few informants also discussed the involvement of patient representatives in model checking, implying that this aimed to ensure that a model structure reflected the patients’ experiences of a disease and its treatment. However, most of these informants suggested that patient involvement occurred only rarely within their modelling processes.

I3: *“I have gone as far as speaking to people who have had that condition in the past and went through the treatment...”*

5.3.4.7 Summary: model checking

Sensitivity analysis and internal validation checks were those most commonly and frequently undertaken by informants. This was in contrast to external validation, which was reported by under half of informants, and structural checking, which was not personally undertaken by any modeller. Informants suggested that the reason for not carrying out these types of checks was a lack of time and resources. Clinician involvement appeared essential in validating a model structure, data and results. The non-academic consultancy informants used formal advisory boards with clinical experts to discuss and encourage feedback on all aspects of a developed model. A few informants emphasised the importance of clinicians validating and formally agreeing to the inputs and results of the model for justification purposes. Informants also suggested that the support of other health economists, either formally or informally, was important in terms of checking the technical aspects of a model. Patient involvement in structural validation was seen to be valuable although few informants actually used patient opinion in structural development.

5.4 Clinician involvement

Informants' discussions of the clinician's role in developing, populating and checking a model structure suggested that clinician involvement was an important aspect of model building. Informants spent considerable time during their interviews discussing issues around clinician involvement. The next sections cover clinician recruitment, number recruited and communication between the modeller and clinician(s), particularly focused on clinician engagement.

5.4.1 Recruitment

Just over half of informants discussed the recruitment of clinicians to their modelling work. For half of these modellers, clinicians were the Principal Investigators (PIs) on projects, with the informants therefore being recruited by the clinician(s) to undertake model development. These informants were almost all UK academics.

I2: *"when it comes to health economists very often we don't come up with the research question, it'll be clinicians...it'll only be later that the health economist will come onto the bid..."*

One of these informants reflected on being involved in a project with this set up, implying that when clinicians are co-investigators it is easier to engage them in model development.

I5: *"it was really helpful and it moved things along very quickly, you didn't feel worried about asking the clinicians and you didn't feel as if you were badgering them..."*

The remaining half of the informants implied that they recruited clinicians to their own projects and modelling work. For almost all of the academics in this group, this appeared to involve rather informal methods.

I21: *"I talk to my supervisor and he talks to them..."*

I9: *"we did cold call people..."*

Most of these informants spoke of the difficulty of gaining a clinician's input to a modelling project that they are not formally involved in, due to clinicians' workloads.

I4: *"it's often quite difficult to get a clinician's time full stop...especially if they have a lot of clinical time with patients, it's almost impossible to get hold of them..."*

The non-academic consultancy informants in this group however, appeared to use more formal methods of recruitment, having access to clinicians who worked for their clients.

Similarly, the two non-academic policy informants stated that their clinicians were recruited through formal advertisement.

I14: *"...one who is working for the company that you're working with because they are quite easily accessible"*

I7: *"it's advertised by the [policy institute]..."*

Around a third of informants discussed the type of clinician that they recruited or worked with, with most suggesting that projects involved leading experts. One non-academic, consultancy informant inferred that recruiting senior clinicians to model development leaves a model less open to change and criticism when it is submitted to a funding body.

I1: *"there were three key clinicians who were really specialist..."*

I13: *"you want as many of the leading people in the field as possible because hopefully you have got all the opinions out prior to the [submission] meeting so you're not going to have some clinician turning up and saying something completely different..."*

5.4.2 Number recruited

Almost all informants discussed the number of clinicians typically involved in model development, ranging from zero up to twelve. Most informants involved two clinicians on average, with three and then four being the next most common practice. Only three informants, all non-academic, indicated that they typically involved more than four clinicians. Two informants, I15 and I17, discussed involving only one clinician; however, only I15 gave

the impression that this was the norm. I17 implied that he only involved one clinician in situations where he was building a simple model in a non-complex disease area.

I17: “so typically two or three I would say but in some cases one, if we’ve spoken to one clinician and it’s appeared to be relatively straightforward...it’s not always been the case that we’ve spoken to a second”

I22 was the only informant to discuss having no clinician involvement in the entire modelling process, however; he gave the impression that this practice occurred on very few projects.

I22: “... about 95% of the projects that I’m on there are clinicians involved on the project...”

A number of informants discussed involving different numbers of clinicians on different modelling projects, with the final number determined by the complexity of the disease area or research question being modelled. One informant suggested that additional clinicians were required for models in disease areas where there was variation in treatment practice.

I20: “depending on the complexity of the question it can be as few as two...other times when we’ve had more complex models we sent them out [to additional clinicians] ...”

I9: “... depends very much on the disease and the intervention...I would argue that [Disease area 1] is the area where you have the biggest variation in clinical practice and I think you need bigger numbers there because even within a hospital trust you’d see different practices in terms of current care...”

One informant mentioned using additional clinicians where treatment effectiveness data was limited.

I13: “... for a standard model that’s got good clinical trial evidence I would want to speak to somewhere between two and four, for a model that has practically no clinical trial evidence I would want to speak to more like six to eight...”

Informants discussed the optimal number of clinicians to involve, with many believing that involving only one limited the generalisability of the model.

I7: “...you’re asking one person to make an assumption for you; in their experience it may be completely different than someone else’s...”

I12: *“you could have a rogue, which has happened to a colleague of mine...one clinician view, fine, went to the board and they said ‘no that’s not right, the way he treats patients is completely different to how we do’...”*

However, a few informants also discussed problems associated with involving too many clinicians, namely that it was difficult to manage and include numerous opinions in structural development.

I17: *“...things quickly get very complicated because you end up with eight clinicians who have eight different views and you go away and try and design a model that takes all of those views into account, and you end up with an over-elaborate model that becomes a bit meaningless if you can’t populate it properly...”*

A small number of informants talked about how they synthesise information given by clinicians if there are disagreements. All of these informants suggested that they would encourage the clinicians to agree on a particular aspect amongst themselves.

I7: *“... so they’ll sit and argue for ten, fifteen minutes on particular fine little points for you”*

I6: *“often they just disagree and you say well ‘do you feel that actually you're wrong on one of these or the other one's wrong?’”*

Only one informant discussed the importance of having a range of clinicians involved in model development.

I8: *“... an A and E nurse, a consultant, maybe a surgeon for the onward treatment, they try to get a big range of clinicians all involved with the pathway...”*

In addition, none of the informants indicated that they recruited clinicians from a range of geographical areas or medical centres. Despite this, a number of informants suggested potential issues with the generalisability of model results if model structures were developed without consideration of the practice of different medical teams.

I6: *“I found that with [Chronic condition 2], different teams do very different things....”*

I4: *“there’s about five different ways of classifying the disease, and different regions of this country, different people use different classification systems...and they don’t really map well together...”*

5.4.3 Communication

Almost all informants discussed issues around communication with clinicians. This included the format for meetings between the modeller and clinician, their typical working relationship, and informant opinion and experiences of clinician engagement in the modelling process.

These issues are discussed next, with the inclusion of informants’ accounts of methods used to encourage clinician involvement in model development.

5.4.3.1 Meetings

Just under half of informants discussed the typical nature of meetings with clinicians. For all of these informants meetings were entirely face-to-face, or involved a combination of face-to-face, telephone and internet correspondence.

I21: *“most of the time face-to-face meetings”*

I8: *“... formal face-to-face meetings, but in between is constant email correspondence going backwards and forwards”*

Two informants suggested that the format of communication depended on the stage of model development being undertaken. It was seen as important by these informants that meetings were face-to-face when developing or checking the structure of a model, with I10 suggesting that this stage required a clinician’s complete attention. I17 gave checking data assumptions as an example of where telephone contact would be appropriate.

I10: *“there’s no substitute for face-to-face meetings for...the structure of a model as a whole, that conversation doesn’t work via email because people don’t read their emails as carefully as they should do...”*

I17: *“we do it by telephone call anyway if it was just a case of checking a few data assumptions”*

The majority of informants discussed having regular and structured meetings with clinicians, with further ad-hoc involvement where necessary. However, a number of these informants implied that the frequency of meetings was dependent on the project, specifically the preferences of the PI and complexity of the disease being modelled.

I5: “... the PI on this one was very good and just set up weekly [meetings] ...”

I13: “the [complex disease area] model that I’ve been working on, we had three or four rounds of clinician review, because that one was really very dependent on the clinician opinion of how it was modelling survival”

5.4.4 Relationship

Over a third of informants offered insight into the working relationship between modeller and clinician. There appeared to be a divide in opinion and practice among informants regarding how much negotiation a clinician had over a final model structure. Almost all of these informants suggested that although they would take into account a clinician’s opinion on structure, as the modeller, they would make the decision as to the information that informed the final model.

I22: “although it’s essential to incorporate them into the process, you do have to...protect your role on the project and basically that’s done via direct rationalisation of the inputs...”

I8: “I think they need to be guided personally...I feel like well it’s my role to tell them what’s important or not”

The majority of these informants implied that the modeller must have the final decision on structural development because clinicians were not typically aware of which information was important to model. Both I20 and I4 suggested that clinicians often wanted to include every conceivable clinical event within a structure, rather than just those that were common and would impact on the model results and overall outcome.

I20: “clinicians will always look to the edge cases...there’s some negotiation about which of those differences [between a typical and less typical patient]are necessary to

include, and which of those differences are so rare that they will only overcomplicate the structure and not give you any new information...”

I4: *“they’ll suggest you put this in the model and it’s like ‘I can’t do that’, I can’t take every possible adverse event and put everything into the model, that’s just not possible...”*

In contrast to the practice of others, one informant gave the impression that the clinicians had the final say on what a structure encompasses.

I7: *“as the health economist you look and go okay ‘this...it’s going to be quite important’ and they come back to us and say ‘no that’s not that important’ or ‘this is’”*

I7, a non-academic policy modeller, offered further context to this by implying that clinicians should dictate the inputs to a model because they have to be happy with using its results in practice.

I7: *“they make [the] decisions because at the end of the day a doctor in a hospital can look at the [model results] and just ignore [them] ...”*

5.4.5 Engagement

The engagement of clinicians in model development was an important issue for almost all informants. These informants gave the impression that they had experienced difficulty in engaging clinicians fully in model building. Many informants implied that a clinician’s lack of engagement was problematic in terms of receiving required information from them.

I9: *“if clinicians aren’t engaged it’s impossible because you can’t possibly understand all the nuances of what you’re trying to model”*

5.4.5.1 Understanding

Most informants gave the impression that a lack of clinician engagement in the modelling process was related to a lack of clinician understanding of modelling and health economics.

The suggestion was that clinicians struggled to understand models and the assumptions attached to them. According to informants, clinicians had difficulty discussing the experience

of an average patient, whose treatment pathways are represented within a model structure. A number of informants suggested that clinicians tended to discuss the experiences of patients on an individual basis, similar to how they assess and treat them.

I11: *“this concept of ‘we are massively oversimplifying what happens in this disease and we’re assuming every patient is the same’, that’s quite difficult for clinicians”*

I14: *“the hard thing with clinicians is getting them to abstract because they see individuals, they don’t see a group...”*

I10: *“...so you might be saying ‘how great a probability of infection would a patient need to have before you gave them [treatment]?’ , GPs just don’t think like that, you need them to think like that for the model but it’s not how they’re used to thinking”*

Related to the tendency of clinicians to discuss the individual experiences of patients, a few informants suggested that clinicians typically emphasised rare and uncommon patient events, which for economic reasons would not be included in a structure.

I4: *“they’ll talk about really rare events and really special cases that in their thirty years of experience they’ve witnessed once and they’ll suggest you put this in the model...”*

I12: *“there’s always a case where ‘oh I had a patient who didn’t behave like that’ or whatever, you’re going to get that...”*

However, around a third of informants discussed working with clinicians who had a good understanding of models and the information required for their development. Most of these informants suggested that clinicians had been more engaged and helpful, if they had previous experience of health economics and/or modelling.

I20: *“[structural development] is an ongoing process that depending on how much familiarity the clinician has with modelling...it can take a long time or a very short time”*

One informant discussed clinicians being more engaged if they had received relevant training. I24 implied that the clinicians who had training tended to understand how models worked, and the key information that was required to build a clinically representative but economically important structure.

I24: “...clinicians who have got training in it, who just totally get it and they’re really involved and they just really describe what your assumption is, they understand it implicitly, they know which ones are a big deal and which ones don’t really matter as much, or aren’t really going to influence your results...”

It was implied by many informants that clinician understanding of modelling facilitated structural development. However, I20 was the only informant to equally stress the importance of the modeller understanding the clinical condition being represented.

I20: “the only way to make [structural development] easier is for the modeller to have a really strong understanding of the clinical reality before trying to build the model... for both sides to have more understanding of what the other party is trying to accomplish”

5.4.5.2 Personality

Around half of informants implied that personality and attitude affected how much a clinician engaged with model development. A quarter of these informants suggested that engagement was wholly related to a clinician’s enthusiasm and motivation to work on a particular project.

I21: “some people put more time [in] and they spend more time on your work”

I9: “it depends on the clinician, how interested they are...”

However, the remaining informants suggested that a lack of engagement was often related to a negative attitude towards modelling. Informants implied that clinicians were reluctant to contribute to models as they were distrusting of how they worked.

I7: “some clinicians are very anti health economics; I’ve actually had one ask me why we bothered, which was lovely, ‘what was the point in health economics because it’s too new of a science to be of use...’”

I24: “I’ve found there are two types of clinician...one just distrusts the whole thing anyway and doesn’t really buy into it and thinks modelling is somehow a quirk, quack science...”

5.4.5.3 Vested interest

Informants suggested that while some clinicians will not engage, others become ‘overinvested’ in model development. Around a quarter of informants reported working with

clinicians who had a ‘vested interest’ in seeing a particular outcome from a model. This resulted in clinicians being difficult to work with, if the model was not generating desired results. The implication from many informants was that clinicians wanted to see a ‘cost-effective’ outcome for the intervention being represented.

I15: *“clinicians are fine as long as if they think it’s cost-effective and it’s cost-effective, great, that’s brilliant, if it’s not cost-effective it’s a big problem, that’s when you have the issues...”*

I23: *“...it was difficult to work with some clinicians because they had very strong opinions about what the outcome should be, so when it came to build the model and present the results they weren’t really approaching it from...an unbiased position. That was challenging because it became difficult for us to engage them in the process because those strong opinions were so prevalent...”*

A few informants, including I16, stated that they had to reiterate to clinicians during the process, the importance of being objective in model development, and having an outcome based on unbiased evidence.

I16: *“you may have medical colleagues who think the new drug is the bee’s knees and my role is to kind of dampen that enthusiasm a little bit or...to justify it, so maybe it is the bee’s knees but you have to get the evidence to show it...”*

5.4.5.4 Methods of engagement

Many informants implied that a lack of clinician engagement in the modelling process made model development difficult. In response to this, almost half of informants discussed the methods used to attempt to engage clinicians further in model building. For two modellers, this referred to activities taking place prior to model development, namely offering clinicians training in health economic modelling to increase understanding and acceptance. I7 stated that this involved outlining the health economic context of models to clinicians and similarly to I24, explaining how models worked and why they were necessary. I24 suggested using existing model structures to allow clinicians to visualise how models synthesise information.

I17: *“it’s telling them why health economics is important, especially in the NHS, and how we do things, how models work...”*

I24: *“...I choose one that has lots of states, inputs and transitions and say ‘this is a previous model that we did in disease X, and to choose what the best treatment is you need evidence, and this is beyond most human minds to be able to do that internally and so modelling is a framework to do that’, and they generally buy into it”*

The remaining informants referred to methods used during model development to engage clinicians. For a few informants this referred to techniques used during communication, namely asking them questions in a direct way to ensure a specific answer.

I12: *“you’ve got to make sure that you keep the questions direct so that they can’t divert off into ‘yes there are 99% of people that die within 5 years, however, there’s one patient who lived for 20 years’...”*

For an additional few informants, showing clinicians a diagram of the model, or its structure within the software, appeared to encourage clinicians to take a more detailed interest.

I13: *“when you’re going through the model physically they get a bit more involved, so it’s just like ‘oh can you click that, can you show me the other’...”*

One of the informants however, reflected that despite efforts to engage clinicians further in model development, they often do not want to spend time learning about models.

I23: *“the feedback that we got from clinicians was...they didn’t really feel like it was necessary to know all that much about models in their day to day practice...they weren’t really interested in pursuing more in that area”*

5.4.6 Summary: clinician involvement

Informants’ methods for recruiting clinicians appeared related to the context of their projects, with some modellers reporting difficulty with gaining external clinical input. Questions were raised regarding the appropriate number of clinicians to include in model development, and whether, for reasons around generalisability of structure, modellers should be looking to involve clinical experts working in different roles and for different healthcare practices.

Informants typically involved two clinicians in structural development, giving the impression

that speaking to only one would bias a model, and too many would overcomplicate a structure and the development process. However, a few informants suggested methods for managing clinician discussion and disagreement, proposing that clinicians should negotiate amongst themselves on the content of a model.

The majority of informants gave the impression that clinician engagement in model development was problematic, as modellers discussed clinicians who distrusted or who did not understand modelling. Informants talked about having to manage and ‘filter’ information given by clinicians into states or pathways that were suitable or important in informing model structure. Some informants suggested that offering clinicians a background or training in modelling would improve their knowledge of the information needed in a model, and may encourage them to trust and invest further in the process. Informants stated that formal face-to-face meetings were optimum in facilitating communication with clinicians, whilst visual aids and ‘hands-on’ involvement may help to enhance clinician understanding. However, some informants suggested that experts might not welcome or have time for training and further engagement in model development.

5.5 Model reflection

All informants offered reflection on what they perceived to be good and poor practice within the modelling process, and what they considered as a good and poor modelling outcome. Informants also offered opinions on the feasibility of developing an ideal model. These issues are reported on next. Perhaps not surprisingly, informants tended to refer to their own practices when discussing good processes, and the practices of others when referring to poor processes or outcomes.

5.5.1 Good and poor modelling practice

Informants offered similar opinions of what constituted good and poor modelling practice within the modelling process. Good and poor process elements were reflective of one another i.e. poor practice was apparent where good practice was lacking. Informants discussed the involvement of experts and also rigour in model development.

5.5.1.1 Expertise

Two thirds of informants cited expert involvement in model development as good practice. For half these informants, this concerned the regular involvement of clinicians. The general impression from many was that having this support facilitated model building and led to a more robust model outcome. Most informants implied however, that regular and engaged clinician involvement was an ideal rather than a typical aspect of their modelling processes.

I23: “we had ‘buy in’ from clinicians ...they were really interested in what was going into the model and what was coming out...”

I19: “... that was my best modelling practice because I had a lot of interaction with experts in the field and everybody put in a lot of effort, it was very rigorous. I still love that model and results that we generated, I believe in them more than other results”

Around a quarter of informants cited other expertise, such as the involvement of additional health economists, as beneficial to the process. Again, many informants gave the impression that their involvement facilitated the process and improved the quality of the model.

I22: “the most successful process included the health economists from the get go... making sure you have the right type of data...and then throughout the process being involved...”

In terms of poor modelling practice, over a third of informants cited a lack of expert involvement, with most informants implying that limited expertise affected model validity. One informant, I20, described his own experience of having difficulty developing a model without assistance from others.

I22: *“a process where there are no other experts on the team who are thoroughly going through and validating the results...that’s when I start to have concerns”*

I20: *“...I was using the wrong program, I didn’t know what I was doing...that was my first model and I was operating without a net, I didn’t have a modeller working with me...”*

5.5.1.2 Rigour

A few informants referred to a good modelling process as one that was robust, defined as such because it included multiple model checking activities.

I12: *“good review processes and good quality control....help eliminate as many errors as possible...”*

I23: *“making sure that it’s validated against clinical opinion...all assumptions are valid”*

In contrast, a poor process was identified as one using unsystematic methods to identify data, and too few model checking activities. Over a third of informants commented on problems associated with models being populated with unrepresentative data that could bias results.

I15: *“why is it so common to find just one data source per parameter? It opens the model to selection bias...”*

I9: *“it’s about cherry picking the values, I’ve seen that happen and have had quite robust arguments with people about why they shouldn’t do that...”*

I19 gave the impression that it was easy for modellers to bias model outcomes through their choice of data inputs, and a perceived lack of scrutiny from others.

I19: *“...I feel like I can get any results from my model, I have so many degrees of freedom, so many parameters to vary, I can vary them still in a plausible range, no one will criticise me to ‘why did you pick that value for instance, that value?’ but internally I know this value is going to give me the results that are publishable...the aspect of modelling that has put me off is the amount of subjectivity that goes in...”*

Just under a quarter of informants suggested that a lack of model checking was definitive of a poor process, and often a poor model.

I5: “...they don't do any good consistency or validity checks or sensitivity analysis at the end...”

I6: “...you look at the published results and they just don't make any sense, validation is a big issue...”

5.5.2 Good and poor modelling outcomes

Almost all informants commented on good and poor model outcomes. The outcomes concerned whether finished models had been built with integrity, had representative structures, and could provide an answer to a specified research question. These issues are explored next.

5.5.2.1 Integrity

Just under half of informants implied that a good model would result from a model structure that had been developed with integrity. For most this involved following an objective and transparent process, to avoid bias, but to inform the audience of biases that were unavoidable.

I19: “... a structure that represents a quantified knowledge in the evidence base, has been impartial, objective and transparent”

I17: “we'd always want to be utterly transparent about what goes into the model so anybody seeing it for the first time can understand exactly what's in there and where the biases might be”

A few non-academic informants emphasised the importance of maintaining integrity and avoiding bias when developing models on behalf of pharmaceutical companies, due to client preferences towards a cost-effective model outcome.

I16: “...sometimes global teams want to have an optimistic model but UK people like me say 'well no that's, NICE will see through that, I'm not going to put my name to that', you have to make the inputs credible, non-biased, defensible...”

A third of informants implied that a lack of transparency in model reporting may lead others to question the quality of a model. The attitude was that modellers do not write-up their

modelling processes in enough detail, leaving users unsure of whether they can trust model results.

I15: *“there’s no information in terms of how did he come up with the other parameters of the model...I find that quite disturbing...And then you have no comments about ‘has the model been internally tested for validity?’”*

I8: *“their reporting, I don’t know if it’s reporting or the construction is so poor that we can’t use them, and that’s a shame...”*

However, a number of these informants suggested that the word limits of journals may account for a lack of reporting, rather than poor practice.

I18: *“my default position is it’s not the author’s doing and it’s on the journal’s side because you are so restricted”*

5.5.2.2 Representative structure

Around a quarter of informants considered a good modelling outcome as the development of a structure that replicated the reality of a disease, thus generating meaningful clinical and economic outcomes.

I21: *“...it should fit the real world according to my knowledge...”*

I23: *“a good outcome would be one that we are confident that it reflects practice”*

Over a third of informants suggested that a model structure and its results must be checked by relevant experts, particularly clinicians, before an outcome can be assumed valid.

I18: *“... I require different people to be happy with the structure, the people that build it but also that validation from technical colleagues...to trust what comes out...”*

I2: *“a model that the clinicians are actually happy with, because it doesn't really matter if the modeller is happy with it, if it's not like real life you're stuffed”*

In terms of a poor outcome, a quarter of informants referred to this as the use of an inappropriate structure. This was defined as either the wrong type of model or the development of a structure that did not reflect clinical reality. A number of these informants

implied that the use of an inappropriate structure related to a modeller's lack of understanding of a disease and associated treatment.

I1: *"they had done a simple decision tree, which doesn't capture the effects of an infectious disease properly"*

I14: *"it didn't reflect what patients went through, people were put on something that cost £6000 a month and in the UK that doesn't happen, it was the wrong comparator..."*

I24 was the only informant to suggest that an inappropriate structure may be the result of modellers' developing structures based on available evidence, rather than clinical reality.

I24: *"people don't include a state because there's no evidence for it...I feel like then the model has been built based on the evidence rather than the clinical pathway, I think it's a major limitation, and quite common"*

5.5.2.3 Answer

Almost half of informants argued that a good model provided an answer to a specified research question. Whilst a few academic informants suggested that the direction of the answer was not important, the majority indicated that they would be looking for a 'cost-effective' outcome. Interestingly, this mostly related to the non-academic consultancy informants, who on behalf of their clients were aiming to demonstrate an intervention to be cost-effective.

I12: *"for the client is an ICER which is less than whatever the threshold"*

I13: *"I know how my clients would define a good modelling outcome, 'drug approved!'"*

A number of informants additionally indicated that the outcome of a model should prove useful in practice, and further, that model users should be able to make decisions from it.

I16: *"...the people making the decisions feel that the model was sufficient to guide them with the decision"*

I14: *"it informs a decision, if it doesn't inform a decision then what's the point?"*

A few informants suggested that model structures and outcomes that were not generalisable were problematic. I23 implied that generalisability was an issue given that model structures typically reflected the practices undertaken by particular experts in a particular location.

I23: “...*limitation in terms of making models that are reasonably applicable in different contexts...if you’re reading a paper and you see a model structure and you’re like ‘oh that’s not what we do here so there’s no point in continuing to read...’*”

5.5.3 Ideal model

Informants were asked about whether they thought it possible to develop an ‘ideal model’. A few informants suggested that there were ideal aspects of a modelling process, such as regular clinician involvement, validation and model checking activities.

I7: “*check it with clinicians as you go along and then obviously it’s validated by someone else who understands the clinical area and understands health economics*”

However, the majority of informants suggested that either there was not the time or the resources for an ideal model development process, or that it would never be possible to develop an ideal model.

I22: “*...re-program the model that’s an ideal for me... I have to admit that’s not always done, that’s just a constraint in resources...*”

I19: “*of course that’s wishful thinking, these people are busy, clinicians are busy, and they are not going to spend [more] time with you...*”

The informants who stated that an ideal model was not possible implied that a structure could never perfectly reflect clinical reality, and that model development was too subjective to generate one ideal model for any decision problem.

I11: “*models are inherently such a simplification that you couldn’t get an ideal*”

I20: “*short of recreating the universe, there will never be a perfect model...*”

I15: “*I feel it might be a lot of an art, some science frameworks behind sections of it but it still feels so much like an art that I’m unable to see what an ideal model would be*”

Two informants gave the impression that it was not efficient to aim for an ideal, implying that extra time and resources spent on model building may not deliver increased accuracy in a model outcome.

I14: *“for the amount of time we spend getting the most we possibly could, to do anymore would take a lot more”*

I16: *“it’s striking that balance between striving for perfection, as close of an approximation to reality as possible, within practical constraints of going to the nth degree and beyond a certain point, you put more in but it’s false precision, you’re conning yourself that you’re getting better results”*

5.5.4 Summary: model reflection

A good modelling outcome was identified as a structure that was clinically representative, had been developed with the regular involvement of clinical experts, and followed a robust process. Only the non-academic consultancy informants gave a preference for a particular answer arising from the model, implying that their clients favoured a cost-effective outcome. Informants felt that modellers could improve their processes generally by using more thorough methods for identifying data inputs, and undertaking further model checking activities. Other improvements concerned taking measures to ensure that model structures and outcomes were generalisable. One informant was of the opinion that subjectivity in model development was problematic, implying that modellers were potentially able to influence model results. However, informants also emphasised the importance of modeller integrity in model development. Despite informants being generally of the attitude that modelling practices needed to improve, a few questioned the practicality and efficiency of committing further resources to model development.

5.6 Modelling guidance

All informants discussed their knowledge of, use of, and attitudes towards published modelling guidance. This was related to the aims of the research, which were to establish how

modellers used existing guidance, and highlight where further guidance might be needed. The research also sought to gauge what guidance the modellers regarded as most useful. These issues around modelling guidance are all reported on next.

5.6.1 Use of guidance

Table 6 provides a summary of the type of guidance used or referred to by the informants.

The most commonly cited guidance was the ‘ISPOR-SMDM’ modelling guidelines, followed by checklist guidance and modelling textbooks.

Table 6: Summary of type of guidance used or referred to by informants

Type of guidance	Number of informants who had used or referred to guidance
ISPOR-SMDM	12
Checklists	9
Textbooks	6
NICE methods guide	6
‘In-house’ guidance	4
CADTH guidelines	3
Methodological papers	2
Taxonomies	2

Around half of informants suggested that they used guidance regularly, although in different contexts. Whilst a few of these informants, all junior academics, gave the impression that they used it to assist their model building, the majority implied that its purpose was to justify their model development. The latter informants suggested that they adhered to and cited the recommendations of well-known guidelines and checklists, to demonstrate that they had followed good practice. These informants were mostly senior, with the majority being non-

academics, adhering to the NICE methods guide whilst developing models for submission to NICE.

I24: *“If I’m going to try and publish this I’ll probably cite that I’ve used the ISPOR guidance, so I just want to make sure that I follow it”*

I17: *“the NICE reference case is obviously something that we’d have to adhere to for modelling so it’s really looking through the report and making sure that everything is matching what NICE wants”*

Around a third of informants stated that they never, or very rarely, use modelling guidance. These were almost all senior informants, with the non-academic ones suggesting that they used ‘in-house’ checklists, ostensibly in place of published guidance.

I1: *“I haven’t looked at guidance. I have my way of doing things and it’s worked for me”*

I11: *“we’ve got quite a comprehensive checklist that we use here...”*

5.6.2 Attitude towards guidance

Around a third of informants appeared to be of the opinion that the current modelling guidance was useful. Most of these informants referred specifically to the ISPOR guidelines, which were considered helpful because they guided the modeller through the entire modelling process, and reminded them of the specific stages and methods that they should undertake.

I9: *“it’s providing a framework to the process of the project...it’s like having a skeleton to hang everything onto to make sure you cover all the right areas”*

I12: *“...used the ISPOR taskforce publications before... [it tells you] make sure you’re including ‘x, y and z’ ...”*

However, a few informants suggested that they found methodological papers most useful, given that they provided specific and detailed guidance on aspects of model building.

I6: *“one thing that is good...the [NICE] DSU does technical report documents...a lot of them are focussed on things like evidence synthesis...not focussed on modelling as a whole but on aspects of it”*

The most common criticism made by almost a third of informants was that guidance was too fragmented. Many of these informants implied that it would be useful if there was more relevant and concise guidance in one place.

I23: *“there’s a lot of material out there, a bit of a more condensed version at a more introductory level would have been helpful”*

I3: *“I haven’t found many guidelines ...like something that you get one document, ‘oh here it is’, you know?”*

Around a quarter of informants were of the opinion that there was not enough guidance available. Informants most commonly referred to a lack of guidance on structure, clinician involvement and the modelling process.

I4: *“there isn’t that much to accompany you in the process of building the model”*

I8: *“...I haven’t seen anything to say... ‘how do you think about a structure of a model’, how do you think about how many states?’ ...”*

I13: *“The [clinical] interpretation step is the bit I think is more difficult and makes a model structure relevant or not, so that’s the really key bit and there really isn’t any guidance...”*

5.6.3 Summary: modelling guidance

The informants’ use of guidance appeared to vary according to their seniority and the context of their work. Junior academic informants were more likely to use published guidance in their everyday practice, suggesting that these were the modellers who found it most useful. Senior and non-academic informants tended to use guidance as justification for aspects of their modelling work, particularly in demonstrating adherence to good practice processes. This seemed particularly important for non-academic informants, who had to follow guidelines outlined by the funding bodies to which they were submitting models. All informants appeared to value process guidance, such as the ISPOR-SMDM guidelines, above any other type. A number of informants gave the impression that current guidance could be improved

with the development of further process guidance, with more detail on structural development, and how clinician opinion should be used.

5.7 Future research

Informants were told at the end of their interviews that the next stage of the research would be case studies with teams of modellers and clinicians, undertaken with the aim to further explore modelling processes and investigate aspects of model development that would benefit from further research and guidance. Informants were asked to identify areas of model development that they considered important to focus on in the case study research, such as stages that they found difficult or where they felt current practice could be improved. Suggestions for future research were entirely from the perspective of the informant, without prompt from the author. Informant suggestions are categorised in Table 7.

Most informants proposed that structural development required further research. These informants gave the impression that modellers did not focus enough on the conceptualisation of a structure, and ensuring that it included all important states or pathways. A number of informants commented on a lack of knowledge regarding how clinicians should be involved in structural development, specifically the process used to translate clinical opinion into structure.

I24: *“...coming up with those diagrams is the most important part and I don't think that we spend enough time on it...”*

I13: *“the current area that is lacking is knowledge about the translation of the clinical evidence into the model itself”*

Table 7: Informant suggestions for future research

Suggestions for future research	Number of informants
How to design and conceptualise a model structure (in conjunction with clinicians)	11
Model validation	6
Communication and engagement between the modeller and clinician	5
How data parameter inputs were selected	5
Evidence synthesis of parameter data	3
Structural uncertainty	3
How to make structural assumptions with clinicians	2
How to elicit expert opinion (including how many clinicians to involve)	1

Clinician involvement appeared central to informants’ suggestions for further study. As well as clinician involvement in making structural and data assumptions, informants considered it important to explore how the working relationship between modeller and clinician could be facilitated.

I2: “...the communication between the clinicians and the modellers. I think that’s quite important...”

One informant suggested that future research should focus on developing guidance to assist clinicians in model development.

I23: “the feedback that we got from clinicians was that they like the idea of models but they don’t really trust them...if there was a lay guide to models that would be really helpful”

Other suggestions concerned exploring good practice for identifying evidence to populate a model, and which model checking activities were most important to carry out.

I5: *“which evidence you incorporate into your model, be it for parameter inputs or structure...it needs more investigation as to what the best ways of doing that is...”*

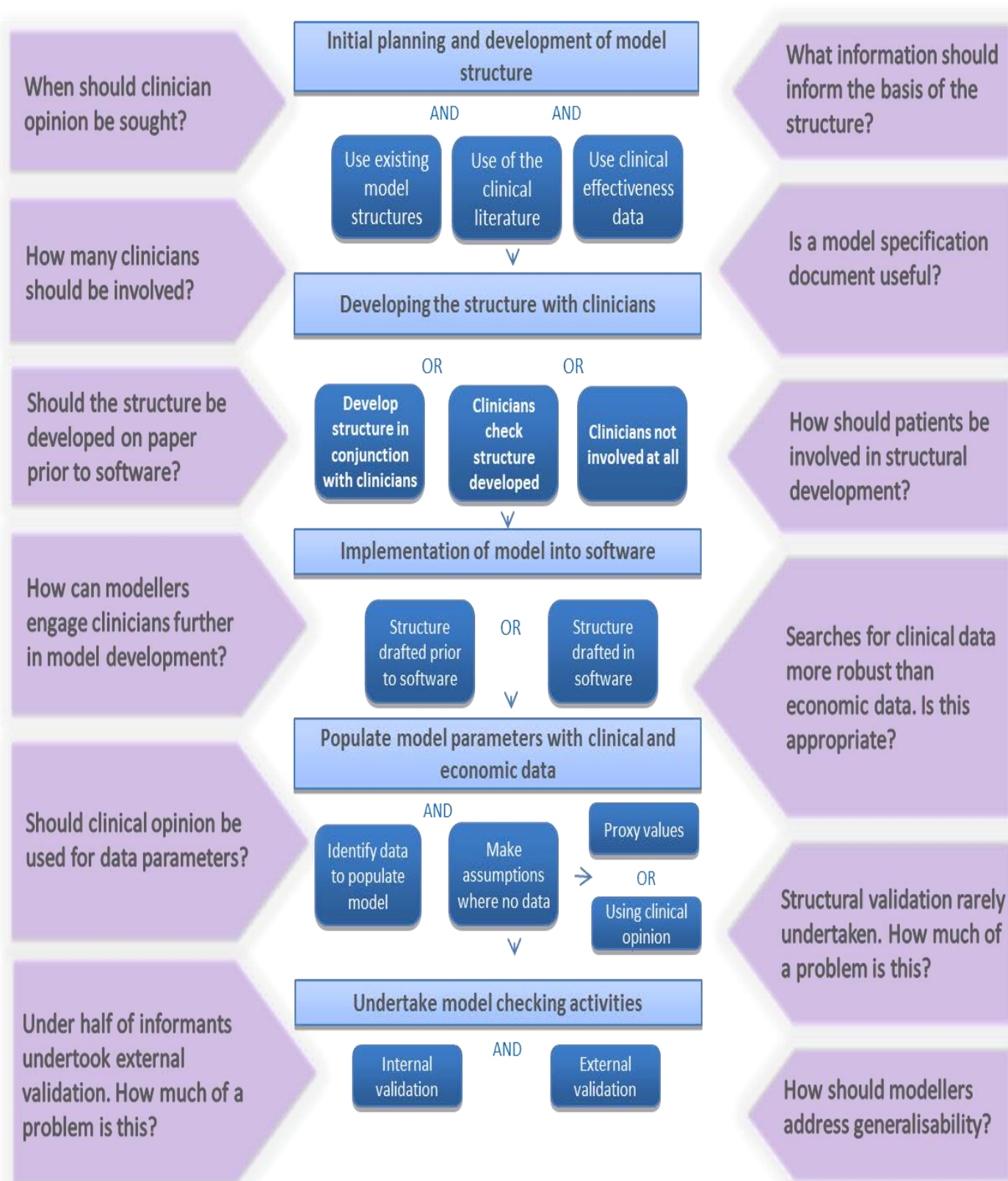
I22: *“...what is best practice for validating a model?”*

5.8 Conclusion

The model diagram outlined in Figure 4 demonstrates the typical process that informants followed when developing a decision-analytic model. Although the overall stages were similar, the diagram illustrates the differences in the methods used within these stages. Differences referred to the information that the modellers used to define a model structure, how clinician opinion was sought, the number of clinicians involved, the steps undertaken to plan and draft a structure, and how data were identified to populate the model. Some differences in practice appeared related to the context of the modellers' work, but others seemed indicative of contrasting perspectives over the methods that should be used. Similarities in informants' processes highlighted aspects that might generally require improvement, namely the use of systematic methods to identify economic data, and the undertaking of further model checking activities, particularly external validation and structural checking. When discussing the shortcomings of model development, informants stated that more robust methods of data collection and model checking were needed, albeit mostly in relation to other modellers' practices. Informants also raised issues around the generalisability of model structures.

The overarching issues related to informants' modelling practices were clinician involvement, and a lack of time and resources. Informants suggested that clinician involvement was essential to model development, but also discussed problems with clinicians' understanding of, and attitudes towards modelling, which impacted on the modellers' ability to conduct the process.

Figure 4: Diagram of the typical model-building process with questions raised



However, informants also proposed that holding face-to-face meetings with clinicians, using model diagrams during conversations, and offering clinicians training in modelling may encourage them to engage further. The second issue was the lack of time and resources available to build models. Informants tended to cite resource constraints as a reason for undertaking less robust practices. This finding suggested that further resources may be required in modelling processes, but informants also argued that it was impractical and inefficient to dedicate additional time and expense to model development.

There were a number of implications from these findings for the second phase of the research. Informants made direct suggestions as to where they felt further research should be focused, including exploring how structural development should be undertaken, and communication between the modeller and clinician managed. Differences in informants' practices raised questions around whether some informants were using more appropriate methods than others, suggesting that further consideration is required in regard to what constitutes good practice in these areas.

The next chapter presents two complete model-building case studies, which followed the modelling processes of two different teams of modellers and clinicians. The chapter provides step-by-step accounts of the stages and methods undertaken to develop two separate models within different model settings. In addition to the modelling practices carried out by the modelling teams, the case studies explore the overarching issues highlighted in the in-depth interviews, and investigate whether similar processes, methods and issues were encountered.

CHAPTER 6: RESULTS OF MODEL- BUILDING CASE STUDIES WITH MODELLERS AND CLINICIANS

6.1 Introduction

The two model-building case studies presented in this chapter aimed to explore and compare the modelling processes used by two separate teams of modellers and clinicians. This phase of the research was designed to build on the findings of the qualitative interviews with modellers presented in the previous chapter, whilst gaining additional insight into how clinicians were involved in model development. Two modelling case studies were selected on the basis that both modelling teams had good reputations for model building (as recommended by the qualitative interview informants), and could potentially provide different perspectives on the modelling process, due to the different contexts in which the models were developed. Whilst Case Study A was concerned with the development of a model in an academic context, Case Study B followed the development of a model designed to inform healthcare policy. Both aimed to compare the cost-effectiveness of alternative diagnostic pathways, within Medical conditions 1 and 2 respectively. The findings for Case Study A are presented first and cover the background to the model, development of model structure, data and populating the model, results, model checking and clinician involvement. The findings of Case Study B are then presented under similar headings.

6.2 Case Study A – Diagnostic model in Condition 1

The next sections present the findings of the non-participant observation, semi-structured interviews and think-aloud interviews undertaken within Case Study A. The research concentrated on those immediately involved in model development, but those who contributed to the model more broadly are also referred to. The background to the model is explored first, including who was involved in the process and the model objectives.

6.2.1 Background

Case Study A followed the development of a model within a UK academic context. Those immediately involved in model building were two modellers (I1 and I25) and one clinician (C1). An extra member of the team, Modeller 2, was identified as an “expert modeller” and held an ad-hoc role in the process.

The modelling element of this study was part of a larger project focused on the conduct of an RCT. C1 was PI of the study and a consultant in the clinical condition being modelled. A wider team of clinical collaborators and statisticians were involved in the RCT and model development, providing clinical input and statistical expertise respectively. The role of the clinical collaborators will be considered in more detail later in the chapter.

The objective of the model was to compare the cost-effectiveness of three diagnostic strategies for identifying problems related to Condition 1 (current practice, Strategy D and Strategy E). The decision problem was modelled in the context of doubt among the clinical community as to the usefulness of the current practice test, which had previously been considered to be the gold standard. C1 believed that Strategy E was the optimum strategy instead.

“...why bother doing a test if you're going to treat patient symptoms, why put them through [an] invasive test and the...result [is] we are still going to treat that [patient reported symptom]” (C1, clinical meeting 1)

6.2.2 Structural development

Most of the time that the modellers and clinician spent together was focused on designing and discussing the model structure. The next sections report on the communication between the two parties, focusing on how the pathways of the model were developed.

6.2.2.1 Beginning structural development

Prior to any formal structural development meetings, I25 reported visiting C1 and the clinical collaborators to gain an understanding of how the current practice option worked in its diagnosis of Condition 1. I25 also undertook general research into the condition, and specific reading into the patient pathways represented in the NICE guidelines. On the basis of this research, I25 developed an initial diagram of the current practice treatment pathways, for discussion in the first structural meeting with C1.

“Initially I met with C1 and other clinicians to make sure I had an idea about what they are doing and after that I did some readings to see what...NICE suggests for the treatment of the specific disease...” (I25, interview)

6.2.2.2 Pathway plotting

Most meetings between the modellers and C1 focused on translating patient treatment pathways into model structure. Communication involved a mixture of open discussion from C1 regarding how patients were treated, and clarification and follow-up questions from the modellers. From C1 the modellers learnt the order of treatments that a particular diagnosis resulted in, and what led patients to receive different treatments to one another.

“I wanted to understand what follows what and...what drives the decisions made in the clinical pathway...” (I25, interview)

In the first clinical meeting, I1 requested that C1 discussed freely what happened to each group of patients after they had been diagnosed using the current practice test.

“It’s important to find out from you what the patient pathway would be, as in the patient would have this and then this treatment...” (I1, clinical meeting 1)

The modellers asked C1 questions to establish why patients were given particular treatments at various points, and learnt that factors such as the result of a test, clinical diagnosis and patient choice would determine the decision, dividing patients in terms of pathways followed.

C1: “...it’s patient choice” **I1:** “...patient choice...it’s not the best treatment for what you’ve got is [Minor surgery 1], it’s you can have [Minor surgery 1] or you can try this...okay”

(Clinical meeting 1)

The modellers were next interested in learning about possible outcomes of treatments, seemingly to identify subsequent pathways, and the patients who should be following them.

I25: “... is having [Symptom 1] but cured of the [Symptom 2] after [Minor surgery 1] considered a successful treatment?” **C1:** “yes, yes”

(Clinical meeting 2)

Most discussion and questions asked by the modellers concerned establishing the chronological order of treatments, particularly whether treatments were sequential or not.

These questions were related to the apparent need to simplify pathways.

Modeller 2: “...the question is whether we actually need to model those alternatives or whether we can...” **I1:** “...put one in and do a weighted average of it, I think it’s the latter from what C1’s saying, it sounds like you’re sending them for minor surgery, it could be [Minor surgery 1] or the other one... the fact is they are substitutes for each other...you don’t do [Minor surgery 1] and then the other”

(Clinical meeting 1)

Also as a means to simplify the pathways included, the modellers emphasised the need to hear about the pathway of an average patient.

“Is there a general case that we could have, assume that everyone will go forward after two treatments unless they’re a special case and go out of our model?”

(I1, clinical meeting 1)

The meetings between the modellers and C1 were focused on determining the pathways for current practice, whilst the pathways for the two comparator arms (Strategy D and Strategy E) were discussed between the modellers only. The modellers developed the comparator

pathways to be analogous to the patients' true conditions, as defined by the current practice test. This was to account for the consequences of patients receiving a misdiagnosis.

I1: *“The test found them positive for [Problem 1] even though we know that's going to be a false positive...and the test found them negative to [Problem 2] so we know that's a false positive, is that how it should be?”* **Modeller 2:** *“It's what determines the effectiveness of the treatments...”*

(Modelling meeting 1)

6.2.2.3 Boundaries of model structure

Decisions regarding the boundaries of model structure were made in conjunction with C1, and concerned the population of the model, the patients excluded, model type, model outcomes and time horizon. Although there was some discussion between the modellers and C1 regarding these elements, it was typically the modellers who made the final decisions on what was included in the structure.

The model population was determined by the RCT study, but also limited by the modellers. This was in terms of the modellers opting to exclude 'smaller' groups of patients from the model, seemingly to avoid having to include additional patient pathways to represent exceptional patients. However, the exact proportion of patients who were considered exceptional appeared to vary. In one example the modellers discussed the importance of keeping 10% of the patient population in the model, but then later excluded up to 25%, despite C1's request to keep them in. Although C1 appeared reluctant to lose this particular group of patients from the main model, I1 made the decision that they should only be accounted for in the sensitivity analysis, which they later were not.

C1: *“...no they are more than an exception to the rule...in the prevalence studies they are quoting that 12 to 25% of the population have [Problem 3]...”* **I1:** *“so my suggestion to you is that our main model excludes this specialist patient group but in the sensitivity analysis we allow the 25% because there are data in the literature....”*

(Clinical meeting 1)

However, there were also examples of C1 wanting to include very small groups of patients in the model, seemingly not recognising the requirement for the pathways of the average patient.

C1: “...I would include them because in my daily clinical practice in the last six years as a consultant I have only seen maybe one a year that they’ve had [Problem 3] and they’ve had [Major surgery 1], so they’re a quite small proportion...” **I1:** “Right, so we just want to be trundling along as if most people are the general case, so actually although you’ve said include them I think we probably mean exclude them because they’re such a minority amount”

(Clinical meeting 1)

The two outcomes of the model, clinical effectiveness in natural units and quality-adjusted life years (QALYs), were selected by the modellers and influenced by the outcomes used in similar models. The time horizon was determined through discussion between the modellers and C1, although again the modellers made the final decision regarding this element. The conversation quoted below demonstrates I1 encouraging C1 to agree to a shorter time horizon than she initially appeared to consider appropriate, due to a perceived lack of data to model a lengthier timeframe.

I1: “...shall we just assume arbitrarily five years...would that be a good time horizon?” **C1:** “five years would be in the medium term, which would be the minimum they would need” **I1:** “so is five years long enough? I mean after that we are not going to get much data” **C1:** “no we’re not”

(Clinical meeting 1)

The selection of a decision tree structure was justified on the temporary nature of the condition.

“Given the short-term nature of the decision problem, the appropriate model structure...is a decision tree” **(Final model report)**

6.2.2.4 Basis of model structure

In addition to C1’s input, other sources such as the NICE guidelines influenced the development of model pathways. However, there was inconsistency in opinion and practice as to when and how the guidelines should be used in structural development. I25 gave the

impression that the structure should fundamentally represent clinical practice as described by the clinician, and that a secondary analysis should be undertaken to assess the effect of any variation in practice suggested in the guidelines.

“[the] pathway would be what C1 said is happening, [that] clinical practice involves several current practice tests but we’d do a secondary analysis with what NICE says, which is only one test in the beginning” (I25, interview)

However, there were examples where the NICE guidelines were used in the main model in place of C1’s experience, as a means of simplifying the structure.

C1: *“...if that’s becoming too complicated you can just say what NICE is saying...we should be giving everybody [Measure 1] of [Minor surgery 1], but we...are giving them a lesser dose...”* **I1:** *“...so there would be no repeat [Minor surgery 1] in that because you would have already done it. That would make it simpler and it’s the NICE recommendation”*

(Clinical meeting 1)

In another example, the guidelines were used to exclude the effects that rationing would have on the treatments that patients received.

C1: *“...because we are struggling with the NHS funds, we don’t necessarily send everybody to [Major surgery 1] because it costs £10,000 per patient...so a lot of our patients drop out...”* **Modeller 2:** *“...for modelling purposes I don’t think we should...”* **C1:** *“yeah, because in an ideal world we should be giving everybody [Major surgery 1]...”* **I1:** *“...so let’s follow NICE guidelines in the model and show the impact of it in terms of cost-effectiveness”*

(Clinical meeting 2)

I1 was very clear however, that clinical probabilities and treatment effectiveness data used to populate the model should not be used to inform pathways. She felt that this could bias the structure, as the modeller would then be aware of how the inclusion of particular pathways could impact on the cost-effectiveness result. I1 argued that a modeller should only have a general idea of the clinical data available to populate a model, to know whether model development was feasible.

“...you should know what type of data you’ve got when you’re making the model structure but you shouldn’t actually see it in terms of probabilities, see it in terms of it giving away information that will lead you to the answer...” (I1, interview)

I1 suggested that developing the structure on the basis of available evidence was not a credible method, with good practice being to draft a structure first and then populate it.

“...it’ll be right up the street of people doing it wrong to see the data and change the question... first you need to populate the model as is” (I1, modelling meeting 2)

6.2.2.5 Making structural assumptions

Making structural assumptions appeared iterative and in response to the need to simplify the model structure, usually due to lack of data. Assumptions were typically suggested by the modellers and confirmed by C1. A major assumption made was that patients would proceed from receiving the least to most invasive interventions, despite this not being reflective of clinical practice. I25 stated that no data were available to estimate the effectiveness of sequential treatments, namely patients receiving conservative treatment after invasive interventions.

“...If you continue to put conservative treatments in the pathway then what would be the effectiveness of the conservative treatment at that point? Do you have that information? You don’t have it...” (I25, interview)

The modellers and C1 also agreed that there would be no recurrence in the patients’ conditions. Again this was not reflective of real life but was seemingly related to a lack of long-term follow-up data, and the shorter time horizon agreed for the structure.

I1: *“so they remain symptomatic”* **C1:** *“and they go back through the loop again”*

I1: *“yeah [but] that’s the end of what we’re doing, we put them through it once...”*

(Clinical meeting 2)

Asking C1 to ‘sign-off’ on assumptions appeared to be an important part of the structural process. The assumptions were emailed to C1 after meetings for official agreement.

I25: “I sent the assumptions to C1 to finalise them and she got back to me and all of them are fine” **I1:** “good, did she come back on an email with that?” **I25:** “Yeah”
I1: “keep that email”

(Modelling meeting 3)

I1 suggested that assumptions must be agreed to and documented to avoid the clinician questioning them later in the process. I1 implied that clinicians often wanted to change assumptions after knowing the results of a model, and argued therefore that clinicians should not be aware of how assumptions might impact on model outcomes at any point prior to agreeing to them.

“...never tell her how the assumptions are going to change the answer, ever, because that’s not what it is about. So I have known people in the past to say ‘oh I’ve seen that answer, let’s have a re-look at that assumption, that was silly’ and you’re thinking ‘well no it was fine when you looked at it not knowing the answer’” (I1, interview)

6.2.2.6 Communication in structural development

Communication between the modellers and C1 in structural development involved either face-to-face meetings or email correspondence. Most pathway plotting occurred in face-to-face meetings with C1, whilst emails were used for ad-hoc queries about the structure, and to confirm assumptions made. Both parties reflected that this format was optimal due to C1’s busy schedule.

“this format works very well because me being a busy clinician I don't have time for face-to-face meetings all of the time, so most problems are sorted over email but there might be a couple of things that really we need to sit down face-to-face to explain and understand...” (C1, interview)

Both I1 and I25 stressed the importance of pathway plotting taking place in face-to-face meetings. I25 indicated that he struggled when asking and receiving answers to structural development questions via email, seemingly due to the complexity of information involved.

“I tried emails, asking questions and I understand that her time is valuable, [but] some of her responses were very brief, more brief than I was expecting, or wanted...I

had to think about....better ways to get the answers....that were relevant for the model. So that's why we had more meetings and more illustrative things... (I25, interview)

Communication focused on the modellers asking C1 questions about the pathways that patients followed. C1 often answered the modellers' questions in terms of the proportion of patients affected, and this appeared to be helpful to the modellers, as it led to discussion over whether groups of patients were large enough to be included in the model.

I25: “...after the [Major surgery 2] can they develop [Symptom 3]?” **C1:** “10%”
(Clinical meeting 2)

Both parties reflected on the use of the model diagram as a key communication tool, giving the impression that it provided a tangible resource for them to relate their discussions and understandings to.

“...just for the sake of the conversation...it is difficult when I'm discussing something and they don't have the idea, but more helpful when they had the picture” (I25, interview)

Problems with communication concerned C1's lack of understanding of the economic terms used by the modellers, which stalled communication between the two parties.

I25: “are the utilities, after the surgery...” **C1:** “...sorry?” **I25:** “...if the person has [Major surgery 1] for example, the utility might be different then, does the surgery...”
C1: “...the, what will be different sorry?” **I1:** “...the quality of life, so do we assume it's full health when they've had surgery...?”

(Clinical meeting 2)

6.2.2.7 Role of the clinical collaborators

The role of clinical collaborators in structural development appeared to be ad-hoc and typically at the request of C1. C1 gave the impression that she felt more comfortable gaining consensus from the collaborators on some issues, particularly where she seemed uncertain or suspected that practice might be different in other clinical centres.

I25: “so what would we do at this point, [Major surgery 2]?” **C1:** “I'll discuss it [with the clinical collaborators], share the burden”

(Clinical meeting 1)

Although C1 appeared keen to gain the input of the clinical collaborators during pathway plotting, I1 suggested it more appropriate to involve them when the structure had been established, and only in a model checking capacity.

“...we wouldn't want to get them to join in at this point because we're still working out what makes sense, so we'll get a model from beginning to end...and then present it”
(I1, clinical meeting 1)

6.2.2.8 Location of treatment pathway

Relating to the generalisability of the developed structure, and the apparent lack of clinical collaboration in structural development, C1 mentioned a number of times during pathway plotting that the practices that she was describing to the modellers may not be the practices that were undertaken in every medical centre.

“...they will have a [Test 1] to begin with, that's in our centre though it's not a standard practice everywhere” **(C1, clinical meeting 1)**

C1 also seemed to be concerned about which treatment pathways to describe to the modellers given that she had worked in two different centres, which had each undertaken different practices.

“...I am pondering over it because...in [Location 1] the consultant would treat the current practice treatment result rather than treating the patient and the symptoms ...but in the [Location 2] because we have facilities for [Test 1]... if we can find [Symptom 2] then we will do [Minor surgery 1]” **(C1, clinical meeting 2)**

The modellers addressed this issue by stating that they would use the most typical practice in the primary model and test exceptions in a sensitivity analysis. However, this raised the issue as to whether C1 would be aware of what the most common practice was across various medical centres. Additionally, alternative practices were ultimately not tested within a sensitivity analysis.

“If your centre is the exception we’ll go with the typical case...we can do sensitivity analysis with exceptions to see what impact it has.” (I1, clinical meeting 1)

6.2.2.9 Implementation

The implementation of the model structure into the software occurred after it had been fully developed and signed-off by C1. I25 suggested that it was optimum to have a ‘good’ diagram of the model structure on paper prior to translating it into the software, as changes could be made on paper with less effort.

“...for me it’s easier to draw it down just to make sure ‘yeah that’s what I want’ and then put it on TreeAge....it’s easier than to work on the decision tree diagram itself because you can just write and erase... (I25, interview)

6.2.2.10 Reflections on structure

All informants were asked for their reflections on the model structure. Opinions appeared generally positive towards the standard of the model, although C1 had some doubt over the assumptions used. Whilst I25 gave the impression that the structural assumptions had made little difference to the overall clinical representativeness of the model, C1 appeared less comfortable with having to simplify some of the clinical aspects. C1 expressed concern that patients with coexisting conditions had been excluded, and therefore the true costs associated with some of the treatments had not been captured.

“...we modelled clinical practice as accurately as we could, we didn’t have to sacrifice assumptions to sacrifice the quality of our model...it is a more simplified version of what happens in practice” (I25, interview)

“In the model we are mostly concentrating on [Symptom 2] and no [Symptom 2] but then there are some patients who have coexisting conditions like [Symptom 1] ... [that] are not being accounted for in the sense that these people will need extra [Medical device treatment] ...and things like some of the [Medical devices] might be more expensive than others, those kind of minutiae are not [there]” (C1, interview)

6.2.2.11 Summary of structural development

The model structure was developed through discussion between the modellers and C1, with the focus being on learning the treatment pathway of the average patient receiving the current practice test. The modellers were interested in knowing about the interventions that particular groups of patients received, and how health outcomes and other factors affected future treatment. Communication was facilitated through face-to-face meetings, diagrams of the model structure, and a mixture of open discussion by C1, and clarification questions from the modellers. In addition to C1's clinical experience, the modellers used the NICE guidelines to inform the structure, although there appeared to be uncertainty around how these two sources should be used in conjunction. However, I1 felt strongly that available data should not influence the structure.

The modellers undertook various approaches to simplifying pathways as structural development progressed, including the merging of non-sequential treatments in the structure, and the exclusion of exceptional groups of patients. However, decisions regarding model boundaries and simplifications of pathways were sometimes inconsistent, and were mostly undertaken by the modellers, involving limited negotiation with C1. C1 did not appear to fully comprehend the need to include the pathways of only the average patient, and later expressed feeling uncomfortable with some of the structural assumptions and simplifications made. This raised the issue of whether communication could have been managed more effectively between the two parties, including further explanation and discussion of structural assumptions with C1. Additional problems with communication referred to C1's lack of understanding of the economic terms being used during discussions. Finally, issues around generalisability of structure were highlighted, as C1 talked about her clinical practices being different to others, and appeared concerned that the information that she was giving to inform pathways may not be representative of the experiences of all clinicians.

6.2.3 Data and populating the model

The data used in the model originated from various sources including the RCT study, literature, relevant registries or databases, and clinician opinion. Presented next are the methods used to identify and check the clinical and economic data used to populate model parameters.

6.2.3.1 Clinical data

The clinical data used to populate the model comprised: 1) the prevalence data that represented the proportion of patients with each of the problems related to Condition 1; 2) the accuracy data that informed the probability of patients being correctly diagnosed with one of these problems; and 3) the effectiveness of treatments given a patient's true condition. Most of these data were informed by the RCT study, specifically the prevalence data and the accuracy of the diagnostic strategies. These data were therefore collected by the clinical trial team and given to I25 in their appropriate form by the statistician. The treatment effectiveness data were informed by clinical literature or expert opinion. The data parameters requiring clinical input were those that referred to patients receiving treatment after a misdiagnosis. I25 stated that despite a review of the literature, there were no robust data on the effects of patients receiving inappropriate treatments. I25 undertook detailed reviews for all of the clinical parameters of the model that were not informed by the RCT, ensuring that he was using the best data available, preferably values from systematic reviews or meta-analyses.

“...it wasn't a systematic review...but it was thorough. I found some studies and then I tried to see which were the most relevant and appropriate sources of information. Thankfully most of them were systematic reviews or meta-analyses of clinical trials, so we had good data” (I25, interview)

6.2.3.2 Economic data

The economic data used to populate the model comprised utility data, which referred to the quality of life of patients, and cost and resource use data. The model used two utility values

within each model pathway, related to the patients' quality of life before and after they had received treatment. The initial utility values originated from the RCT and differed depending on the specific problem that a patient had. The utility values after treatment were an outcome of the model pathways, and referred to whether or not patients were "subjectively" cured of their symptoms. These endpoint values were informed by the literature due to a lack of long-term follow-up data. I25 first undertook a systematic review of the literature, but after finding no suitable studies, searched the 'Cost-Effectiveness Analysis Registry' for appropriate utility weights. This search was not systematic, but used relevant search terms to identify eighteen studies that could potentially inform QALY values for patients cured of their symptoms. I25 used a weight from only one study because the others were deemed inappropriate.

"Utility weights from other studies were not considered either because they were focused on a population with [Problem 4] who were undergoing [different treatments], or because they were focused on a [different population]" (Final model report)

I25 reported being unable to find appropriate utility values for patients not cured of their symptoms. Therefore, the modellers made the assumption that patients maintained the initial QALY value taken at the beginning of the RCT study.

I25: "... [the population] in the beginning of the study have [Problem 1]" **I1:** "...of course they do" **I25:** "If they remain with [Problem 1] they will have the same quality of life" **I1:** "...but if they get better you take them to full health?" **I25:** "No 0.92 and this is from the literature" **(Modelling meeting 4)**

I25 gave the impression that identifying utility values had been difficult and reported being unable to find and include quality of life decrements, which referred to patient utility after receiving appropriate or inappropriate treatments.

"I couldn't find decrements for the interventions; the best thing I could do is assume that if they are still with [Problem 1] they have a quality of life of 0.87..." (I25, modelling meeting 4)

Identifying appropriate cost and resource use data to populate the model appeared to begin at the pathway plotting stage. On a number of occasions, I1 asked for more information or clarification on treatments, particularly on the resources used and staff time involved, to determine costs.

I1: *“Do they have to have an anaesthetic?”* **C1:** *“...we are doing it under local anaesthetic now increasingly, but some patients will still go to theatre because they want a general anaesthetic...”* **I1:** *“...like a minor...”* **C1:** *“...operation...”* **I1:** *“...in terms of costs anyway, once you’ve given somebody a local anaesthetic they’re a day case, it’s a minor operation?”* **C1:** *“Correct”* **(I1, clinical meeting 1)**

Cost and resource use data were identified using the HRG (Healthcare Resource Groups) database. I25 discussed searching for specific HRG cost codes informed by a review of the NICE guidelines, and also searching the database manually.

“...I did a literature search to see what NICE suggested, they had some HRG codes in their reports so most of them came from there, and then if I couldn’t find a cost I did some searching within the HRG database” **(I25, interview)**

The information gathered about the treatments appeared to assist I25 in determining which of the HRG categories were most suited to the data requirements of the model.

“...so within its HRG category they have all the surgeries for example, they all fall within [the same category] so I searched there to make sure ‘this one’ was the type I wanted...” **(I25, interview)**

However, I1 implied that when using HRG codes to identify costs, it was not necessary to know too much detail on an intervention, as the categories each represented a broad definition of a particular treatment.

I25: *“Do we need to ask what surgery is?”* **I1:** *“No, let's just call it surgery for now...it's surgery...in that area, in terms of a cost you're going to be looking at minor surgery or major surgery so you don't need to know what it is, you need to say 'is it defined as minor or major?' because when you look at the HRG costs it will be merged...”* **(Modelling meeting 1)**

6.2.3.3 Data and clinician involvement

C1 and the clinical collaborators were involved in identifying and providing data to populate the model, with C1 offering references for clinical data not available in the RCT, and a number of the treatment effectiveness values being informed by the opinions of C1 and the collaborators. I25 elicited the effectiveness values from the clinicians via email, asking for the mean, mode and confidence intervals for each treatment effect estimate. He gave the impression that expert opinion was not a preferred method of data collection.

“Because I couldn’t find anything in the literature I [sought clinical opinion] ...the reason is that it had to be in the structure of the model and we needed that value...it was a last resort really” (I25, interview)

In terms of the cost data, C1’s role appeared to be more in a model checking capacity. I25 reported presenting C1 with relevant HRG cost categories and asking her to select the most appropriate. He implied that this method worked well.

“...because I had the description and comments about the HRG categories, that helped her realise ‘okay yeah it’s not that’ or ‘this intervention takes more than 20 minutes’ for example, and so it was very easy for her to say ‘yeah, that’” (I25, interview)

However, C1 did not appear to be involved in identifying or checking utilities. Perhaps related to C1’s lack of involvement in this aspect, I1’s comment below appeared to be indicative of what she considered to be a clinician’s role in populating the model.

“when it becomes utilities I’d say ‘yes that is your (I25) responsibility’ because the clinicians won’t know where utility data come from, or cost data or resource data, but [treatment] effectiveness data, the fact that they are doing this study means that that effectiveness data is buzzing around in their heads...where there’s a gap it’s their job to plug it” (I1, modelling meeting 1)

6.2.3.4 Summary of data and populating the model

I25 reported robust methods for identifying data to populate clinical parameters, with data originating from either the RCT or detailed literature reviews. The only exception to this was

the treatment effectiveness values informed by expert opinion. The economic data were informed almost entirely by the literature, either using utility values from relevant studies or costs from the HRG database. The processes undertaken to retrieve these data did not seem to be as thorough as those used for the clinical parameters, as non-systematic searches were ultimately used to identify values, and assumptions around utility data were made without validation from C1. However, I25 reported difficulty with finding relevant and appropriate quality of life values. The modellers and C1 seemingly held different roles in data collection, with I1 implying that it was the clinician's responsibility to identify clinical data, but that the clinician's role was limited in terms of the data used to populate economic parameters.

6.2.4 Results

After the model was populated with the clinical and economic data it was 'run' in TreeAge software to generate results, which were then presented to C1 for her feedback. For context, the model found the current practice test to be the most cost-effective. Whilst I25 expressed that the results were in the direction expected, C1 appeared surprised and unhappy with the outcome.

“This is not what I was expecting, I was hoping my hypothesis was right that the current practice test doesn't add much to our management and it's just a time wasting step” (C1, clinical meeting 3)

I1 reported that after hearing the result, C1 wanted to use new data in the model, to see how it would affect the model result. However, both modellers were strongly of the view that alternative data could not be used, as it would mean changing structural assumptions around the model population that had previously been agreed on. I1 suggested that C1 was unhappy with this decision.

“...there were some new data [but] it wasn't appropriate for us to put in our model....and C1 just didn't understand why we couldn't update our model with this

new data, and what it was, the data she'd found...applied to a different population...and she had forgotten that" (I1, interview)

However, a secondary analysis was undertaken to further test C1's hypotheses and seemingly address her concerns.

"...so we did a secondary analysis to see whether the interventions were more cost-effective in subgroups...patients with [Problem 4], Strategy E is not cost-effective, but it is cost-effective when you give it to patients with [Problem 1]... C1 was happy with that" (I25, interview)

In addition to the result not being as C1 had expected, further discussions between her and I25 identified disagreement and apparent miscommunication over how the Strategy E comparator was defined.

C1: "See this is what I disagree with because...Strategy E is not what the clinicians are saying, this is what the patients are saying...I don't know whether I have been on the same page here?" I25: "I thought it was..." C1: "No...the patient tells us what they have and we just tick it" (Clinical meeting 3)

6.2.5 Model checking

Most checking activities were undertaken by I25 after the model results had been generated.

These checks focused almost entirely on sensitivity analysis. The sensitivity analyses undertaken were mostly on the data in the model, and included increasing the cost of current practice, and reducing the test's accuracy. The analyses appeared aimed at gauging whether the direction of results would change if data values were varied to be less favourable towards current practice. I1 stressed the importance of the sensitivity analyses being sensible and justified, based on alternative relevant sources of data or using other legitimate scenarios supported by C1.

"But that's only if patients can be asymptomatic and actually everyone thinks that they are unlikely to become asymptomatic. You can always find a scenario that makes something better off...you do a sensitivity analysis to assess the robustness of the assumptions you've made, not change the assumption to something less robust..." (I1, modelling meeting 5)

The sensitivity analysis also tested uncertain data parameters, particularly those informed by clinical opinion. Only one of the sensitivity analyses involved a change to the structure of the model, in terms of seeing the effect that having only one diagnostic test in each treatment pathway had on the model result. This was to take into account the practice outlined in the NICE guidelines. The think-aloud interview demonstrated how I25 worked through each of the sensitivity analyses, stating what he expected to happen to the cost-effectiveness results of the model prior to running each scenario.

“By lowering current practice sensitivity...in [Outcome 1] ...I’m still expecting this effectiveness to be higher than Strategy E but the cost-utility might come down to .50, .40, it will still be dominating but at a higher cost...Let’s see, so current practice test, lower cost, less effective...Makes sense doesn’t it? Yep. I mean it was dominating but now it is less dominating...” (I25, think-aloud)

Whilst some results of the analyses appeared intuitive to I25, he struggled with the interpretation of others. For example, in realising that where all patients were assumed to receive the same intervention for a particular problem, the effectiveness of Strategy E increased because misdiagnosed patients now received more appropriate treatment.

“Why did current practice become less effective? [Long pause] There was a high likelihood of a wrong diagnosis for Strategy E, but if you assume now that these patients will be getting [Minor surgery 1] instead of having the option, patients receive the intervention they would get if they were correctly diagnosed. So you would expect the effectiveness to increase...” (I25, think-aloud)

I25 did not only use the sensitivity analyses to test the effect of alternative data and scenarios on the results, but also as a means of internally validating the model and checking that it was programmed correctly and working as it should.

“Run something just to see if the model works... that’s a check, whether it happened as expected...” (I25, think-aloud)

6.2.5.1 External validation

Despite considerable internal validation of the model, I25 discussed not undertaking any external validation due to a lack of similar models to compare the structure and results to.

“...there was a decision tree in similar symptoms but in a different population group; [there] wasn't something similar so I didn't do any external validation” (I25, interview)

6.2.5.2 Clinician involvement in model checking

The model checking undertaken by C1 took place throughout the development process, in terms of her informing and checking iterations of the model structure, confirming data values and giving feedback on results. However, C1's opinions appeared to become less influential as model development progressed, with I1 implying that the results were only presented to C1 to see if there were any further sensitivity analyses that could be undertaken.

“The only reason for doing it...they might say something that we ought to test...so it's worth them seeing the results...” (I1, modelling meeting 4)

6.2.5.3 Summary of results and model checking

The presentation of the results demonstrated a breakdown in communication between the modellers and C1, in terms of the definition of Strategy E and C1's lack of understanding and acceptance of the model conclusions. C1 appeared to have somewhat of a vested interest in the model outcome, in not wanting the current practice test to be cost-effective, and also appeared unaccepting of the fact that structural assumptions could not be changed after they had been agreed. Model checking focused on sensitivity analysis, with the checks aimed at exploring whether the results were robust to less optimal scenarios. However, the sensitivity analyses undertaken were almost all on the data in the model, despite discussions with C1 over the testing of alternative structures and the inclusion of additional patients. In addition to undertaking limited structural checking, I25 stated that he did not carry out any external validation due to a lack of similar models and studies in the literature.

6.2.6 Clinician involvement in model development

Clinician involvement was determined to be a key factor in the model's development, particularly in pathway plotting. However, problems were encountered and reported with their participation in the modelling process. The next sections explore issues around C1's and the clinical collaborators' involvement, specifically the number of clinicians involved, clinician understanding of modelling, and the potential for training clinicians in model building.

6.2.6.1 Number of clinicians

The involvement of only one clinician in immediate model development raised issues around the generalisability of the structure. I1 suggested that involving more clinicians would have allowed the modellers to learn about alternative clinical practices that could have potentially informed the model.

“The more people you involve, the more heterogeneity you get and that's important because you've got to know that there are other ways of skinning a cat or whatever”
(I1, interview)

This said, I1 defended the practice of working with only one clinician, arguing that there were other factors that informed the development and validation of the structure, including the NICE guidelines and the treatment pathways followed in the RCT.

“...it's not based on just one person's idea ever. There's historical models to look at...there's evidence from the literature and there's a trial that has already been peer reviewed” **(I1, interview)**

I1 also referred to the involvement of the clinical collaborators as being important to the validity of the structure, and indeed their opinion was sought at various points of model development. However, collaborator input appeared inconsistent, as I25 reported only receiving contributions from eight of the twenty-two collaborators. Further, C1 suggested that she struggled to get comments from the collaborators on structural issues, which often led to her reluctantly making decisions alone.

“...everybody's so busy it's difficult to get colleagues to respond and sometimes because of the time pressures I have to make executive decisions, and I wish this was not just left to me...” (C1, interview)

6.2.6.2 Clinician understanding of modelling

On a number of occasions, C1 appeared to have difficulty understanding and accepting aspects of model development. For example, she seemed unhappy with having to exclude patients from the model, having to make structural assumptions, and struggled with the interpretation of the model results. Despite C1 sometimes appearing confident in her understanding of modelling, expressing the view that making assumptions was necessary where there were limited data, she gave the impression more frequently that she found it difficult to comprehend how the model was working.

“...sometimes there are uncertainties that can't be resolved with the existing evidence and then you just have to make a lot of assumptions. It's not a weakness of the model; it's the weakness of the available literature” (C1, interview)

“I don't know anything...I don't have any knowledge of health economics so if you guys (I1 and I25) think it's okay then that's fine, whatever you explain to me seems fine” (C1, clinical meeting 3)

6.2.6.3 Guidance for clinicians

Without prompt from the interviewer, C1 suggested that clinician guidance would be a useful addition to the modelling process. C1 reflected that it would have been beneficial to have had some information prior to model development, on health economic terms, how a model works, and what her role would entail.

“it's quite easy to presume that the clinician knows about health economic modelling when I'm just ignorant of even the basic terminology and so I don't know whether...we could have a one or two hour brief in the beginning to explain what health economics is, what modelling is, and what they expect from us...” (C1, interview)

The modellers appeared to have different attitudes towards the suggestion that guidance should be given to clinicians. Whilst I1 gave the impression that she thought it unnecessary for clinicians to understand the model, at least not until they needed to use the results, I25

argued that the clinicians should know what information is important to communicate and include in the structure, particularly what will impact on the cost-effectiveness result.

“The difference is we had to understand their bit to do our bit, they can do their bit and assume that we know what we are doing for our bit, where it becomes an issue...we come up and say it’s cost-effective, it should be done in practice and then they think ‘God we need to understand how you can say this’...” (I1, interview)

“ I think it is necessary for them to understand...the health economics pathway and mostly what drives the costs and what is important in the clinical effectiveness of the treatment...the adverse events, things that drive the costs or the outcomes” (I25, interview)

The modellers’ opinions on the necessity of clinician training appeared to be related to their perspectives on whose role it was to train the clinicians. Whilst I25 stated that it was the modellers’ responsibility to guide clinicians through model development, I1 suggested that the clinicians should be educating themselves, giving the impression that it would be too much of a commitment to provide guidance to every clinical expert.

“...it’s our responsibility; we should be guiding the discussion...” (I25, interview)

“If she was motivated to, she could come on our courses and learn about modelling... there are ways of them finding out without me having to teach them, you don’t want to meet every collaborator you’ve got and sit down and tell them why you’re needed on their study...” (I1, interview)

6.2.6.4 Summary of clinician involvement in model development

Issues around clinician involvement in the process concerned the lack of clinical contribution to structural development, and C1’s lack of understanding of various aspects of model building. C1 stated that she struggled to follow the model’s development, and indeed, later had difficulty understanding and accepting the model results. Again, problems around the generalisability of the structure were raised, as both I25 and C1 reported limited collaborator involvement in the process, and gave the impression that more immediate clinician involvement would have been beneficial. Further, C1 suggested that she would have found her role in the process easier if she had previously received some training in health

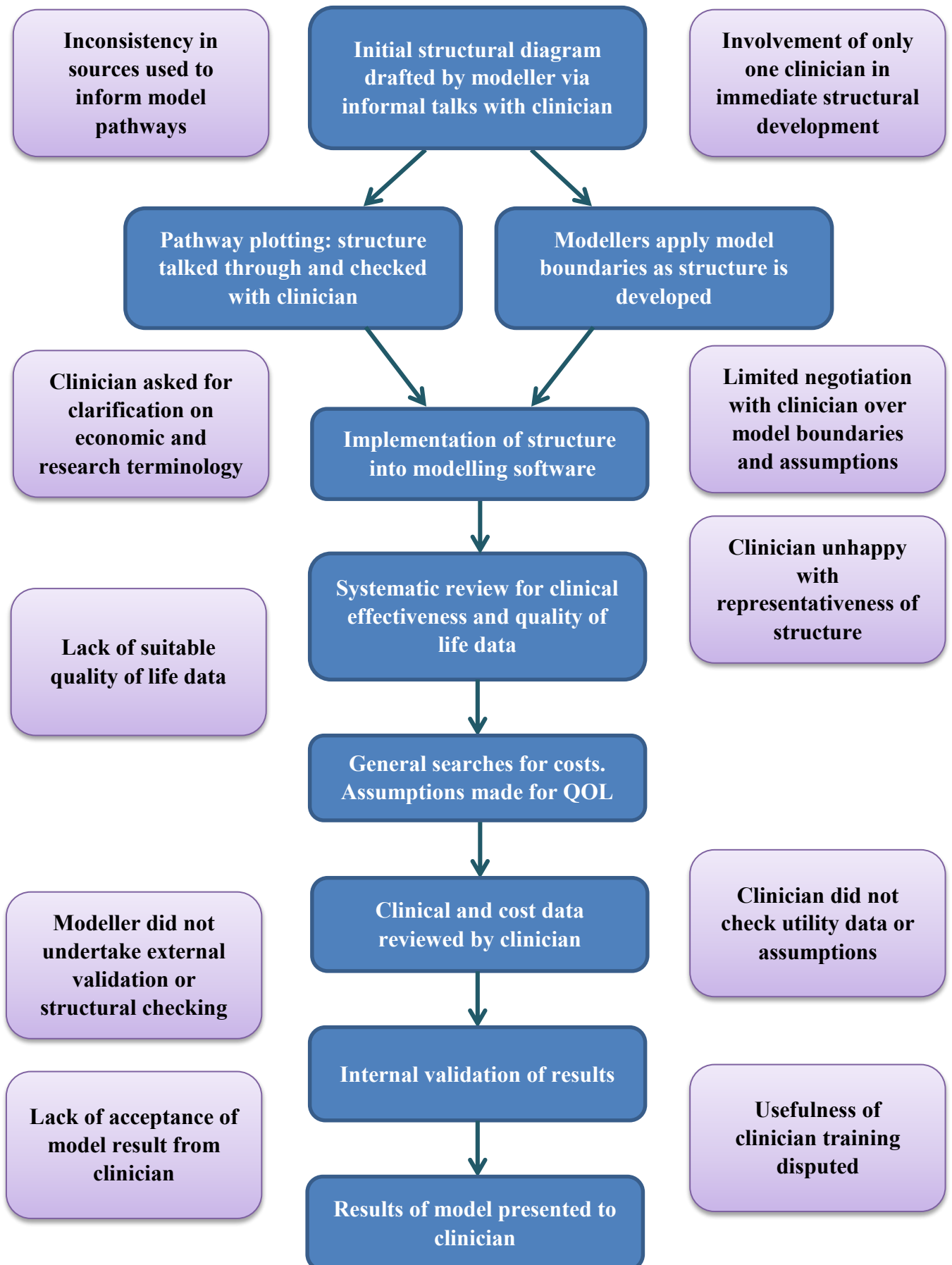
economics, to allow her to understand model development as it progressed. Although I25 agreed that it would have been useful for C1 to have known which information was most important to communicate when developing the structure, I1 argued that offering training to clinicians was impractical, unnecessary and outside of the modeller's role.

6.2.7 Conclusion

The first part of this chapter has presented the findings of the model development undertaken within Case Study A. The analysis of the observations and interviews has highlighted the stages and methods used, in addition to the issues and problems encountered. These are demonstrated in the diagram in Figure 5.

The issues that appeared most important were those related to structural development and clinician involvement in the process. Much of the structural development focused on discussion between the modellers and C1, with the pathways being drafted according to C1's experiences of clinical practice, in addition to the NICE guidelines and other clinical literature. The modellers and clinicians had open discussions around model structure, and commented that face-to-face meetings and reference to a diagram of the model facilitated structural development. C1 and the clinical collaborators were also involved in identifying some of the clinical data for the model, and C1 had the task of checking the cost inputs used.

Figure 5: Modelling process for Case Study A, with issues highlighted



Issues included inconsistency in the sources used to inform model pathways, questions around whether involving only one clinician in structural development was generalisable, and communication between the modellers and clinician. Communication was problematic in a number of respects, with the modellers not negotiating with C1 over model boundaries, and C1 not being accepting of the actions undertaken to simplify the structure. Further, there was some inconsistency in how model boundaries were being applied. C1 also struggled to follow model development, and found interpretation of some of the economic terms difficult. It became apparent that both C1 personally, and the development of the model would have benefited from greater communication from the modellers, potentially beginning with some form of training prior to the process.

Instances of the modellers using less robust practices were also identified. Economic data used to populate the model were retrieved through unsystematic searches, and structural assumptions and clinical opinion were used in the place of arguably more reliable sources. Further, I25 undertook limited structural analysis, and no external validation on the model and its results. However, some of these practices were not intentional, but instead based on a lack of suitable data and literature. For example, the modellers could not identify relevant utility data to populate the model, particularly long-term quality of life outcomes, making structural assumptions necessary. I25 also reported being unable to undertake external validation due to a lack of similar economic studies.

The modelling process conducted within Case Study B is explored next, under the headings of background, structural development, data and populating the model, model checking, results and clinician involvement in model development.

6.3 Case Study B – Diagnostic model in Condition 2

The next sections present the findings of the non-participant observation and semi-structured interviews undertaken within Case Study B. The background to the model is explored first and covers the objectives of the model, who was involved, and the training that clinicians received prior to model development.

6.3.1 Background

The focus of Case Study B was the development of a model in a UK healthcare policy context. Those involved in immediate model development were two modellers (I18 and I26), two project managers, three researchers and around twenty clinical experts. For the purpose of this research, the reflections and opinions of the two modellers and three of the clinicians (C2, C3 and C4) were concentrated on. The remaining members of the modelling team and clinicians were referred to in terms of their contributions to model development meetings.

The objective of the model was to compare the cost-effectiveness of five alternative diagnostic strategies for identifying Condition 2 (current practice and Diagnostic strategies 2, 3, 4, and 5). Each strategy used a different combination of Diagnostic tests A, B and C. The medical condition in which the model was developed was selected because it was considered a priority by the modelling team and clinicians. The clinicians reported variation in the way that Condition 2 was treated, and suggested that current practice was inefficient and had questionable accuracy. It was mooted that identifying the most cost-effective strategy for detecting Condition 2 would result in important cost savings for the health service.

“[Condition 2] was picked because it was quite a high prevalence so it could have a big impact [financially]...” (I18, interview)

6.3.1.1 Clinician training

Training was offered to all clinicians involved in model development. Clinicians were given the option to attend a training session prior to the modelling process, and were also able to access the material online. Training focused on giving the clinicians a background and context to modelling, and information on how models worked and generated cost-effectiveness recommendations. The training was delivered through presentations given by the modellers, and included tasks for the clinicians to complete. The background information covered why health economics and cost-effectiveness were important, and explained how models trade-off the costs and health benefits of interventions. The presentation seemed to be directed at addressing any apprehension that the clinicians had around using cost-effectiveness, and addressing the ‘myth’ that modelling was only concerned with costs.

“Myth 1: ‘Health economics is about saving the government money’

- *Health economics helps the NHS use its limited budget to maximise health outcomes for the whole population [and] identify interventions offering the best value for money”*

(Clinician training presentation)

The tasks were aimed at providing the clinicians with an understanding of how model structures were developed, as clinicians were asked to populate a structure, calculate cost and QALY output, and determine the resulting ICER. The clinicians also had to identify the potential impact of assumptions made in a hypothetical model. An example assumption and the associated implication are given below:

Assumption	Implication
<i>“The model has a time horizon of 1 year”</i>	<i>“By accepting this analysis, we accept that the relative and cumulative difference in costs and difference in QALYs will not change after one year of the intervention...a potentially serious limitation”</i>

(Clinician training booklet)

At the end of the session, clinicians were given an additional booklet, which contained definitions of a number of economic concepts, including terms referring to the effectiveness of interventions, for example, utilities, and those necessary in the interpretation of model results.

6.3.2 Structural development

The design and development of the model structure was initiated by I26, and discussed and validated by the clinicians. Most discussion regarding pathway plotting took place within one clinical meeting, although checking and changes to the structure were addressed in future meetings within the process. The next sections report on how the pathways of the structure were generated.

6.3.2.1 Beginning structural development

The initial development of the model structure was embarked on by I26 prior to any specific discussions with the clinicians. I26 drafted the structure based on the findings of an economic literature review, using existing model structures in Condition 2 to inform how patients were divided in terms of the model pathways. This involved separating patients first by their true conditions, and next by whether or not their conditions were identified by the relevant diagnostic test. The impression was that beginning a decision tree by delineating patients according to their true diagnosis was a common way to structure diagnostic models.

“...based on what I’d seen with other diagnostic models...so splitting out the population by whether they’ve got [Problem A] or not, and then if that’s identified or not....” (I26, interview)

I26 modelled the remaining patient pathways on the information provided in the clinical literature, looking particularly at which treatments patients received after a positive or negative diagnosis for Problem A. I26 implied that the treatment branches of the pathways were important in representing the consequences associated with patients receiving

misdiagnosis and mistreatment.

“...with a diagnostic model it’s basically just looking at the costs of who gets identified, and the main things are missed [Problem A] and the people who you are over-treating” (I26, interview)

6.3.2.2 Pathway plotting

After developing a draft of the model structure, I26 presented it to the clinicians for their feedback. Using a diagram of the model, he guided the clinicians through the progression of patients in each arm of the decision tree.

“...so initially patients are split by whether they’ve got [Problem A] or not, and then if they’ve got [Problem A] whether they have [Symptom A], and then comes in the Diagnostic test B or C, so whether it’s positive or negative, and then treatment follows on from that. So if they’ve got a positive diagnosis and they’ve got [Problem A], they have [Treatment 2] followed by [Treatment 1]...”

(I26, Clinical meeting 2)

The clinicians were then asked a number of questions on the structure. Initial questions focused on the nature of treatments that different groups of patients received. This was particularly to establish how interventions should be represented in the tree, but also where there were differences in patient costs and outcomes, pertaining to where patients needed to be divided within the structure.

I26: *“... (for patients with) missed [Problem A]...will [Treatment 2] be required, so would that be different [Treatment 2] for those identified with [Problem A] or would it be reduced outcomes for delayed [Treatment 2]?”* **C5:** *“...the [Treatment 2] is much more complicated....because you go onto get secondary problems... and (there are) major cost implications for treating those, and loss of quality of life...”*

(Clinical meeting 2)

Further questions were aimed at confirming whether pathways were valid in terms of the way in which patients were treated at various points.

I26: *“...is there a reason why we don’t have a no further treatment option, so after the initial indeterminate Diagnostic test A, we just take that to be negative and send*

people home?” C4: “...well if it’s negative but you’re still suspicious clinically then you can’t discharge that patient, that patient has to go through a patient pathway”

(Clinical meeting 2)

Discussion in pathway plotting

Discussion between clinicians over model structure allowed for further validation of the patient pathways, and highlighted variation and disagreement in the clinical practices undertaken. The excerpt below demonstrates clinician disagreement in terms of defining the current practice pathway.

C5: “when they come through the department I think ‘well the Diagnostic test A is positive, there’s a [Problem A]’, is it [Symptom A] or not? Or the Diagnostic test A is negative and I’ve got a low level of clinical suspicion, in which case there’s a pathway back out the door really...” C3: “...from [my] point of view once you’ve [tested] someone for [Problem A], we are obliged to put them on a pathway...I don’t think it would ever be ‘you’ve had a negative Diagnostic test A, go home’ because you need to take the negative findings and determine them” C5: “... you always refer the patient into some form of follow-up?!” C3: “Yeah...”

(Clinical meeting 2)

On a number of occasions the clinicians reported using different practices to one another in terms of how they treated patients. When determining which clinical experience should be represented within the model structure, the clinicians often decided amongst themselves on the most common practice to include.

C9: “.... it varies from clinic to clinic as to whether they come back, have an examination and then Diagnostic test A or whether they come back, complete [Treatment 1], have Diagnostic test A and then get an examination....But I think quite a lot of people get a second Diagnostic test A even if they are asymptomatic because they are not examined until Diagnostic test A is done...” C3: “So I think the answer is ‘yes, the second Diagnostic test A gets done before the negative clinical examination, and in the presence of two negative Diagnostic test As and a negative examination and follow-up, those patients will get discharged”

(Clinical meeting 2)

6.3.2.3 Boundaries of model structure

The boundaries of the model structure were determined at the beginning of the process, alongside the development of a model protocol. The protocol was generated through discussion between the clinicians, and designed to capture the components that the clinicians' considered important to include in the structure. The population was determined to be patients who had suspected Condition 2 and had received an indeterminate first Diagnostic test A. The clinicians determined this to be the point at which adverse events could be avoided if patients were given optimum treatment. Clinicians were instructed by the modellers to capture all relevant comparators, and the five diagnostic strategies included represented all of the possible combinations of diagnostic tests that patients with suspected Condition 2 could receive.

The time horizon and type of model used were determined by the modellers, although the former was discussed with the clinicians for their agreement. The important factor in determining the time horizon appeared to be the nature of the condition, and how long the symptoms or the outcomes of the treatments remained.

***I26:** "about the time horizon, so in the original protocol... [We had] [condition specific] outcomes, so does the model need to have a long time horizon to incorporate those?" **C5:** "Yeah but I'd have no idea what that time horizon should be" **I26:** "Lifetime?" **C3:** "I think lifetime" **C5:** "I think lifetime would be the sensible thing to go on....even people with subtle [condition specific outcomes] allegedly go onto have quite significant disability later on in life..." (Clinical meeting 2)*

The use of a decision tree structure appeared to be at the preference of I26, who suggested that he selected the simplest structure that was able to represent the decision problem.

"...I thought about the progression of [Condition 2], which makes you think about a state transition model, but then it was going to be quite difficult to identify different states... I thought the decision tree captured the main important aspects" (I26, interview)

6.3.2.4 Basis of model structure

The model structure was based on a combination of I26's research of the literature, clinician input, and knowledge of the data available to inform parameters. C4 was of the opinion that pathways should reflect clinical experiences and how patients were treated in reality.

"...it's feeding in your own personal experience of working through that medical process in real life....it's deconstructing what you do on a day to day basis and trying to codify that in terms of the model" (C4, interview)

However, C2 suggested that pathways, even if reflective of clinical practice, should still be supported by evidence.

"...that tension between what I think is happening, what I think is feasible and actually....will there be the evidence to support that?" (C2, interview)

When determining the focus of the model, clinicians were encouraged to consider whether there was evidence to inform the structure, and a search was undertaken to check for relevant data.

"...we had looked through the literature so we knew what the evidence base was for the question and that's where we said 'we have some evidence for this but perhaps this is where we should build a model so we've got more evidence...'" (I8, interview)

However, the general attitude was that the structure should not be influenced by the clinical evidence available, as C9 relayed the importance of the boundaries of the structure being agreed upon prior to model development, to avoid bias. Similarly, I8 gave the impression that structural development must prioritise developing pathways that are clinically representative from the perspective of the clinicians.

"We have to define the protocol first and then go away and find the clinical literature; otherwise it will be biased by the literature" (C9, clinical meeting 1)

"...I don't care if there are data available because I think 'well that can be sorted out'...you can put a data entry in at zero to 1 and just ignore that part of the structure. But if [an aspect of the structure] suddenly becomes important you can't just add it in, so I also feel for completeness, for representation to the people who are trying to understand this problem... [It's] quite important..." (I8, interview)

There were circumstances during pathway plotting where the clinicians suggested excluding pathways from the model, on the basis that there would be no evidence to populate or support them. However, I8 emphasised the importance of keeping pathways in the structure that might impact the overall cost-effectiveness result. In the example below, I8 stressed the need to maintain the aspect of the structure that accounted for the adverse events associated with each comparator, despite the clinicians' perception of a lack of evidence on this element. The impression was that I8 considered this important for the comparison of the differences in cost-effectiveness between the strategies.

I26: *“Can those with a false positive diagnosis have [Treatment 2], those that don't have [Problem A]?”* **C4:** *“It could happen; it's a bit of a whoops moment”* **C10:** *“For the purposes of evidence I think you can say no...”* **C5:** *“...I've certainly heard of a case though...like hen's teeth, rare, but the ones you see stick in your mind”* **C10:** *“I think on the basis of evidence you can say no....”* **I8:** *“...but that box still needs to be explored to compare Diagnostic test C against how many people would end up there with Diagnostic test B or Diagnostic test A”* **C9:** *“Yeah with Diagnostic test A you'd get people in that box...”* **(Clinical meeting 2)**

6.3.2.5 Making structural assumptions

I26 suggested potential structural assumptions to the clinicians for their feedback. Most structural assumptions appeared necessary due to a lack of long-term outcome data, particularly on quality of life. The modellers proposed limiting the structure and finishing patient treatment and recovery at the point at which there were no longer data available to populate parameters.

I26: *“... assume that everyone with [Condition 2] has a quality of life based on that patient reported [Measure] of [Value 1] for the first year and then they go to perfect health if it is identified, if it is not identified then assume [Value 1] for the lifetime...”* **C10:** *“So does that make the assumption that treatment works and continues to work for the rest of that patient's life?”* **I26:** *“Yeah so with missed [Problem A], patients will stay at the 1 year [score] for the rest of their life”*

(Clinical meeting 4)

Clinicians discussed the validity of the suggested assumptions, and in most cases agreed with what the modellers were proposing. Where there was disagreement between the clinicians, there were instances of the modelling team making the final decision on the assumption used. In this example, the modellers opted to include the perspective that seemingly allowed for the consequences of misdiagnoses in the current practice arm to be accounted for, instead of patients receiving further testing as argued by C6.

I26: *“Another assumption was that in current practice if patients had an indeterminate Diagnostic test A.... [If] they needed [Treatment 2] then they were considered to be missed and they would have continued reduction for their lifetime”*
C3: *“...there is the potential for them to be erroneously discharged because the Diagnostic test A is not 100% sensitive...”* **C6:** *“Is there? Because a lot of centres including us do initial Diagnostic test A, follow-up Diagnostic test A and then on to Diagnostic test C...”* **C3:** *“But it is done on clinical grounds so if it’s tender you [keep them in], if it’s not tender and you’ve got two negative Diagnostic test As you discharge them...”* **(Clinical meeting 4)**

Despite clinicians being generally accepting of the assumptions proposed, I8 emphasised the importance of having them officially ‘signed-off’. The clinicians were therefore asked to agree to relevant sections of previous meeting minutes.

“we’ve got to get all the clinicians to buy in and so that’s why we use our meeting time because that’s the only time we’re on record...and when they all say ‘yes’ as a group that means we can sign it off and we’ve got agreement on that particular assumption from the entire group” **(I8, interview)**

I8 gave the impression that clinicians were asked to sign-off on assumptions because they had a tendency to question them at the results stage.

“... That’s when it all comes out of the woodwork, until they’ve seen those results and they disagree with those results, they don’t question what’s going in and then all of a sudden you get a flurry...” **(I8, interview)**

6.3.2.6 Communication in structural development

Communication in structural development worked in terms of the modellers asking the clinicians questions to check aspects of the structure, and the clinicians subsequently

discussing the validity of the pathways. I26 suggested that he used the method of developing the structure and presenting it to the clinicians because it was the most efficient approach in the context of his workload and limited clinician contact time.

“I’ve just tried to do it as quickly as I can really because we are doing [numerous models]I just present as much as I can to them in each meeting” (I26, interview)

There were issues with gaining access to clinical opinion in structural development, specifically in terms of establishing an economic subgroup. The modellers reported that a subgroup would typically involve three to four clinicians, who would offer more in-depth advice on, and validation of, structural inputs. However, I26 reported only being able to recruit one individual, due to the clinicians’ conflicting schedules.

“...I haven’t had a huge amount of contact; it’s been mostly with C5 because it’s been difficult to get people together at the same time...” (I26, interview)

Despite this, both I8 and I26 suggested that C5’s sole involvement in the subgroup was unlikely to have a negative impact, given that C5 typically took on a leading role in discussions. However, the lack of negotiation with others was sometimes problematic, as C5 made suggestions for the comparators of the model that other clinical experts later disagreed with.

“... people mentioned that we should look at a strategy of immediate Diagnostic test C or Diagnostic test B without the initial Diagnostic test A, C5 thought this was wrong [because] the patient [would] have to go home and come back for an additional visit. This was raised with [the other clinicians] (and they) said these were feasible strategies” (I26, interview)

Other issues with communication concerned the clinicians’ lack of understanding of the initial nodes in the decision tree, which divided patients according to their true conditions. This was clarified through an explanation from I26.

C3: *“I’m not sure you need the negative nodes in this pathway because if it is [Problem A] it’s either [Symptom A] or not [Symptom A] and then they go on to the [different treatment] options...”* **I26:** *“the first bit is whether they’ve got a true*

[Problem A] or not, that's what you know, so then the positive [Diagnostic test] is whether it's picked up" C3: "So you have an Diagnostic test C or Diagnostic test B, the patient really has [Problem A] but there's a possibility that the [Diagnostic tests] might not have picked that up" I26: "Yep" (Clinical meeting 2)

6.3.2.7 Implementation

I26 suggested that he planned the structure on paper prior to its implementation in the modelling software. I26 stated that developing a plan and protocol of the model at the beginning of the process ensured that he had a complete structure to later program.

"...it's planning for problems, I find it hard to look ahead at things that are going to crop up, that's why I think it's important to have a plan of the model....rather than just start doing it and see what happens" (I26, interview)

6.3.2.8 Reflections on model structure

All informants were asked to reflect on the model structure. The attitude of the modellers seemed positive, as both I8 and I26 gave the impression that they were happy that the structure was providing a clinically valid representation of the decision problem. However, C4 suggested that he was struggling to understand why the structure was representing clinical events that in his opinion were unlikely to occur.

"some of those scenarios seemed to be things that just wouldn't happen in real life i.e. if you went down a certain pathway you could perhaps end up in a scenario in which a surgeon would operate on a [Problem A] that didn't exist..." (C4, interview)

6.3.2.9 Summary of structural development

Structural development was generally initiated by the modellers, and reviewed and discussed by the clinicians. The modellers tended to ask questions aimed at validating pathways, and I26 reflected that this method was efficient in the context of limited time and access to clinical opinion. Discussion between clinicians appeared to be a key communication tool in terms of ensuring the clinical representativeness of the pathways, but also intrinsically addressing the generalisability of the structure, as the clinicians resolved disagreements by negotiating on what they considered to be the most common or valid clinical practice. The process appeared

to benefit from the large group of clinicians involved in immediate model development, but having only one clinician involved in the subgroup demonstrated that one opinion may not be representative of all. Some tension was identified between the modellers and clinicians in terms of the aspects that were included in the structure. Whilst the clinicians were fundamentally concerned with ensuring that the structure was clinically valid, the modellers additionally were striving to keep pathways that could potentially demonstrate the comparative cost-effectiveness of the strategies being evaluated. The clinicians appeared, at times, to struggle with understanding the components of the structure that were important to the economics of the model.

6.3.3 Data and populating the model

Data used to populate the model were identified using various means, including systematic reviews of the literature, ad-hoc literature searches, searches in relevant databases, and clinician opinion. The next sections report on the methods used to identify and check the clinical and economic data in the model.

6.3.3.1 Clinical data

The clinical effectiveness data, which referred to the diagnostic accuracy of the three tests used in the strategies, were searched for using a systematic review. The accuracy value for the first diagnostic test was taken from a relevant paper retrieved. The modelling team had to undertake further ad-hoc searches to identify a value for the second diagnostic test, as this did not feature as a comparator in the original paper. The third test, Diagnostic test C, was assumed to be the reference standard, and as such the sensitivity and specificity values were set to 100%.

The probability values originated from a variety of sources. Two of the values, including the prevalence of Condition 2, were informed by clinical papers identified through general

searches, whilst the remaining inputs came from clinician opinion. I26 gave the impression that the use of unsystematic searches for these values was not ideal, but necessary due to time constraints.

“...we don't really have time to do a systematic search for each parameter in the model so we just do a general search and try and find the best evidence we can, that's the main thing I would improve” (I26, interview)

6.3.3.2 Economic data

The process for identifying costs and utilities for the model was initiated with a systematic review aimed at finding economic evaluations to inform the structure of the model, and potentially data to populate parameters. However, the search returned only one relevant paper, which included no quality of life outcomes. Utility data were therefore identified using a “quick search” of the literature, retrieving a relevant patient reported value to inform the quality of life of a patient one year post Condition 2. Due to a lack of long-term quality of life values, it was assumed that patients would remain with the one-year value for life if their problem was not identified, or return to full health if they were diagnosed appropriately. Although these values and assumptions were reviewed by the clinicians, I26 later identified the lack of appropriate utility data as the greatest source of uncertainty in the model.

“I think the long-term benefits, the health outcomes, they came from papers but they were small studies and it was based on assumptions from the clinicians as well, so in assuming that the health state carries on...” (I26, interview)

Cost and resource use data were identified using NHS reference costs and the HRG database. I26 searched for relevant costs within the database according to the condition being modelled and type of medical care. The cost data were presented to the clinicians and questions were focused on whether the clinicians recognised the costs to be valid, and whether appropriate HRG codes had been used. The clinicians who answered these questions seemed familiar with the codes, and were aware of the differences between cost categories.

I26: “...there are two different costs, there’s category 1 and category 2. I’m not sure what the difference is. I was hoping someone here might know?” **C10:** “It is comorbidities, so the comorbidities push you into category 2...” **I26:** “So do we want category 1 or?” **C5:** “Category 1 would be more representative...”

(Clinical meeting 4)

While most of the cost values were considered appropriate, some of the clinicians questioned the transparency and accuracy of the NHS reference costs. However, all parties seemed to accept that these data and the data source were the best available.

C7: “For Diagnostic test C does it refer to [Type 1] Diagnostic test C or [Type 2] Diagnostic test C? There will be a cost difference.” **I26:** “In the costs it just says Diagnostic test C, it doesn’t specify” **C8:** “It’s interesting though because it would be significant” **C10:** “Significant on every level because the [Type 1] is cheaper to maintain” **C8:** “Well actually ... the capital costs are lower but the running costs are the same or higher” **I26:** “So you think that cost would be more?” **C10:** “No I think that the figures you’ve found are what we should use, there’s nothing better published”

(Clinical meeting 4)

6.3.3.3 Data and clinician involvement

Clinician involvement in populating the model mostly concerned the checking of data values, with all of the clinical and economic data inputs being reviewed by the clinicians prior to their inclusion in the model. Clinicians were also asked to recommend sources for data parameters, mainly in circumstances in which a systematic review had not been undertaken. I8 gave the impression that the clinicians would have more confidence in the model results after being involved in data collection.

“If you have any papers around this then please add them on, we are not doing a systematic review and so if we have other numbers these will impact on the results...I’d hate to think that you won’t believe in the model outcome because of the numbers being used” **(I8, clinical meeting 3)**

In circumstances where data were not available in the literature, clinicians were asked to provide data estimates for parameters. Around one third of the base case inputs were informed by clinician opinion. The modellers asked open questions regarding the values to use, and

based final estimates on consensus among the clinical experts. Where there was disagreement, as in the example below, the modellers eventually used the suggestion supported by most clinicians, including C5, in the main model, and increased the accuracy of the follow-up test in the sensitivity analysis.

I26: “...so we don’t have data on the accuracy of follow-up Diagnostic test A, can we make assumptions about that?” **C5:** “I’m not aware of any evidence to show that the follow-up Diagnostic test A is any more accurate than the first” **C3:** “It’s probably a little bit more sensitive because you do get some [Problem As] that don’t show up and then do show [at the follow-up]” **C5:** “Do you? After eight years of looking for them I never saw one.... [I’d] assume that the second Diagnostic test A is no more accurate than the first”

(Clinical meeting 2)

6.3.3.4 Summary of data and populating the model

The modelling team used a combination of systematic reviews, general searches and clinician opinion to identify parameter inputs for the model. Systematic reviews were used in the first instance to search for clinical effectiveness data and economic evaluations that could potentially inform economic data parameters. However, both searches were limited by a lack of relevant evidence. General searches were undertaken to retrieve clinical probabilities and utilities and cost data. Collecting quality of life data was problematic, as I26 stated that the lack of long-term utility data was the greatest source of uncertainty within the model. Further, the clinicians were concerned with the accuracy of NHS reference costs, although they were also of the opinion that the HRG database was the best and most practical resource currently available. The clinicians were involved in checking data, highlighting relevant data sources to inform parameters, and making data assumptions for the model where there was no literature available. Discussion between the clinicians appeared to be essential in ensuring the validity of data assumptions.

6.3.4 Results

After populating and running the model, the model results were presented to the clinicians.

I26 initiated the presentation with a reminder of the data inputs to the model, and an explanation of the concepts used to interpret the results.

“...often we present cost-effectiveness as a cost-effectiveness ratio and compare it to the £20,000 per QALY threshold set by NICE, but to present it in a simpler way using the net benefit approach, we take that £20,000 QALY threshold, multiply it by the number of QALYs we have for that strategy and that gives us what we value those QALYs at in monetary terms. Then we take off the cost of the strategy and the optimal strategy will be the strategy with the highest net benefit” (I26, clinical meeting 5)

For context, the main model demonstrated that Diagnostic strategy 5 was the most cost-effective. The modellers discussed the results with the clinicians, allowing them to ask questions to aid their understanding of the findings, and explaining how the differences in costs and effectiveness between the strategies impacted cost-effectiveness overall. However, a number of the clinicians appeared to struggle with interpretation, specifically of the research and economic terminology. As a result, the modelling team had to provide clarification.

I26: *“...this is the base case analysis... the maximum net benefit is from the Diagnostic strategy 5, so that’s based on the analysis where missed [Problem A] has reduced quality of life over a lifetime....”* **C9:** *“So base case analysis means?”* **I26:** *“...not the sensitivity analysis”* **C2:** *“can I ask why the Diagnostic strategy 2 and Diagnostic strategy 3 cost is higher? Is that presuming there is a follow-up appointment as well...?”* **I26:** *“...they have another attendance”* **I8:** *“Your Diagnostic strategy 5 accuracy estimates are incredibly high because that is your reference standard, so you’re avoiding having people go off on the wrong route... that’s worth the additional cost of Diagnostic test C”* **C9:** *“...I may be being stupid here....tell me what base case means....?”* **I8:** *“....base case is based on your best possible estimate...then what you do is change a parameter a little bit either way just to see how the conclusions might change”*

(Clinical meeting 5)

The model results seemed unexpected to the clinicians as they questioned the differences in cost-effectiveness between strategies, and subsequently the data inputs used. The clinicians appeared unhappy with the value used for the accuracy of Diagnostic test B, and the

assumption that Diagnostic test C was the reference standard. These were inputs that the clinicians had previously agreed to. As a result, the clinicians requested that the accuracy of Diagnostic test B was increased to 100%, and the impact of this tested in a sensitivity analysis.

C9: *“I’m surprised that Diagnostic test B was so poor?”* **I26:** *“Diagnostic test B was poor based on missed [Problem A] because it was 86% sensitive”* **C10:** *“...only 86% sensitive?”* **C5:** *“Compared to Diagnostic test C but you could argue Diagnostic test C over-diagnoses”* **C10:** *“So should we say actually Diagnostic test B is 100% sensitive for [Condition 2]?”* **C5:** *“...You can do that on the sensitivity analysis...”*

(Clinical meeting 5)

Other issues were not as straightforward to resolve; during a preliminary results meeting, clinicians questioned the validity of the model pathways, specifically the current practice arm. The clinicians were concerned with the credibility of the assumption that patients would leave the model after receiving a second negative diagnostic test, thus experiencing permanently reduced health outcomes. Interestingly, this was the structural assumption that had earlier caused disagreement among the clinicians (see section 6.3.2.5), with the final decision being made by the modellers, rather than explicitly between the clinical experts.

C9: *“[Current practice] ...if you just do that alone then we are going to miss people, that’s what we are saying isn’t it? In practice I don’t think that happens because most people would never do just Diagnostic test A and follow-up...”* **C9:** *“... [If] you’re still symptomatic you’re going to go on and [have]... Diagnostic test C”* **C10:** *“so this model is saying that we are not doing that, this is saying that we accept the findings of the follow-up Diagnostic test A and send them home”* **C9:** *“...if we are saying that normal practice misses [Problem A] and someone says ‘well how do you know that?’, we’ve got no evidence...what we have got evidence to say is that up to the point that you’ve taken the second Diagnostic test A you may not have found everything, but normal practice isn’t to discharge at that point unless there’s no clinical problem...”*

(Clinical meeting 6)

The clinicians’ concern appeared to be that the clinical evidence did not adequately support the developed pathways, and subsequently the adoption of a new diagnostic strategy.

However, the clinicians were happy that the result could be justified on the basis of cost difference alone.

R1: “we have done our best to find how many patients would have been missed in the current practice strategy; we would have included it if we found it.... [But] it looks like current practice is more expensive anyway...” **C9:** “So we can justify it on the basis of the cost without looking at the long-term consequences?” **I26:** “Yeah” **C9:** “And if there are long-term consequences because we miss some people who go through without ever having [further diagnostic testing] then that would only be to our advantage.”

(Clinical meeting 6)

6.3.5 Model checking

Most model checking was undertaken by I26, and mainly involved sensitivity analysis. I26 undertook deterministic sensitivity analysis on almost all of the data parameters, to test the impact that it would have on the model results. Many of the alternative values used were informed by clinician opinion, and some of the scenarios tested were at the clinicians’ request, for example, increasing the accuracy of Diagnostic test B to 100%. The clinicians appeared to benefit from the opportunity to discuss the outcome of the sensitivity analyses with each other, seemingly aiding their acceptance and interpretation. The excerpt below demonstrates the clinicians’ discussion of the effect that decreasing patients’ reduced quality of life had on the model.

C10: “at the moment we are saying that patients had a [false negative] diagnosis, they remain with these symptoms. But then I26 has modified that by saying for only 1 year, or 2 years....” **C9:** “...basically the long-term effects are not in play?” **C10:** “Well they’re just in play for current practice really” **C2:** “so the difference between Diagnostic strategy 4 and Diagnostic strategy 5 is just the difference between the costs, they’re both the same as each other, the mean cost is just lower for Diagnostic test B...And that’s why Diagnostic strategy 4 is coming out better” **C10:** “And the longer you say the harm lasts, Diagnostic strategy 5 comes out better”

(Clinical meeting 6)

The clinicians appeared to have little problem with the interpretation of the sensitivity analyses, and a few of the clinical experts demonstrated the ability to predict the effect that changes would have on the results prior to the check being undertaken.

“ And if you put that at 100% the results would change...the sensitivity analysis is going to turn it on its head, so that Diagnostic strategy 5 looks great if you use the base case but as soon as you say that Diagnostic test B is more sensitive, Diagnostic strategy 4 becomes more cost-effective...” (C4, clinical meeting 4)

I26 also undertook extreme value analysis as a means of internally validating the model.

Following a validation checklist, I26 populated clinical and economic parameters with “null and extreme values” to ensure that when the model was run, the cost-effectiveness result was changing in the direction expected. Other internal validation checks involved ensuring that data inputs were consistent with clinical outcomes, and that the modeller was able to explain model results.

6.3.5.1 External validation

External validation checks were also included on the checklist, with I26 comparing the outcome of the current model with the results of the published economic evaluation retrieved through the systematic review. It was reported that the conclusions of the published model were consistent with those in the published paper, although the identified study had not compared all five of the diagnostic strategies that had been included in the current model.

6.3.5.2 Clinician involvement in model checking

Clinician involvement featured as a fundamental aspect of model checking, as the clinicians were given the task of validating pathways, data inputs, model results and sensitivity analyses. In terms of structural validation, the impression given was that it was the role of the clinicians to ensure that the structure represented the clinical problem appropriately. It was stated in the model report that the structure had been clinically validated through continuous and iterative

consultation with clinical experts. The perspective that the clinicians were responsible for structural validation was also mirrored in comments from the modellers.

“Structure...I see that as a planning one, just through looking at it with the clinicians and trying to sort it out...through presenting ‘this is where I’ve got to so far, this is what’s coming out so far, what do you think?’” (I18 interview)

“...all I’ve done is taken the structure to the clinicians and used their knowledge to make sure that the structure is what it should be” (I26, interview)

However, I18 implied that further structural validation and sensitivity analysis should have been carried out, but that it was not because the checks were very difficult to undertake.

“...actually that’s one thing that we test the least because it’s so complicated and yet the structure dictates so much in terms of what is important, so I find that quite, but it’s just a really complicated issue...” (I18, interview)

6.3.5.3 Summary of results and model checking

The model results were presented to the clinicians for their feedback, with opportunity for the clinicians to clarify the findings with the modellers and each other. The clinicians gave the impression that the model result was unexpected, and subsequently questioned the data inputs and structural assumptions used. It appeared that there was some breakdown in communication between the modellers and clinicians over what was considered valid to include in the model, perhaps caused by a lack of explicit agreement on structural assumptions. However, discussion between the two parties facilitated a solution in terms of identifying a justification for the model results, and the clinicians’ requesting that alternative values be tested in the sensitivity analysis. Further, although initially a problem, the clinicians’ understandings of economic terms used in the interpretation of the results were enhanced through the clinicians’ ability to seek clarification from the modellers.

Model checking mostly focused on sensitivity analysis, with clinician understanding of the results of analyses again being enhanced by the opportunity for discussion with others. I26 undertook some external validation, and structural validation was carried out through iterative

checks of the structure with the clinicians. However, I8 suggested that further structural checks should have been undertaken, but stated that these were avoided because of their complex nature.

6.3.6 Clinician involvement in model development

The clinicians were involved in the structural, data and results stages of model development. Their involvement appeared to be valuable to the construction and checking of the model, but there were also some problems in communication between the modellers and clinicians. The next sections explore issues around the clinicians' involvement, and reflect on the number of clinicians involved, their understanding of modelling and the usefulness of clinician guidance.

6.3.6.1 Number of clinicians

The entire process of model development involved up to twenty clinicians, but the number attending each meeting varied. Nevertheless, there were a number and variety of clinicians involved in each of the decisions around the structure and inputs to the model. Both the clinicians and modellers were of the opinion that this was essential to the model's validity, credibility and generalisability.

“There’s a good mix of people, we’ve got people who specialise in all areas” (I26, interview)

“It’s clear from discussions around the table that things are very different across the whole country and I think...it’s actually a really important part of it in terms of validating the outcome, which is actually the strength and breadth of both experience and geography of clinicians” (C2, interview)

“...it would immediately undermine the credibility of what we do so I think it’s important to have these discussions, even though they can feel quite uncomfortable at times” (C3, interview)

C3's comment suggested that discussions between larger groups of clinicians could be difficult, as it appeared to increase the likelihood of disagreements. Although discussion was deemed to be valuable, the modellers gave the impression that the clinicians' conversations

could sometimes be difficult to manage. Both I8 and I26 reflected that there tended to be a few clinicians who regularly dominated discussions. I26 suggested that further measures needed to be taken within modelling meetings to ensure that every clinician had the opportunity to contribute.

“We’ve got a couple of argumentative clinicians in there....there’s always one or two like that in every group” (I8, interview)

“We get one person who specialises in a certain area and has a lot more input than other people, so I suppose it would be better if we tried to make sure that more people had an input. I don’t know how we could do that. We’d have to sort of make people say something ...” (I26 interview)

6.3.6.2 Clinician understanding of modelling

On occasion, the clinicians appeared to have difficulty in understanding aspects of the model building, for example, the economic components included and the terminology used. The clinicians also suggested that they struggled to understand how structural development worked and why the model was being developed as it was. C4 thought that his lack of understanding would later impair his ability to interpret the model result.

“... understanding about how it’s all pulled together, I’m not clear whether that’s something I’ve missed or something that I would never be expected to be told” (C2, interview)

“... those models are such an unfamiliar way of looking at information for me. The feeding the information in is something that I can understand, I don’t have an understanding of the shape that the model is taking yet and I suspect that even when the finished product comes out that it’s going to take some explaining...” (C4, interview)

I26 also suggested that the clinicians struggled to follow model development, particularly in understanding why certain aspects had been included in the structure. He proposed that the clinicians had this difficulty because they were only asked questions on specific aspects of structural development. I26 gave the impression that building and conceptualising the

structure in conjunction with the clinicians would have enhanced the clinicians' ability to understand the model.

"[getting] the structure to match what the issues are...I think the problem is the difference between the understanding of the clinicians and the modellers, and they don't always know what we are doing" (I26, interview)

"You do get some good feedback on specific bits of it but then they don't see the actual thing happening, so maybe if they did actually see it being built as it goes then that might [help]" (I26, interview)

6.3.6.3 Guidance for clinicians

All of the clinicians were offered training prior to commencing their roles, but none of the three clinicians interviewed attended due to work commitments. C3 gave the impression that it would have been useful if the training were more flexible and potentially web-based, to fit in with their schedules. While C3 suggested that knowledge of how models worked would have been helpful, C2 stated that she required clarification of economic concepts.

"...a dictionary of terms, almost a glossary of terms...." (C2, interview)

"I think some sort of interactive online training or a video might be quite useful, just about health economic models and the rationale of using different types with some examples, that would be quite good I think because then it would be something we could do in our own time rather than having to come down for a day's training" (C3, interview)

Both C4 and I26 also stressed the importance of the clinicians receiving continuous training, and being reminded of important aspects at various points of development.

"...they basically just forget when they've seen it once, so they need to keep going over it..." (I26, interview)

"...I've heard the technical terms used lots of times and I understand what most of the terms mean ...but until you start using something all of the time you never understand it..." (C4, interview)

C2 gave the impression that the nature and amount of training that clinicians required would depend on their previous experience of health economics and modelling.

“I think I’ve probably come into it a bit more confidently than others because of my research background, because one of the projects that I’ve been involved with...it includes health economics, an economic analysis section, so I’m used to the terminology. I think otherwise if someone was talking about QALYs and I’d never done any work on that before it would have been completely over my head” (C2, interview)

6.3.6.4 Summary of clinician involvement

The large number and diversity of clinicians contributing to model development was perceived to be essential to the validity and generalisability of the model, but the modellers gave the impression that it was more difficult to manage discussion between larger groups of clinical experts. The clinicians stated that they struggled to follow structural development, particularly why the model was being developed as it was. I26 proposed that the clinicians’ comprehension of what was important, and occurring in model building, could be improved through their involvement in the conceptualisation and development of the entire model structure. The clinicians interviewed thought that training would have further aided their understanding of model development, but none of those interviewed attended the training offered due to work commitments. The suggestion from the clinicians and the modellers was that training should be flexible, preferably web-based, and given throughout model development.

6.3.7 Conclusion

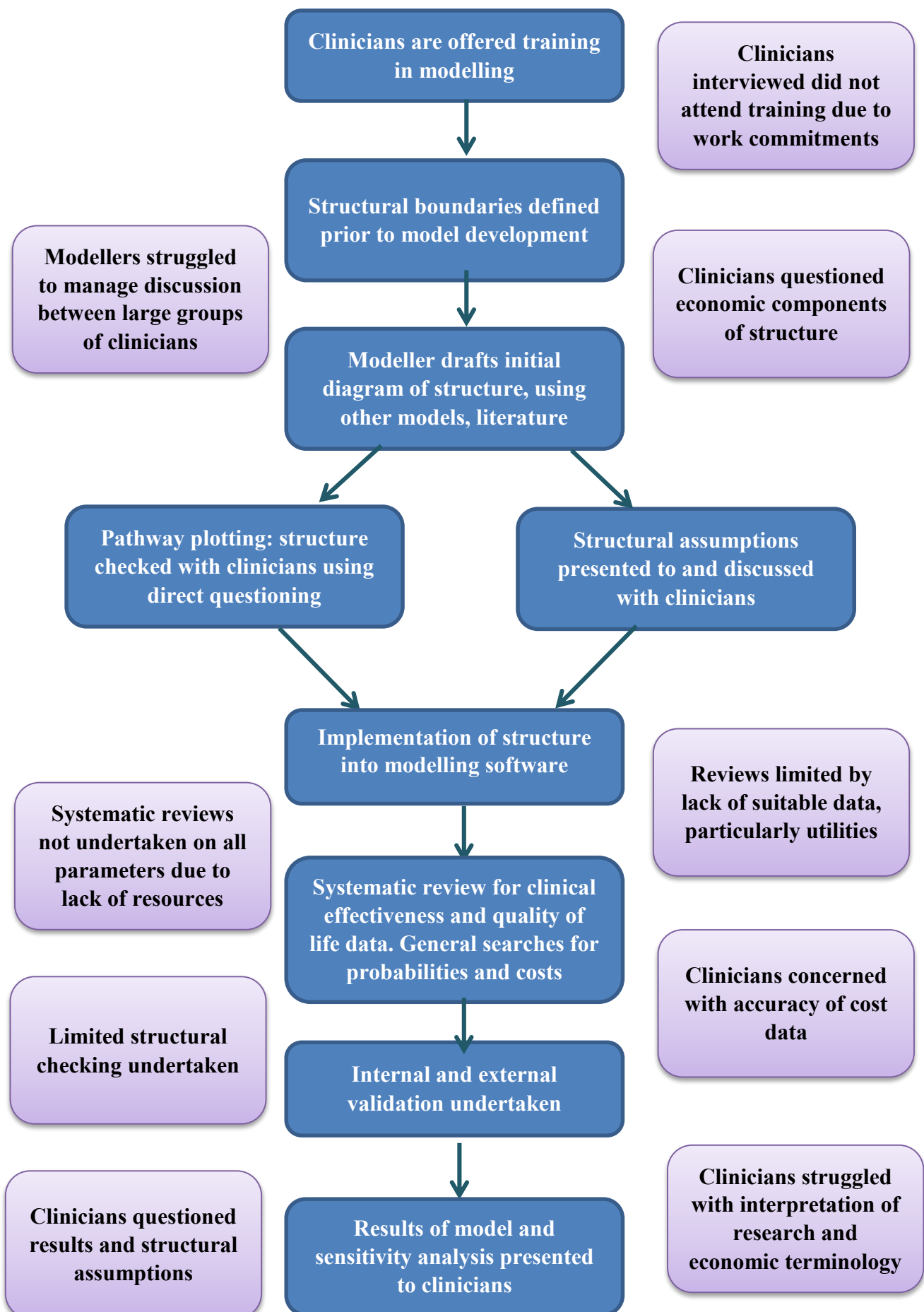
The second part of this chapter has presented the modelling process used by the modellers and clinicians in Case Study B. The findings of the observations and interviews have highlighted the modelling practices undertaken, and the issues and problems encountered in model development. These are demonstrated in the diagram in Figure 6.

Issues highlighted concerned the clinicians’ involvement in structural development. Although clinician input was fundamental to various aspects of model development, there were problems around the clinicians’ understanding of modelling. The clinicians questioned the

inclusion of economic components to the structure, and struggled with the interpretation of some of the economic concepts used. I26 attributed this lack of understanding to the clinicians' limited overview of the model-building process, and how the model was put together as a whole. Other issues related to communication, as clinicians questioned structural assumptions and model pathways fairly late on in the modelling process. This was seemingly due to the lack of explicit agreement between clinicians over structural assumptions earlier on in model development. Although discussion between clinicians was seen as beneficial to the model and clinician understanding, a further problem with communication concerned managing discussions between large groups of clinical experts.

Problems related to the data used in the model included the use of general and ad-hoc searches as opposed to systematic reviews. The modelling team undertook general searches on a number of parameters within the model, recognising this as a limitation of their practice due to the lack of time available. The modellers reported on the lack of long-term quality of life data, giving the impression that this was the source of the greatest uncertainty within the model. The clinicians were also concerned with the accuracy of the cost data used, suggesting that the NHS reference costs did not highlight the differences in cost between similar types of treatment. In terms of model checking, the modellers undertook structural validation only insofar as asking the clinicians to check various elements of the structure, with I8 suggesting that further structural checks could have been carried out. However, I8 also stated that alternative forms of structural validation were too complex to include in the process. A final practical issue in terms of the overall process related to the inability of the clinicians interviewed to attend the training offered by the modelling team.

Figure 6: Modelling process for Case Study B, with issues highlighted



After reviewing the processes used by the modellers and clinicians in Case Study A and Case Study B, the next chapter looks at their practices in relation to one another, to gauge whether the teams used similar methods and encountered the same problems and issues. There is also a focus on what can be inferred from the similarities and differences in the practices undertaken. The next chapter summarises the findings of both stages of the qualitative research, and considers the implications of these for current model development and future research.

CHAPTER 7: DISCUSSION OF FINDINGS

7.1 Introduction

The objective of this research was to explore the processes used to develop health economic decision-analytic models. These are models used by healthcare decision-making bodies such as NICE, to assess the costs and outcomes of health interventions, and provide an assessment of which offer the most value for money (NICE 2013). The research was undertaken in light of literature that highlighted errors in the processes and results of published decision-analytic models (Chilcott et al., 2010; Karnon et al., 2007; Roberts et al., 2006), and a systematic review suggesting that there was limited guidance available on particular aspects of model development (see chapter 2). The aims of the thesis were therefore to investigate good practice in model building, and explore issues and problems that were occurring in the process. Further, the intention was to examine model development from the perspective, and through the experiences, of those who were involved, namely modellers and clinicians. This was possible through the use of qualitative methods, specifically in-depth interviews with modellers, and case studies with teams of modellers and clinicians, involving non-participant observation, think-aloud and qualitative interviews. A qualitative approach was selected following a review of qualitative studies in other areas of health services research and health economics (see chapter 3), which suggested that in-depth and exploratory methods were particularly useful in highlighting and understanding methodological issues occurring within healthcare processes.

This chapter begins by presenting a summary of the results of each of the two phases of qualitative research undertaken within this thesis, and provides a synthesis of the main findings across the in-depth interviews and case studies. The findings of this thesis are then

discussed in relation to the results of similar qualitative studies, and following this, the implications of the findings for various aspects of model development are considered, and recommendations for good practice are made. The final sections of the chapter outline suggestions for future research and reflections on the methods used.

7.2 Summary of findings of the qualitative interviews with modellers

A total of twenty-four in-depth interviews were conducted with modellers in the UK and in Canada. The modellers interviewed included those from academic and non-academic backgrounds, and those who were in senior and more junior roles within their organisations. Interviews were detailed and inductive in that informants were encouraged to talk at length about their experiences, and raise issues that they felt were important to the aims of the research. The use of constant comparative methods of analysis allowed the practices of the informants to be compared for similarities and differences in terms of the methods undertaken; common problems were also examined.

The interviews captured rich data on the processes used by individual informants. Modellers were following similar stages in model development, and clinician involvement was considered to be a key aspect of the process, particularly in receiving input to model structure. Although modellers undertook similar stages within structural development, different approaches were taken to these stages. Informants discussed different ways of involving clinicians, including when they sought clinician opinion, and how many clinicians they talked to. A few modellers did not involve clinicians until late in model development, which was reflected on as poor practice. While most informants involved only two clinicians in model development, some spoke to just one, or even no clinicians, with the latter practices raising concerns among the informants around the generalisability of developed structures. This was

in addition to the lack of measures reported by informants to ensure that a range of clinicians, i.e. those from a variety of backgrounds and locations, were involved.

Other differences in structural development related to the sources used to inform model structure. Modellers had conflicting opinions about whether the data used to populate a model should influence structural development, and whether existing structures should be used to directly inform a new model. The academic and non-academic consultancy modellers had slightly different foci in terms of what they considered important to include in a structure. Academic informants suggested that the structure should be based on all clinically valid health states, whilst some non-academic informants implied that the structure should contain the states that were most important to patients, namely those that had most impact on their clinical and quality of life outcomes.

Common problems were identified through informants' discussions of their practices, and their perceptions of where and how processes could improve. The overarching issues that were reported related to communication between modellers and clinicians, and a lack of time and resources to undertake robust practices. Informants cited problems with engaging clinicians in model building, suggesting that clinicians' lack of understanding and acceptance of modelling made it difficult to involve them in any depth. Informants also criticised aspects of others' processes, such as using less robust methods for identifying data to populate a model, and carrying out limited model checking activities. However, some of the modellers interviewed appeared to lack robust practices in these areas, as external validation was carried out by fewer than half of informants, and structural checking (outside of clinician validation) was not undertaken at all. Those who gave a reason for their limited model checking cited a lack of time to carry out additional activities.

7.3 Summary of findings of the case studies with modellers and clinicians

Two case studies were undertaken with two teams of modellers and clinicians, to observe modelling processes, and identify any issues with the methods used. The case studies were selected because they offered insight into potentially different processes, capturing the development of an academic model and policy model from beginning to end. The teams of modellers sampled were from among those who had a good reputation for model building, on the basis of recommendations from the interview informants. It was anticipated that problems encountered by these modelling teams were also likely to be experienced by others. Both case studies used non-participant observation and qualitative interviews to gain a detailed understanding of each model's development. Case Study A additionally used the think-aloud method to explore thought processes around model checking.

The approach to inquiry was inductive; however, the case studies were also designed to investigate further the issues raised during the first phase of the research. Particular attention was paid to the role and contribution of clinicians in the modelling process, given that this was identified in the qualitative interviews to be an important but problematic aspect of model building. The use of the framework method for analysis allowed case study data to be coded according to pre-existing concepts, in addition to permitting the development of new codes and themes. Case studies were analysed separately, but framework analysis facilitated the comparison of the modelling methods used, highlighting similarities and differences in practices, and common problems identified. Although the findings of the individual case studies provided useful insights into issues with model development (see chapter 6), comparing the findings allows broader inferences to be made about the potential benefits and drawbacks of the different modelling approaches taken. These are explored in the next section.

7.3.1 Comparison of the case studies

The case studies were similar in terms of the overall stages of model development followed. The modelling processes included initial structural development by the modeller, clinician discussion and checking of pathways, populating the model with data, model checking activities and presenting the results to the clinician(s). However, different methods and approaches were undertaken by the two modelling teams.

Modellers from both case studies used similar methods to inform their model structures, as initial structures were developed with reference to clinical literature and/or other economic models. In both cases, structures were drafted into diagrams of patient treatment pathways, to be discussed with the clinician(s). Clinical input was then used in both case studies to verify and develop model pathways, with the NICE guidelines additionally being drawn on by the team in Case Study A.

Much of structural development in both case studies focused on gaining clinician opinion and feedback on model pathways. In Case Study A this input came from just one clinician, whilst up to twenty clinicians were involved in discussions of model structure in Case Study B.

Modellers' discussions with clinicians addressed similar structural issues; however, clinician opinion was sought in different ways. In Case Study A modellers asked the single clinician to talk through and clarify all patient pathways, whilst Case Study B modellers concentrated on asking questions on specific aspects of structure.

Both teams of modellers developed the structure prior to using the data that would populate the model, with the impression from both case studies being that the structure would be biased if it were to be based on this data. For this reason, the team in Case Study B decided on the boundaries of the structure prior to knowing the data available. In Case Study A, structural boundaries were applied iteratively, with modellers, for example, excluding groups of patients

as model development progressed. However, these boundaries were applied rather inconsistently, as the modellers varied their definitions of what they considered to be a non-typical patient.

The approach to communication with clinicians differed across the two case studies, as modellers in Case Study B encouraged clinical experts to have an influential role in decisions around model development. Clinicians were asked to determine model boundaries and discuss and validate structural assumptions. This contrasted with Case Study A, in which the clinician was permitted little negotiation about, and given limited explanation of, simplifying assumptions made. This approach to communication seemed to impact the clinician's understanding and acceptance of the model, as she later reported feeling unhappy with the clinical representativeness of the structure.

Discussions between the clinicians in Case Study B seemed to work well in enhancing the generalisability of structure, as variation in clinical practice among the clinicians was highlighted. The involvement of only one clinician in immediate structural development in Case Study A raised issues around generalisability, as the clinician found it difficult to give a representative overview of treatment pathways. In contrast to Case Study B, discussion between clinicians could not be used to resolve uncertainty around treatment pathways, or to enhance the clinical validity and generalisability of structure. Although a wider group of clinical collaborators were involved in Case Study A, these experts were reported to have had limited input to pathway plotting.

The methods used to identify data to populate the models differed according to the context of the projects; however, there were similarities in the search strategies employed, as both modelling teams attempted systematic or 'thorough' searches of the literature to identify clinical (effectiveness) and economic data. General searches were used in Case Study B to find clinical probabilities, with the main modeller commenting that the team did not have the

time and resources to conduct systematic searches for all parameters. The teams encountered similar problems in being unable to find appropriate utility values for their models, and as a result incorporated a similar assumption that patients who were not cured of their conditions would maintain their initial quality of life for the entire model. Cost data were identified through NICE reference costs in both processes, with the clinicians in Case Study B criticising this resource, implying that cost figures were not specific enough.

Most model checking within both case studies focused on sensitivity analysis, with internal validation also carried out. External validation was undertaken in Case Study B, but not in Case Study A due to a reported lack of relevant economic studies. In both case studies, limited or no structural checking was undertaken outside of clinicians' structural validation. The modellers in Case Study A discussed undertaking various structural sensitivity analyses, but only eventually tested one alternative structural scenario. In Case Study B, the senior modeller stated that alternative structural checks should have been carried out, but were not, due to their complexity.

When the results of the models were presented to the clinician(s), both case studies encountered problems with the clinicians' understanding and acceptance of the outcome. In both cases, this appeared to be connected to problems with communication between modellers and clinicians. In Case Study A, the clinician expressed dissatisfaction over the results, stating that the outcome was not as she was expecting. Although this appeared partly due to the clinician's preference for a particular result, factors pertaining to the modellers' lack of negotiation with the clinician over important structural decisions, and the clinician's lack of understanding of the model and its development throughout, may have contributed to her inability to accept and perhaps also anticipate the model results. When the preliminary results were presented to the clinical experts in Case Study B, the clinicians questioned pathways included in the model, and it became apparent that there had been a lack of explicit agreement

between clinicians on an important structural assumption. Although the clinicians had discussed the validity of this assumption during structural development, there was outstanding disagreement between them, with the modellers eventually making the final decision on what was included in the model. This issue would have likely been avoided if the modellers had encouraged clinicians to come to an agreement on this assumption earlier on in model development.

Clinician understanding of model development was an overarching issue in the two case studies. Both sets of clinicians struggled to follow model development, and there was evidence in both case studies that a lack of understanding about what was important to include in the structure hindered discussions around pathway plotting. In Case Study A, there were examples of the clinician asking to include rare patient pathways, with the main modeller later implying that structural development would have been more straightforward if the clinician had been more aware of the information that was important to model. The Case Study B clinicians struggled to understand the inclusion of economic components in the structure, specifically the aspects that were considered by the modellers to have the potential to demonstrate the comparative cost-effectiveness of strategies. The main modeller from Case Study B later reflected that using conceptual modelling methods, and developing the structure in conjunction with clinicians, would have enhanced the clinicians' ability to understand the model. An additional problem with clinician understanding in both case studies concerned confusion among the clinicians over the research and economic terminology used during discussions.

Whilst clinician training formed the initial stage of model development in Case Study B, the modellers and clinicians in Case Study A were divided about its value. The clinician from Case Study A introduced the idea of clinician guidance without any prompt from the author, stating that it would have been beneficial to have had some background information on how

model development worked, so that she was aware of what she was expected to contribute. The main modeller had a similar view that it would be useful to provide clinicians with an understanding of what information was important to communicate during pathway plotting. However, the senior modeller argued that it was not the modellers' responsibility to provide training for the clinician, suggesting that this was unnecessary and too resource intensive.

7.4 Synthesis of findings from the different aspects of the research

A number of similar overarching themes were raised across the two phases of the qualitative research. These were issues around structural development, clinician involvement and the lack of time and resources available to model.

The findings collectively demonstrated that modellers followed similar stages in their modelling processes, but used different methods for structural development. Informants discussed, and/or were observed, using a variety of sources to inform structure, including clinician opinion, other model structures, NICE guidelines, and clinical data. This called into question whether there needs to be an agreed hierarchy of sources for informing model structure. Within the case studies, modellers used diverse methods to determine structural boundaries, with the Case Study A modellers applying restrictions iteratively during pathway plotting, and those involved in Case Study B agreeing on boundaries prior to model development. Arguably, the latter practice appeared more robust, as inconsistent criteria appeared to be used in Case Study A to exclude patient groups from the model. Modellers across both phases of the research drafted their structures differently. Interview informants were split as to whether they produced an initial draft of the structure in conjunction with clinical experts, or they first developed a structure alone, to later receive clinician validation. Both case studies used the latter method, although they undertook different approaches to discussing the initial draft with clinicians. Despite the different approaches taken, most

modellers engaged in a number of stages of planning before software programming, suggesting that model conceptualisation prior to implementation was important.

A final structural issue identified in both phases of the research was the generalisability of model structures. Many interview informants commented that involving too few clinicians in pathway plotting would bias the structure, but typically included two or even fewer clinicians in their processes, and did not discuss recruiting clinicians from diverse locations. The involvement of only one clinician in immediate structural development within Case Study A was problematic, as it was unclear how typical the practices of this clinician were in relation to other clinical centres. In Case Study B, there were numerous differences in the treatment pathways followed by the clinicians involved, which led to discussions among the clinical experts to determine the most representative practices to include in the model. This raised questions around whether good practice should constitute the involvement of a minimum number of clinicians, purposively selected to cover a variety of contexts.

Clinician involvement was a major theme of the research, with the importance of the clinicians' role in pathway plotting being highlighted initially in the qualitative interviews, and explored further within the case studies. However, clinician involvement was presented and observed as problematic within both phases, particularly in terms of clinician engagement in the process, and communication between modellers and clinical experts. Interview informants referred to experiences of working with clinicians who had a negative attitude towards cost-effectiveness modelling, and those who struggled to understand the importance of communicating the pathway of an average patient during structural development. The case study research identified similar issues, with clinicians in both case studies having difficulty understanding and accepting the developed models. However, observations made within the case studies suggested that more could have been done to aid clinician understanding. In Case Study A, the clinician was given very little direction from the modellers regarding what

information was important to communicate and include in structural development, and in Case Study B, clinicians were not given an overview of how the structure was developed, and why particular elements had been included. This raised questions as to whether modellers should be approaching structural development differently, perhaps involving clinicians in the initial conceptualisation of a model, and offering clinicians training as a standard element of the modelling process.

Observation and insight into communication between modellers and clinicians indicated that modellers held different views on the role that clinicians should have in model development. The majority of qualitative interview informants gave the impression that clinician input was important to structural development, but that it was the modeller who had the final say over what was included in the structure. The role of the clinicians differed in the two case studies, with a greater role for the group in Case Study B. The modellers limited negotiation with the clinician in Case Study A appeared problematic, which raised questions around how communication and decision-making should work between modellers and clinicians in model development.

The issue of the time and resources available to undertake modelling was emphasised in both phases of the research. During interviews, modellers suggested that they did not have the capacity to carry out external validation and structural checking, or systematic searches for economic data. The case study research similarly found that data inputs were not from systematic searches, although this was mostly as a result of a lack of available data, particularly on quality of life. On this note, further consideration may be required as to the impact of structural assumptions made in both processes due to the lack of available utility data. As with the interviews, model checking in the case studies focused on internal validation, as both sets of modellers carried out limited structural checking, and external validation was not undertaken in Case Study A. These findings raise questions around

whether modellers are prioritising the most important aspects of model development in the face of limited resources. In Case Study B, clinicians questioned the accuracy of NHS reference costs, suggesting that it may be important to consider whether existing and commonly used data sources are sufficient for model development.

7.5 Findings compared to existing literature

The systematic review undertaken in chapter 3 demonstrated that there was a limited number of papers that had used qualitative research to understand methodological issues within decision-analytic modelling. Of these, only two studies focused on exploring the entire model-building process (Chilcott et al., 2010; Squires, 2014).

The findings generated within this thesis both contribute and differ to those reported in other qualitative studies. Similar to this research, Chilcott et al. (2010) and Squires (2014) found that modellers were using different methods to one another in the development of their model structures. Chilcott et al., (2010) noted that modelling practice varied in terms of how far modellers planned and conceptualised a structure prior to software implementation, and Squires (2014) observed that modellers had different approaches to deciding which factors should be represented within the structure of a public health model. The findings of these studies, and of this thesis, may also potentially be related to the outcome of the systematic review presented in chapter 2, which found that there was little detailed guidance available for modellers on how to plan and develop model structures.

A major concern of a number of the qualitative papers therefore was the lack of a standard and explicit model conceptualisation stage within modelling processes (Chilcott et al., 2010; Kaltenthaler et al., 2011; Squires, 2014). This thesis has similarly highlighted a number of arguments in favour of using conceptual modelling methods. The thesis finds reasons in support of those given for conceptual modelling in the other qualitative studies, specifically

that the use of conceptual modelling methods can enhance the validity and credibility of a model, and improve communication and understanding of model development between the parties involved (Chilcott et al., 2010; Kaltenthaler et al., 2011; Squires, 2014). However, as a result of the exploration of communication between modellers and clinicians, the findings of this research additionally suggest that conceptual modelling techniques are important in the engagement of clinicians in model development. The research found that clinicians lacked understanding and acceptance of models and model results, perhaps in part due to their lack of involvement in model conceptualisation and decisions around model structure. Further discussion and recommendations around the use of conceptual modelling methods are given in section 7.6.1.

An additional finding of this research present in other qualitative papers was the acknowledgement of time and resource constraints on model development. A number of qualitative studies also reported that a lack of time and resources were limiting the practices of modellers, as Kaltenthaler et al. (2013) and Kaltenthaler et al. (2014) suggested that modellers were using 'rapid review methods' rather than systematic searches to identify evidence for model parameters. Squires (2014) found that the complexity of a structure was dependent on how much time a modeller had, and Chilcott et al. (2010) reported that modellers were not carrying out more complicated model checking activities, such as reprogramming and structural sensitivity analysis. Most of these papers gave the impression that time and resource restrictions, and their effect on the practices undertaken by modellers, were an inevitable aspect of model development. However, similar to the conclusions drawn from this research, Chilcott et al. (2010) questioned whether modellers were prioritising the most appropriate modelling activities in the time available. This thesis has provided insight into why modellers are using less robust practices, with modellers either perceiving particular tasks to be too complex, or being unable to find appropriate evidence for the model.

Discussion and recommendations around the lack of time and resources available in model development are given in section 7.6.3.

This research has made an original contribution in relation to its exploration of the involvement of clinicians in model development. Other qualitative studies commented relatively little on the role of clinical experts, and none involved clinicians directly in their research. Additionally, the systematic review of modelling guidance found that there was very little guidance available on the role of clinicians in the modelling process, particularly on how modellers should communicate with clinicians, and how many clinicians should be involved in model development. Most qualitative papers that mentioned clinicians, referred to their input very briefly, in stating that their input was important to structural development and model checking, and also that it was beneficial to have their contribution to the modelling process early on (Kaltenthaler et al., 2014, 2013, 2011). Chilcott et al. (2010) and Squires (2014) expanded on this, emphasising the advantages of using conceptual modelling methods as a means of facilitating communication with clinicians in structural development. However, the observation of communication between modellers and clinicians in this research has permitted new insights into what worked well, and arguably not as well, in terms of clinician input to structural development. Further, the observations and interviews with clinicians and modellers identified problems related to clinician understanding of model building, and generated suggestions around the potential benefits of including clinician training as a standard element of the modelling process. Recommendations on communication and training for clinicians are given in section 7.6.2.2.

Another important finding of this research was that modellers may be involving too few clinicians in model development. Again, few of the other qualitative papers discussed the appropriate number of clinical experts to include in the modelling process, and of those that did, recommendations seemed vague or impractical. Chilcott et al. (2010) noted that several

of the modellers interviewed demonstrated a preference for involving multiple clinical experts, and Squires (2014) suggested that involving as many stakeholders as possible (including clinicians) in the development of public health models was optimal. The latter recommendation raises questions around how feasible it is to attempt to recruit ‘as many experts as possible’ within a time and resource constrained environment. Discussion and recommendations around numbers of clinicians are given in section 7.6.2.1.

7.6 Implications of the findings for model building practice

The objective of the research was to make recommendations for improvements to model development, and highlight where future research was required to encourage good practice in modelling processes. The aim was not to provide prescriptive or all-encompassing recommendations for improvements to current modelling methods, but to consider whether conclusions can be drawn from the research that can assist modellers in their practices.

Although this research has involved both a broad assessment of perceptions, and a detailed in-depth analysis of the processes used to develop decision-analytic models, the findings will not be representative of the practices of all modelling teams, and it is likely that further primary research would highlight alternative perspectives on optimum methods and approaches to model development. However, inferences can be drawn about which practices appeared better within the processes that were observed, and according to the informants’ opinions. The next sections consider the implications of the findings for structural development, for clinician involvement, for the lack of time and resources available in model development, and for future modelling guidance.

7.6.1 Implications for structural development

Both phases of the research have demonstrated that modellers and modelling teams are using different methods and approaches to developing their model structures. Recommendations

made as a result of the research concern the use of conceptual modelling methods, and offers a suggested hierarchy of sources to inform model structure.

7.6.1.1 Model conceptualisation: planning the structure and determining model boundaries

The differences in the structural development methods used by informants, in addition to their reflections on the different approaches to drafting the structure, have given some indication of what might be considered good practice in this aspect of model building.

First, the majority of informants engaged in some form of structural planning, producing model specification documents and/or early diagrams of model structure, indicating that this was a standard and good practice element of model development. Informants reflected that these documents and diagrams were important in facilitating the discussion of a model between modellers and clinicians, to assist understanding of what a structure encompasses and to gain clinician agreement on a proposed structure.

Second, gaining input from clinicians early on in the modelling process was perceived as good practice, as the majority of informants involved clinicians from the outset of structural development. The few interview informants who did not reflected on this as being poor or unusual practice. However, modellers followed different practices in terms of whether they drafted an initial structure in conjunction with clinical experts, or developed a structural diagram first, to be validated by clinicians. To suggest which of these approaches constitutes better practice is difficult, as informants discussed the advantages of having an initial structural diagram, but the case studies suggested that clinicians struggled to understand and engage in model development because they had not been involved in initial model conceptualisation and structural decision-making. For the benefits associated with both approaches, modellers could, prior to their first meeting with clinicians, draft a broad outline of their understanding of what happens to a patient with a particular disease, and discuss each

aspect thoroughly with the clinical experts, to further develop the structure according to the clinicians' knowledge and experiences. This is largely similar to the practice undertaken in Case Study A.

Third, the modelling teams in the case studies took different approaches to determining structural boundaries. In Case Study A, applying boundaries to the model as structural development progressed highlighted inconsistency in practice, as the modellers applied contradictory criteria to exclude patients from the model. This inconsistency has potentially undermined the model's quality, as the modellers would have difficulty justifying their decision to exclude larger patient groups, but keep smaller patient numbers in the model. Arguably, it would have enhanced the credibility of the process if modellers had determined the proportion of patients that they considered to be exceptional in advance of model development. It is thus recommended that model boundaries are defined at the outset of the modelling process, as was observed in Case Study B, to avoid opportunities for, and accusations of, bias and lack of rigour.

Finally, the modellers in the qualitative interviews were found to be using and prioritising different information in the development of structure. This referred particularly to the divide between the academic informants who stated that they structured their models to reflect patients' experience of a disease, and non-academic consultancy informants who suggested that they developed a model to include the pathways and health states that were most important to a patient. Clearly, there is subjectivity in how modellers choose to build their models, and this is not necessarily poor practice; however, it can lead to modellers generating different cost-effectiveness results for similar decision problems. As a result, it seems important to recommend that modellers acknowledge the effect that decisions around structure can have on the outcomes of their models.

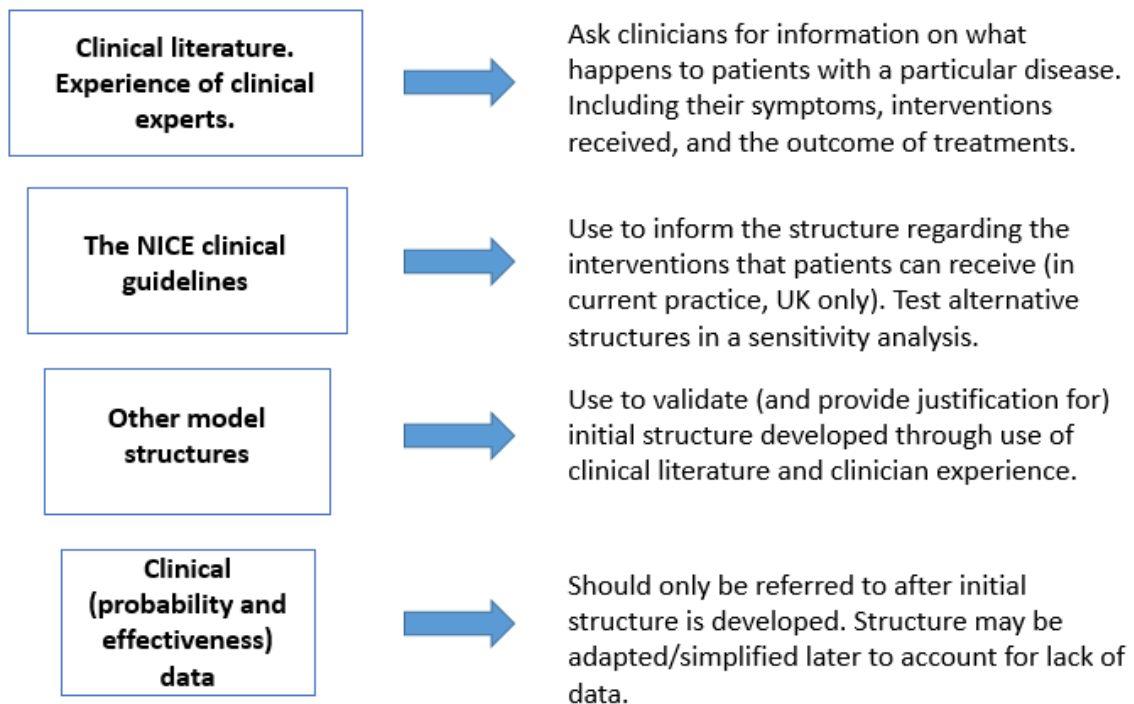
To incorporate these good practice recommendations, this research advocates the use of model conceptualisation methods. Conceptual modelling involves the separation of the problem-orientated model from the design-orientated model, as all factors potentially relevant to a decision problem are understood and represented in a conceptual diagram, prior to the development of the structure that will directly inform a cost-effectiveness recommendation (Kaltenthaler et al., 2011). The process of development from the problem-orientated model to the design-orientated model is documented, and the conceptualisation of the problem-orientated model, and decisions around the boundaries of a design-orientated model, are made in conjunction with clinical experts (Kaltenthaler et al., 2011). In line with the recommendations of this research, conceptual modelling methods thus facilitate the involvement of clinicians from the beginning of structural development, and require model boundaries to be determined explicitly, and at the outset of the development of the design-orientated model. Further, conceptual modelling necessitates that modellers are transparent about how they decide what to include in model structure, and the comparison of the problem-orientated and design-orientated model facilitates discussion around the impact of the exclusion of particular patient groups, pathways and health states from the final model (Kaltenthaler et al., 2011). Using conceptual methods, modellers can therefore continue to develop structures on the basis of the information that they perceive to be important, but should be open about how what has been included in and excluded from the structure, may influence model results. Modellers should read existing guidance on conceptual modelling methods (Kaltenthaler et al., 2011; Roberts et al., 2012; Squires, 2014).

7.6.1.2 Basis of structure

The majority of informants across both stages of the research suggested that the pathways or health states within a model should represent a patient's experience of a disease. Informants indicated that knowledge of a particular clinical condition to inform model structure was

generated through familiarisation with the clinical literature and discussions with clinicians. In terms of poor practice, many modellers were of the opinion that the evidence used to populate the model should not influence structural planning. Despite this, there was a fairly large group of Canadian modellers from the qualitative interviews, who stated that they developed their structures based on available data. With clear opinions from informants on the appropriateness of competing sources to inform model structure, but notable inconsistency in practice, this research has generated a suggested hierarchy of sources for structural development, similar to those that rank the validity of clinical evidence (Evans, 2003). A diagram of a suggested hierarchy of sources to inform model structure is given in Figure 7. At the top of this hierarchy, as observed in the practice of the majority of informants, is the use of clinical literature and clinician experience to inform model structure. Second in the hierarchy are the NICE guidelines, which modellers in the UK can use to gain knowledge on treatments that patients receive within current practice. To promote consistency, it is suggested that modellers avoid using clinician input and the NICE guidelines interchangeably, as was observed in Case Study A. It is instead recommended that modellers use the NICE guidelines in support of clinicians' versions of patient treatment pathways, and where the NICE guidelines suggest alternative pathways, test the impact of these in a structural sensitivity analysis. This approach should generate a more representative model, and valid cost-effectiveness recommendation, as the primary model is developed, and results calculated, on the basis of what is happening in actual clinical practice.

Figure 7: A suggested hierarchy of sources to inform model structure



Informants in both stages of the research discussed using existing models to inform their own model structures; however, informants were split as to how much influence existing structures had. The majority of interview informants used other structures as an inspiration rather than a basis for their models, but the modellers in Case Study B, for example, drafted the first version of their model entirely on previous work. Given this difference in methods, and thus the lack of clarity regarding what constitutes best practice, it is recommended that, rather than basing new models on existing structures, modellers use others' models to validate and justify decisions in structures that they have already drafted. If, as a result, modellers felt that they were missing important information from their model design, they could add to their structures accordingly. These methods will prevent new models being developed wholly on the basis of existing structures that may be outdated, be of poor quality and/or contain methodological errors.

Finally, the clinical data being used to populate the data parameters in a model should only be referred to after a structure has been developed. Developing a structure on the basis of available data was seen by informants as having the potential to bias a model result, as modellers could become aware of how including certain elements in the structure would impact the model outcome. Basing a model on only the data available will also likely lead to important components of a decision problem being absent from a structure. The sentiment from the majority of informants was that a structure should be developed initially on the basis of clinician experience, and structural and/or data assumptions should be made later in the process if data are unavailable to populate existing parameters.

7.6.2 Implications for clinician involvement

A number of problems were highlighted with clinician involvement in the modelling process. These specifically referred to the generalisability of model structures, the number of clinicians involved in structural development, and clinicians' lack of acceptance and understanding of modelling, seemingly related to communication between modellers and clinicians. The implications of these issues for modelling practice are considered next.

7.6.2.1 Number of clinicians

The findings of this research suggested that some modellers were involving too few clinicians in their processes; however, there was no indication from the research as to an ideal number of clinical experts to include. There were clear advantages to involving a large and diverse group of clinicians in the modelling process, as demonstrated within Case Study B. This was as discussions between clinicians could be used to resolve uncertainty and increase the representativeness of structural pathways, in addition to the validity and generalisability of the model. However, there are practical implications associated with involving greater numbers of clinicians in model development, including difficulties with managing discussion between

large groups, and recruiting (additional) clinical experts to a project. Arguably, the former can be addressed through greater facilitation from modellers within clinical meetings; however, the latter needs further consideration, particularly given clinicians' busy workloads. In the face of difficulty in involving larger numbers of clinicians in model development, it may be more efficient to recommend that recruiters strive to include a purposive sample with maximum variation, as sometimes advocated within qualitative research, thinking about all of those that can give a knowledgeable but potentially different perspective on the same issue (Merkens, 2004).

7.6.2.2 Communication and training for clinicians

The direct observation of meetings and discussions within the case studies, and the in-depth interviews undertaken with modellers, provided detailed insight into communication between modellers and clinicians. In terms of what worked well, face-to-face meetings, and the use of visual methods, such as diagrams of model structure, were reported as useful for engaging clinicians and enhancing their understanding. Likewise, discussions between clinicians, and the ability to have their questions answered by modellers, helped clinical experts to overcome difficulties with their interpretations. However, the findings highlighted a number of problems with how communication was facilitated in model development, and the clinicians' lack of understanding and acceptance of modelling.

To address these problems, conceptual modelling methods are again recommended to facilitate communication and the engagement of clinicians in the modelling process. As both sets of clinicians in the case studies struggled to follow model development, it was suggested by the clinicians and modellers that greater clinician involvement in designing the model would have improved clinicians' understanding, and thus made structural development easier for both parties. Further, conceptual modelling facilitates the involvement of clinicians in determining the boundaries of structure, providing that clinicians' have input both to the

problem-orientated and design-orientated model. Modellers in both phases of the research involved clinicians in structural decision-making to different extents, but the observation of communication, and the interview with the clinician in Case Study A, suggested that this clinician struggled to engage with the model because she had been permitted limited negotiation over what was included and excluded from the structure. For reasons around clinician engagement, and given the emphasis in both phases of the research on the importance of clinical experts offering sign-off on the clinical validity of a model, it seems essential that clinicians are involved in all decisions around structure. To facilitate this, clinicians can be offered training in model building, to teach them about the information that is important to represent in a cost-effectiveness model, and to provide them with the capacity to discuss and negotiate with modellers over structural inputs.

7.6.2.2.1 Clinician training

The suggestion that clinician training could be used to increase clinician understanding and engagement in the modelling process was made within both phases of the research by both modellers and clinicians, and was also an established aspect of model development in Case Study B. The usefulness of clinician guidance was mostly mentioned in relation to problems around communication between modellers and clinicians, as modellers suggested that it would be advantageous for clinicians to know what information was important to contribute to discussions. The clinicians also gave the impression that they had difficulty following and engaging with model development due to a lack of understanding of how a model was put together. The potential benefits of offering clinicians training are that clinicians may be more trusting of, and engaged in, model development, and have a greater understanding about what they need to contribute, leading to model development being carried out more efficiently. The recommendations for guidance for clinicians that emerged from both phases of the research were:

- An outline of what health economics is, why cost-effectiveness models are needed, and how models work, for example, how they synthesise cost and effectiveness information
- An outline of the typical stages involved in model development, for example, through a diagram of the modelling process
- The information that is important to a model in terms of the cost-effectiveness result, for example, the structure needs to be based on the clinical pathways of the average patient, rather than very rare clinical pathways, but also needs to include potentially important economic events
- Why assumptions need to be made and what is their potential impact (as outlined in the training for clinicians in Case Study B)
- A ‘dictionary of terms’ that gives definitions of the economic and research terminology commonly used in model development.

However, there was some resistance from informants regarding the introduction of clinician training to the modelling process. While some questioned the practicality of offering guidance to all clinicians because of time and resource constraints, others queried the likelihood of clinicians committing to undertaking training. The research findings have demonstrated that the clinicians interviewed were keen to receive some form of training, but that this training needed to be flexible. There were suggestions from the clinician informants that initial training sessions could be available online for them to access at their leisure. It was also suggested that training be offered throughout the process, for example, by reminding clinicians of the important information to include in a structure prior to structural discussions, and by developing a document of economic and research terminology that clinicians can refer to within meetings. Further, it may be most effective to develop ‘universal’ training for clinicians that could potentially be used within any modelling process, to prevent modellers

from having to spend time preparing training material for every model's development. Future research should focus on developing training and guidance material that is accessible, and that will facilitate model building for both modellers and clinicians.

7.6.3 Implications of a lack of time and resources for model development

Modellers in both phases of the research discussed being unable to undertake particular aspects or stages of model development due to a lack of time and resources. The modellers typically prioritised structural development, internal validation and sensitivity analysis on their models, and were less likely to carry out systematic reviewing for economic data parameters, external validation, and structural checking of the model. Further, modellers within the qualitative interviews gave the impression that they had to weigh up the benefits of carrying out additional modelling activities against the costs associated, and how far undertaking the task was likely to improve the quality of a model. The tasks that modellers were least likely to carry out thus appeared to be those that they perceived as the most difficult and/or time consuming. The implication of these findings is that modellers would benefit from an increase in the time and resource available for model development, however, given that this is unlikely to be feasible, future research should focus on exploring whether modellers are prioritising the aspects of model building that are the most important to the validity and robustness of a model.

Modelling practices were also limited by the data available. In the case study research, modellers discussed attempting to carry out good practice methods, for example, systematic searches for clinical and economic data, but finding no suitable evidence. This particularly applied to long-term quality of life data, with both sets of modellers having to make structural assumptions to account for this, specifically that patients who were not cured of their conditions maintained their initial quality of life for the entire model. As an area for further research it would be interesting to look at how commonly this assumption is made in

modelling, and whether it is, in general, consistent with clinical reality. Further, it seems important to raise awareness within the health economics and wider academic community regarding the shortage of relevant evidence, particularly long-term quality of life data, and whether this can somehow be considered and addressed within future research studies.

Finally, a number of modellers questioned the validity of, and level of detail offered within, traditional data sources for decision-analytic models, specifically the use of NHS reference costs. An alternative would be for modellers to use 'bottom-up' or micro costing methods, which require primary data collection, and the identification of costs and resource use at the unit level i.e. for individual patients (Morris et al., 2007). However, the micro costing approach is time consuming and resource intensive, and thus the benefits of its use in model development will again need to be considered against the importance of carrying out other modelling activities.

7.6.4 Exploration and documentation of the practices of modellers

Alongside the objective to highlight issues and problems being encountered within model development, this research has sought to explore and provide accounts of the current processes and methods used by modellers. The systematic review in chapter 2 found that current guidance on model building was generally lacking in detail, and particular aspects of the process were missing from the advice given, including methods for translating clinical and economic information into a model structure, and how to involve clinical experts in model development. In response, this thesis has presented an in-depth insight into how modellers carry out these, and other aspects of model development, which in the face of limited methodological guidance, could provide a useful resource for modellers. Those looking for an accompaniment to model building could potentially use informants' accounts of their methods, particularly for advice on what each stage of the process typically encompasses, what modellers perceived as good and poor practice, and the problems that they may

encounter. The think-aloud and observations undertaken of modelling processes in the case studies have also allowed for the documentation of the more nuanced aspects of model development, such as the techniques modellers used to communicate with clinicians over model pathways, and how modellers might interpret the results of sensitivity analyses.

Rather than offering prescriptive guidance and/or a summary of recommendations, a useful output of qualitative research into model building may be ‘guidance’ in the form of the reporting and/or a synthesis of model development processes, with some analysis and reflection from the researcher as to what appears to be good practice, and what common problems are. The benefit of this is that the level of detail captured on the modelling methods discussed and/or observed can be maintained as a resource for the reader. This case-study style guidance is likely to be valuable to anyone who is new to modelling, and who requires a thorough overview and insight into how the stages of model development work. Equally, this format may be more appealing to those who are experienced in modelling, providing a means for these modellers to compare and assess their processes against those of others. This is given that many of the senior modellers interviewed in the first phase of the research suggested that they did not (need to) use modelling guidance. However, the practicality of publishing such an in-depth and detailed resource would require further consideration.

The idea that modellers can potentially learn from the documentation of others’ processes suggests that it would also be beneficial for modellers to provide a more in-depth write-up of their methods, particularly on the aspects that are missing or lacking in detail within current guidance. This would be a useful and straightforward way for modellers to share and reflect on their modelling processes, and facilitate discussions around good practice, and the constructive criticism of the methods that modellers are using. Again, the practicalities of this would need consideration, given the restricted word limit of journals and other publications;

however, the creation of a separate online forum for modellers to share their methods is a potential starting point.

7.7 Future research

The findings of the qualitative research undertaken within this thesis have given rise to recommendations for future work. These include how to involve greater numbers of clinicians in model development, developing clinician training and guidance material, ascertaining which aspects of model development should be prioritised, and the practicalities of how detailed guidance on modelling processes can be shared. These suggestions have largely been based on issues that informants have discussed as being important, or which were repeatedly observed to be a problem within the modellers' processes. However, an important requirement for the validity of qualitative research is to also pay attention to 'negative cases', or findings that differ from the main themes that have emerged (Mays and Pope, 2000). An example of this was the suggestion by a few informants within the qualitative interviews that involving patient representatives in model development, particularly in checking model structure, was potentially important. Although patient involvement was carried out very rarely within informants' processes, and discussed by only a few modellers, the suggestion was that gaining the input of patients may improve the validity of a model, and thus it is recommended that future research explores how patients can be incorporated into everyday modelling practice. This avenue for future work also seems key, given the increasing emphasis on patient and public involvement within health research and in the development of clinical guidelines (INVOLVE 2015; NICE 2015).

7.8 Reflections on the method

Having considered how the findings of this thesis both support and add to those of other studies, and the implications for modelling practice and future research, this thesis will next

reflect on the methods that have been used. This section will consider the appropriateness of the research design, including its strengths and limitations, and reflections on the way in which the researcher, and the approach taken, may have impacted on the research findings. A full description of the methods used within this research is available in chapter 4.

The first phase of the research was qualitative interviews with modellers from various backgrounds, whilst the second phase involved two case studies with teams of modellers and clinicians, using semi-structured interviews, non-participant observation and think-aloud as methods of data collection. A clear strength of this approach was the level and breadth of detail that the combination of these methods were able to capture, thus allowing a thorough understanding of how model development was being undertaken, and what the common problems were within the modelling process. These rich insights have enabled recommendations to be made for modelling practice on the basis of considerable evidence, and permitted the in-depth documentation of informants' modelling activities, which could provide a valuable resource for modellers. In addition, the detailed presentation of the data collected within the thesis, and the analysis and interpretations that the researcher made as a result, are important requirements for demonstrating rigour in qualitative research, in allowing others to see how important themes, and the conclusions of the research were generated (Pope and Mays, 1995).

An in-depth and exploratory approach to research was taken to enhance the validity of the research findings. The research was designed to facilitate a broad and deep understanding of model development and the issues associated, through information on, and exposure to, as many aspects of the modelling process as possible. Thus, the qualitative interviews undertaken in the first phase of the research were face-to-face and flexible, to allow rapport to be established between the researcher and modellers, and to encourage informants to speak at length about their modelling experiences and the issues that they felt to be important. In both

case studies, the aim was to gather data on all stages of model development, and thus all email correspondence, documents and meetings between modellers and clinicians were observed. The use of observation allowed the communication between modellers and clinicians to be studied, which provided an original insight into how their discussions contributed to model development, and highlighted a number of problems with the way in which communication was managed. In Case Study A, the novel method of think-aloud was additionally used to generate data on tasks that modellers worked on independently, specifically the interpretation of sensitivity analyses.

However, a limitation of the case study research was that the researcher was unable to observe all aspects of model development within the two processes, particularly stages that the modellers undertook individually. Although attempts were made by the researcher to be present for these modelling activities, this proved difficult, as individual tasks tended not to be pre-planned. This was particularly an issue within Case Study B, as the modellers were involved in multiple model-based projects, and thus were unable to specify in advance when they would be working on a particular project and activity. Due to time constraints, the modellers tended to carry out stages of the process in a fragmented and ad-hoc manner, and as a result, think-aloud was not used within Case Study B, as the method requires that an informant invests time in verbalising their thought processes whilst completing a task (van Someran et al., 1994). The impact of observing mostly planned modelling activities was that more in-depth data were collected for particular stages of the modelling process, specifically for structural development, which involved regular meetings between modellers and clinicians. However, the qualitative interviews undertaken with modellers in the case studies allowed for the detailed exploration of the stages that could not be observed, through in-depth discussion of the methods used.

The triangulation of different qualitative methods was a further strength of this research. This is in relation to the combination of methods used within the case studies, and across the two phases of the empirical work. The use of non-participant observation in addition to the qualitative interviews in the case studies allowed the researcher to question the informants on what was observed, to gain their reflections on the methods used, and an impression of what they considered to be good and poor practice. Additionally, it was possible to relate informants' comments during interviews to the observations of what had occurred within actual modelling practice, offering context to, and possible explanation for, informants' perspectives. For example, in Case Study A, the clinician stated that she was unhappy with the representativeness of the model, and on review of the discussions between the modellers and clinician, it became apparent that this disapproval was likely to be connected to her lack of involvement in structural decision-making. The additional use of observational methods therefore allowed inferences to be made about why issues were occurring, which facilitated considerations around how problems could be addressed.

In terms of the overall research, the combination of interview and observational based methods has enhanced the breadth and validity of findings, as the case studies were able to capture data on issues that informants may forget or decide not to discuss during interviews. Within the case study phase, more data were collected on problems associated with the modelling process, as the modellers interviewed within the first phase of the research tended only to discuss issues with other people's processes in any depth. The triangulation of qualitative methods has provided a comprehensive evidence base for the research findings and its implications, as the main themes and issues that emerged were corroborated across both phases of the empirical work (Mays and Pope, 2000).

A robust sampling strategy was employed for both phases of the research, as twenty-four modellers from different geographical and professional backgrounds were interviewed in

phase one, and the case studies were carried out within diverse contexts. The aim was to gain a breadth of perspective, but also select informants and case studies that were ‘information-rich’, and thus able to generate detailed and credible data on model development (Patton, 2002). A strength of this research was that it went outside of the UK, permitting an international perspective, and outside of academia, as around a third of interview informants were from policy or consultancy backgrounds, and one of the case studies was undertaken within a policy environment. This gives the research an advantage over the qualitative modelling studies reviewed in chapter 3, as these mostly concentrated on academic modelling processes, and did not involve modellers from other countries. Further, none of the other papers included clinical experts in their studies, which has enabled this research to make an original contribution in terms of its focus on, and recommendations for, clinician involvement in model development. Arguably, this research represents the most extensive qualitative analysis of the model-building process undertaken to date, as other studies used relatively small sample sizes, and/or did not specify a sampling strategy that was aimed at ensuring that findings were transferrable across modelling teams and processes. The broad sampling strategy adopted within this research has enhanced the generalisability and transferability of its conclusions, as the recommendations have been generated across a number of different contexts (Firestone, 1993; Krefting, 1991).

A possible criticism of the case study research, however, is that further insights and potentially different processes and issues may have been highlighted if additional case studies had been carried out. For example, both case studies followed the development of decision tree models, meaning that some of the explanation of the methods used in model development will be specific to this type of structure. Although it would have been advantageous to have had the time to undertake a further case study, the research was successful in achieving diversity in the case studies selected, which allowed inferences to be made around the benefits

and drawbacks of the different approaches used in their model development. Additionally, the modelling teams in the case studies were chosen because they were recommended by the interview informants as having a good reputation for model building. In terms of the generalisability of the research recommendations, it was anticipated that issues present in both of these case studies were also likely to be problematic for modellers and modelling teams who were less experienced at modelling.

A rigorous approach was taken to analysis in both stages of the research. All interview recordings were transcribed by the researcher verbatim, line-by-line coding of interview transcripts was carried out, and analysis was undertaken in conjunction with the research supervisors to enhance the thoroughness and reliability of the research findings (Krefting, 1991; Pope and Mays, 1995). Additionally, grounded theory procedures were drawn on, and methods of constant comparison used, for the systematic analysis of data, as all new data were compared to the properties of emerging categories to enhance understanding of research themes, and increase the validity and sensitivity of the meanings assigned to the data (Kolb, 2012; Strauss and Corbin, 1990). In the case studies, the framework method was used, with the initial stage of analysis requiring the researcher to become 'immersed' in the data to ensure understanding, and the coding and charting stages providing a detailed and systematic process for categorising and comparing case study findings (Gale et al., 2013).

A strength of the analytical approach undertaken throughout the research was the use of inductive reasoning, alongside an exploratory approach to data collection. The use of the constant comparison method for the analysis of the qualitative interviews, specifically the continual comparison of data, facilitated the emergence of important themes and issues, and allowed future sampling and data collection to be aimed at generating a deeper understanding of these (Strauss and Corbin, 1990). The use of the framework method for the case study analysis then enabled these themes to be explored further using additional exploratory

qualitative methods, whilst also permitting new issues to arise. The inductive approach thus ensured that rich data were available on the issues most important to the research, as discussed by, or observed in, the practices of informants. This enhanced the credibility of the research, as findings and recommendations were defined by and closely related to the data (Kolb, 2012).

Finally, reflexivity is an important consideration of qualitative research, in terms of assessing how the background and preconceptions of a qualitative researcher affect the process and conclusions of a piece of research (Malterud, 2001). Haynes (2012) suggests that a researcher should question and be aware of their motivations for undertaking particular research, and how any existing assumptions or connections to the topic of study might be impacting on empirical practice and findings. As a PhD student undertaking a doctoral project, the research had an academic objective, aimed at increasing my own knowledge of modelling practice and issues, and contributing to any existing literature around good practice in model building.

With no previous experience as a health economic modeller, only knowledge gained from a Health Economics and Health Policy MSc course, I feel that I had no preconceptions of how model development should be undertaken, only a general idea of the stages involved.

Prior to starting the research, I had some experience of interviewing gained through a dissertation undertaken on a previous Master's course. However, I feel that my interview technique improved significantly whilst carrying out the qualitative interviews, as I became more confident in forming probing questions and using non-verbal cues, to encourage the informants to speak in more depth about particular topics. Further, I felt that it generally became easier to establish rapport with informants after I had undertaken a couple of interviews, as by then I knew the topic guide well, and so could therefore concentrate on listening and developing future questions based on informants' responses (Rubin and Rubin, 2005). The implication of this is that as I became more practiced and confident in conducting

the interviews, and also in my modelling knowledge, it is likely that richer data were collected for informants. Therefore more detailed data may be available for modellers interviewed later in the research.

Finally, it is important to acknowledge that my presence during the observations for the case studies may have affected the practices that informants carried out, given that the informants were aware that they were being observed. Although informants may have actively engaged in more robust practices because of their involvement in the research, the case studies both still generated data on problems occurring within model development.

7.9 Conclusion

The aim of the research undertaken within this thesis was to explore decision-analytic modelling processes, and investigate both good practice, and common problems, in model development. This research has made a number of original and important contributions to the health economics modelling literature, both methodologically, and in terms of its recommendations for future model development and for further research.

To the author's knowledge, this research has been the most extensive qualitative exploration of the modelling process undertaken to date, making its findings generalisable and useful to modellers and model building across different professional and international contexts. The research has used a robust and novel combination of qualitative methods, and demonstrated the benefits associated with using a qualitative approach to understand issues in model development. This thesis therefore provides an established and rigorous methodology for those wanting to undertake further research to improve the methods used in modelling processes and/or to develop modelling guidance. The research has also expanded on the papers available on good practice in the modelling guidance literature, and previous

qualitative studies aimed at exploring and identifying issues with decision-analytic model development.

The research contained within this thesis has resulted in findings on, and implications for, structural development, clinician involvement, and addressing the lack of time and resources available for model building. Outputs include recommendations for the use of model conceptualisation methods, the development of a hierarchy of sources to inform model structure, and suggestions for future research around improving model data sources, and for exploring which modelling activities to prioritise in the face of limited resources. The involvement of clinical experts in this research has been particularly novel, and has generated a number of recommendations aimed at facilitating and optimising the input of clinicians to future model development. These include recommendations around the number of clinicians that should be involved in model building, and suggestions for improving communication between modellers and clinicians in structural development, specifically through the use of model conceptualisation, and the introduction of clinician training to modelling processes. The in-depth and detailed exploration of the processes used by modellers has additionally provided a resource to guide other modellers in model development, as rich data were generated on modelling methods, reflections on good practice, and problems encountered during model building.

The contribution of this research is important in the context of the reliance of organisations such as NICE and CADTH on the cost-effectiveness results generated by decision-analytic models. The outputs of economic models are used by these organisations to inform decisions around the allocation of finite healthcare resources, and therefore there is a requirement for these models to be valid, and of good quality, to produce optimum cost-effectiveness recommendations. Modellers may also require guidance in their everyday modelling practices, to ensure that all models potentially informing healthcare decision-making are

developed to a high standard. This thesis has made a substantial contribution to this endeavour by generating recommendations for model building practice, aimed at enhancing the quality and validity of future models and modelling processes.

APPENDICES

Appendix 1: Search strategy for the systematic review of modelling guidance

1.1 Searching of the Electronic Bibliographic Databases:

1.1.1 The Cochrane Database of Systematic Reviews:

This initial search was iterative in that where too few or too many reviews were recovered, the search terms and strategy was altered to find more relevant reviews.

1st August 2012

Search 1:

Searched: '*health economic model building*' (retrieved 2 articles, none of which relevant to the proposed research topic)

Search 2:

Searched: '*modelling guidance*' (retrieved 8 articles, none of which were relevant to the proposed research topic)

Search 3:

Searched: '*modelling*' (retrieved 15764 articles, the majority of which were likely to be irrelevant to the context of the proposed research topic)

Search 4:

Searched: '*modelling*' and then for articles under the subheading '*methods studies*' (retrieved 1443 articles, 1 of which were deemed relevant to the proposed research topic)

24th September 2015

Updated search: Searched: '*modelling*' and then for articles under the subheading '*methods studies*' (retrieved 1445 articles, 2 of which were deemed relevant to the proposed research topic)

1.1.2 Search of the electronic bibliographic databases for relevant articles:

The search strategy for each of the electronic bibliographic databases was iterative, as search terms were developed to retrieve an optimum number of relevant papers. Only the final (original and updated) searches for the MEDLINE® and HMIC databases is documented below, although the development of the final search terms followed a similar process to that of EMBASE.

1.1.2.1 Original searches: EMBASE (EMBASE Classic + EMBASE 1947 to 2012 July 31)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis
2. Model building *or* modelling process
3. Guidance *or* guidelines *or* best practice *or* methodology
4. 1 *and* 2 *and* 3.

The search returned 16 papers.

Search 2:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis
2. Model building *or* modelling process
3. Guidance *or* guidelines *or* best practice *or* methodology *or* methods *or* procedures *or* checklist *or* critical appraisal
4. 1 *and* 2 *and* 3

The search returned 38 papers.

Search 3:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis
2. Model building *or* modelling process *or* decision-analytic modelling *or* decision-analytical modelling
3. Guidance *or* guidelines *or* best practice *or* methodology *or* methods *or* procedures *or* checklist *or* critical appraisal
4. 1 *and* 2 *and* 3

The search returned 112 papers.

Original search of EMBASE returned total of 112 papers.

1.1.2.2 Updated search: EMBASE (EMBASE Classic + EMBASE 1947 to 2015 September 24)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis
2. Model building *or* modelling process *or* decision-analytic modelling *or* decision-analytical modelling
3. Guidance *or* guidelines *or* best practice *or* methodology *or* methods *or* procedures *or* checklist *or* critical appraisal
4. 1 *and* 2 *and* 3

The search returned 144 papers.

Updated search of EMBASE returned total of 144 papers.

1.1.2.3 Original searches: Ovid MEDLINE® (1946 to July Week 4)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis *or* Economics, Medical
2. Model building *or* modelling process *or* modelling *or* modelling *or* decision analysis *or* decision support techniques *or* model
3. Guidance *or* guidelines *or* practice guideline *or* best practice *or* methodology *or* methods *or* checklist *or* critical appraisal *or* modeling methodology
4. 1 *and* 2 *and* 3

The search returned 431 papers.

Search 2:

The search aimed to identify papers related to Chilcott et al (2010), which was identified in MEDLINE.

The search returned 3 papers.

Original search of Ovid MEDLINE® returned a total of 434 papers

1.1.2.4 Updated searches: Ovid MEDLINE® (1946 to July Week 4)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis *or* Economics, Medical
2. Model building *or* modelling process *or* modelling *or* modelling *or* decision analysis *or* decision support techniques *or* model
3. Guidance *or* guidelines *or* practice guideline *or* best practice *or* methodology *or* methods *or* checklist *or* critical appraisal *or* modeling methodology
4. 1 *and* 2 *and* 3

The search returned 664 papers.

Search 2:

The search aimed to identify papers related to Chilcott et al. (2010) using the Ovid Medline® ‘find citing articles feature’.

The search returned 3 papers.

Search 3:

The search aimed to identify papers related to Philips et al. (2006) using the using the Ovid Medline® ‘find citing articles feature’.

The search returned 39 papers.

Search 4:

The search aimed to identify papers related to Roberts et al. (2006) using the using the Ovid Medline® ‘find citing articles feature’.

The search returned 29 papers.

Updated search of Ovid MEDLINE® returned a total of 735 papers.

1.1.2.5 Original search of HMIC Health Management Information Consortium (1979 to July 2012)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis
2. Model building *or* modelling process *or* decision-analytic modelling *or* modelling *or*

modelling *or* decision analysis *or* model

3. Guidance *or* guidelines *or* best practice *or* methodology *or* methods *or* checklist *or* critical appraisal

4. 1 *and* 2 *and* 3

Original search of HMIC Health Management Information Consortium returned 148 papers.

1.1.2.6 Updated search of HMIC Health Management Information Consortium (1979 to September 2015)

Search 1:

1. Health economics *or* health technology assessment *or* cost-effectiveness analysis

2. Model building *or* modelling process *or* decision-analytic modelling *or* modelling *or* modelling *or* decision analysis *or* model

3. Guidance *or* guidelines *or* best practice *or* methodology *or* methods *or* checklist *or* critical appraisal

4. 1 *and* 2 *and* 3

Updated search of HMIC Health Management Information Consortium returned 265 papers.

Appendix 2: Data extraction form - systematic review of modelling guidance

Description:	
Record number	
Author	
Article title	
Citation	
Type of publication	
Source of finding e.g. electronic database	
Study characteristics:	
Aims / objectives	
Process / stages of process	
Stages referred to:	
Understanding the context of the model	
Structuring the model	
Data (populating the model)	
Model implementation	
Dealing with uncertainty / model checking activities	

Appendix 3: Research protocol for the systematic review of modelling guidance

No.	Criteria
1	<p>Review question:</p> <p>What guidance currently exists for the process of decision-analytic model building?</p>
2	<p>Inclusion criteria:</p> <p>Any paper or form of literature that is focused on offering some level of advice (i.e. guidance, guidelines, methods, critical appraisal, or checklist) on at least one aspect of the model-building process i.e. a stage in the development of a model.</p>
3	<p>Methodology:</p> <p>Identifying the research evidence</p> <ul style="list-style-type: none"> a) Electronic databases: Cochrane Library (Cochrane Database of Systematic Reviews (CDSR)), MEDLINE®, EMBASE and HMIC. b) Forward citation searching on key papers within electronic databases c) Reviewing reference lists of relevant papers for further relevant key papers (systematic) d) Searching of published textbooks for relevant information e) Searching the NICE and CADTH websites guidance.
4	<p>Potential search terms:</p> <p>Health economics, cost-effectiveness analysis, model building, modelling process, guidance, guidelines, best practice, critical appraisal, checklist.</p>

5	<p>Paper selection:</p> <p>Paper selection will be based on the inclusion criteria. The initial selection will be made based on the title and abstract. The full-text version will be obtained for papers that appear to meet the inclusion criteria or avoid the exclusion criteria. This also applies where a decision could not be made on the basis of the information available in the title and abstract.</p>
6	<p>Data extraction:</p> <p>Publication information, study characteristics and findings related to the research question will be recorded in a data extraction form.</p>
7	<p>Synthesis of results:</p> <p>Narrative synthesis will be used.</p>
8	<p>Exclusion criteria:</p> <p>Papers that simply compare different methods for model building are to be excluded from the review, as are those that comment only on the mathematical construction of decision-analytic models. For practical reasons, the exclusion criteria extends to articles published in a language other than English.</p>
	<p>Miscellaneous</p>
9	<p>Documenting the searches:</p> <p>All of the electronic database searches are documented in Appendix 1.</p>
10	<p>Protocol modifications:</p> <p>Any modifications to the protocol will be recorded in the report.</p>

Appendix 4: Search strategy for the systematic review of qualitative modelling studies

1.1 Searching of the Electronic Bibliographic Databases:

1.1.1 The Cochrane Database of Systematic Reviews (Search for title, abstract and keywords (searched July 2015)).

Search 1: Qualitative modelling – returned 11 papers, 0 relevant (*Cochrane Database of Systematic Reviews: Issue 7 of 12, July 2015*)

Search 2: Qualitative modelling – returned 72 studies, 0 relevant (*Cochrane Methodology Register: Issue 3 of 4, July 2012*)

Search 3: Qualitative model building – returned 0 studies (*Cochrane Database of Systematic Reviews: Issue 7 of 12, July 2015*)

Search 4: Qualitative model building – returned 3 studies, 0 relevant (*Cochrane Methodology Register: Issue 3 of 4, July 2012*)

1.1.2 Systematic search of relevant electronic bibliographic databases

EMBASE Classic and EMBASE 1974 to 2015 July, Ovid MEDLINE ® 1946 to July Week 4 2015 and HMIC Health Management Information Consortium 1979 to May 2015.

Pilot search 1: Qualitative AND Modelling

Returned 1971 results, search terms were considered to be too broad.

Pilot search 2: Qualitative research AND Health economic modelling

Returned 2 results, none of which were relevant, search terms considered too narrow.

Pilot search 3: Health economics or cost-effectiveness analysis or health technology assessment

AND

Qualitative research or qualitative analysis or qualitative or qualitative methods

AND

Model building or modelling process

The search returned 3 results, with 1 paper meeting the inclusion criteria at the title/abstract stage of the screening process. Search terms were broadened for this search to involve the context of the research question however; it was proposed that expanding the search terms might result in further relevant papers being identified.

Main search 1: Find citing articles search on Chilcott et al. (2010) [paper identified in previous electronic database search]

Returned 3 results, none of these papers were relevant to the research question.

Main search 2: Reference list searching of Chilcott et al. (2010).

Search returned 0 relevant papers.

Main search 3: Health economics **or** cost-effectiveness analysis **or** health technology assessment **or** cost-effectiveness modelling **or** cost-effective

AND

Qualitative research **or** qualitative analysis **or** qualitative methods **or** qualitative

AND

Model building **or** modelling process **or** modelling **or** modeling

Returned 105 papers, 4 of which were identified as relevant to the research question at the title/abstract stage of the screening process.

- Kaltenthaler et al. (2014) ‘Identification and Review of Cost-effectiveness Model Parameters: A Qualitative study’
- Chilcott et al. (2010) ‘Avoiding and identifying errors in health technology assessment models: A Qualitative study and Methodological review’
- Kaltenthaler (2011) ‘The UK NICE single technology appraisal process: A qualitative study based on manufacturer’s submissions’ (*Conference abstract*)
- Kaltenthaler et al. (2013) ‘Reviewing the Evidence to Inform the Population of Cost-effectiveness Models within Health Technology Assessments’

Main search 4: Find citing articles search on Kaltenthaler et al. (2014).

Search returned 0 papers.

Main search 5: Reference list searching of Kaltenthaler et al. (2014).

Search returned 3 relevant papers, however two of these were duplicates of those already retrieved through electronic database searching. The remaining paper was original and relevant to the research question.

- Kaltenthaler (2011) NICE DSU Technical Support Document 13: ‘Identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models.’

Main search 6: Web searches for other relevant papers by Chilcott et al. (as lead or co-author).

Search returned 0 additional and relevant papers.

Main search 7: Web search using Google Scholar for other relevant papers by Kaltenthaler (as lead or co-author).

Search returned 3 original and relevant papers.

1. Kaltenthaler et al. (2012) ‘A qualitative study of manufacturer’s submissions to the UK NICE single technology appraisal process’
2. Squires et al. (2013) ‘Health Economic Modelling: A Conceptual Modelling Framework for Public Health Economic Models’ (*conference paper*)
3. Chilcott et al. (2010b) – ‘Choice and judgement in developing models for health technology assessment; a qualitative study’. Discussion paper. (*Unpublished*).

Hazel Squires was contacted directly for source of conference paper, which returned Squires (2012) ‘A methodological framework for developing the structure of public health economic models’ (*PhD thesis*). Conference paper therefore considered superseded by thesis as on same piece of research. Kaltenthaler (2011) conference abstract also superseded by Kaltenthaler et al (2012) full-text paper. Chilcott et al (2010b) superseded by full text HTA report.

Main search 8: Reference list searching of Squires (2012).

The search returned 2 relevant papers, 1 of which was a duplicate. The second paper was a PhD thesis and the author was contacted for a copy.

Result: Paisley (2012)

Total of 7 papers retrieved after removing work that was superseded by full-text papers.

Appendix 5: Data extraction form- systematic review of qualitative modelling studies

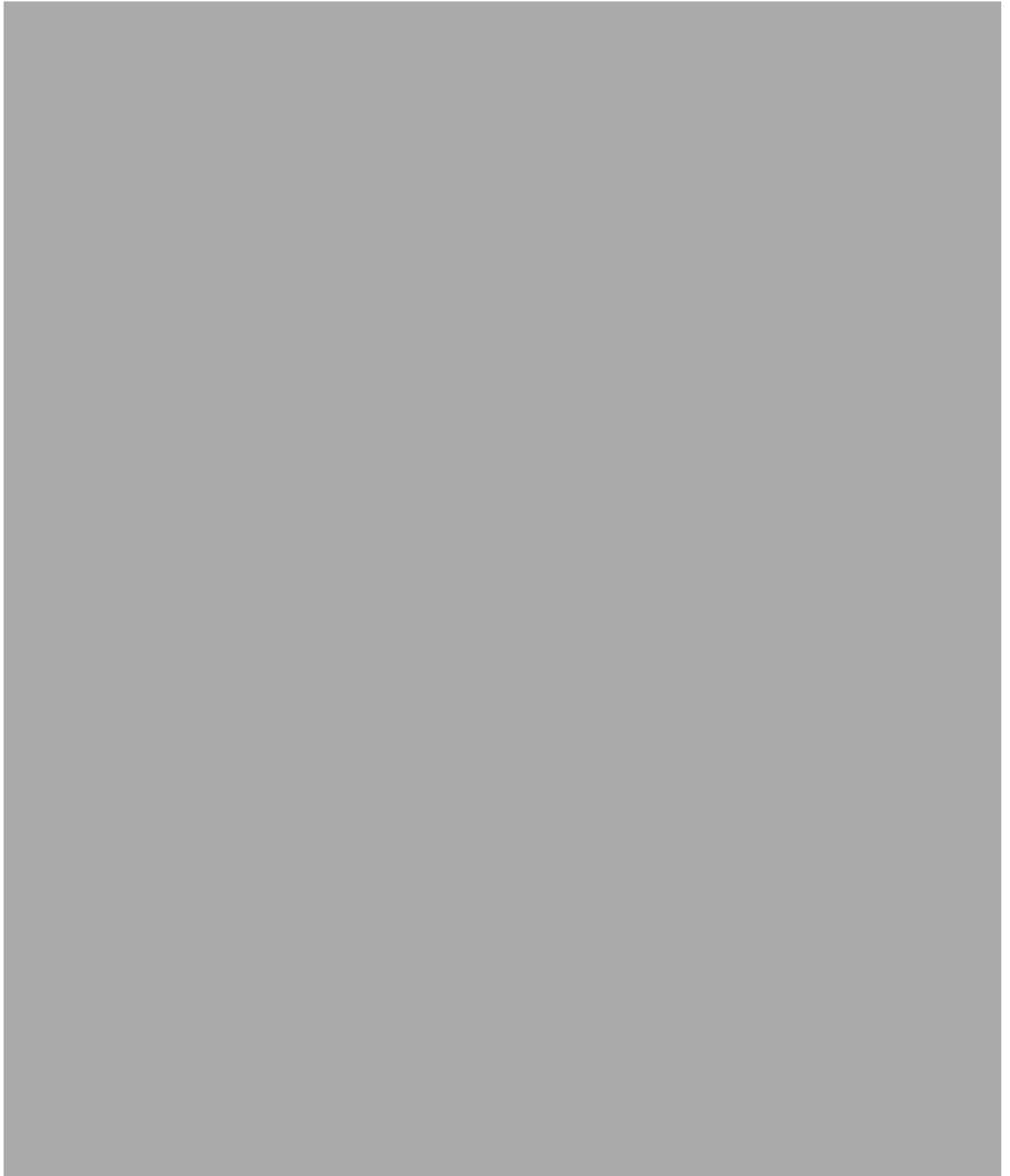
Description:	
Record number	
Author and year	
Article title	
Country/setting	
Citation	
Type of publication	
Source of finding e.g. electronic database	
Study characteristics:	
Aims of the paper	
Qualitative methods used	
Findings:	
<i>Current practice requires improvement</i>	
<i>Guidance and further research is needed</i>	
Summary of quality concerns	

Appendix 6: Research protocol for the systematic review of qualitative modelling studies

No.	Criteria
1	<p>Review question:</p> <p>How have qualitative methods been used to understand and/or improve health economic modelling?</p>
2	<p>Inclusion criteria:</p> <p>Any form of literature that has used qualitative methods to investigate the elements or processes involved in the development of health economic models.</p>
3	<p>Methodology:</p> <p>Identifying the research evidence:</p> <ul style="list-style-type: none"> • The searching of major electronic bibliographic databases for potentially relevant papers • Forward citation searching within electronic databases • The systematic scanning of the reference lists of key papers • Online searches using Google Scholar.
4	<p>Potential search terms:</p> <p>Health economics, cost-effectiveness modelling, health technology assessment, qualitative research, qualitative analysis, qualitative methods, modelling process, model building.</p>

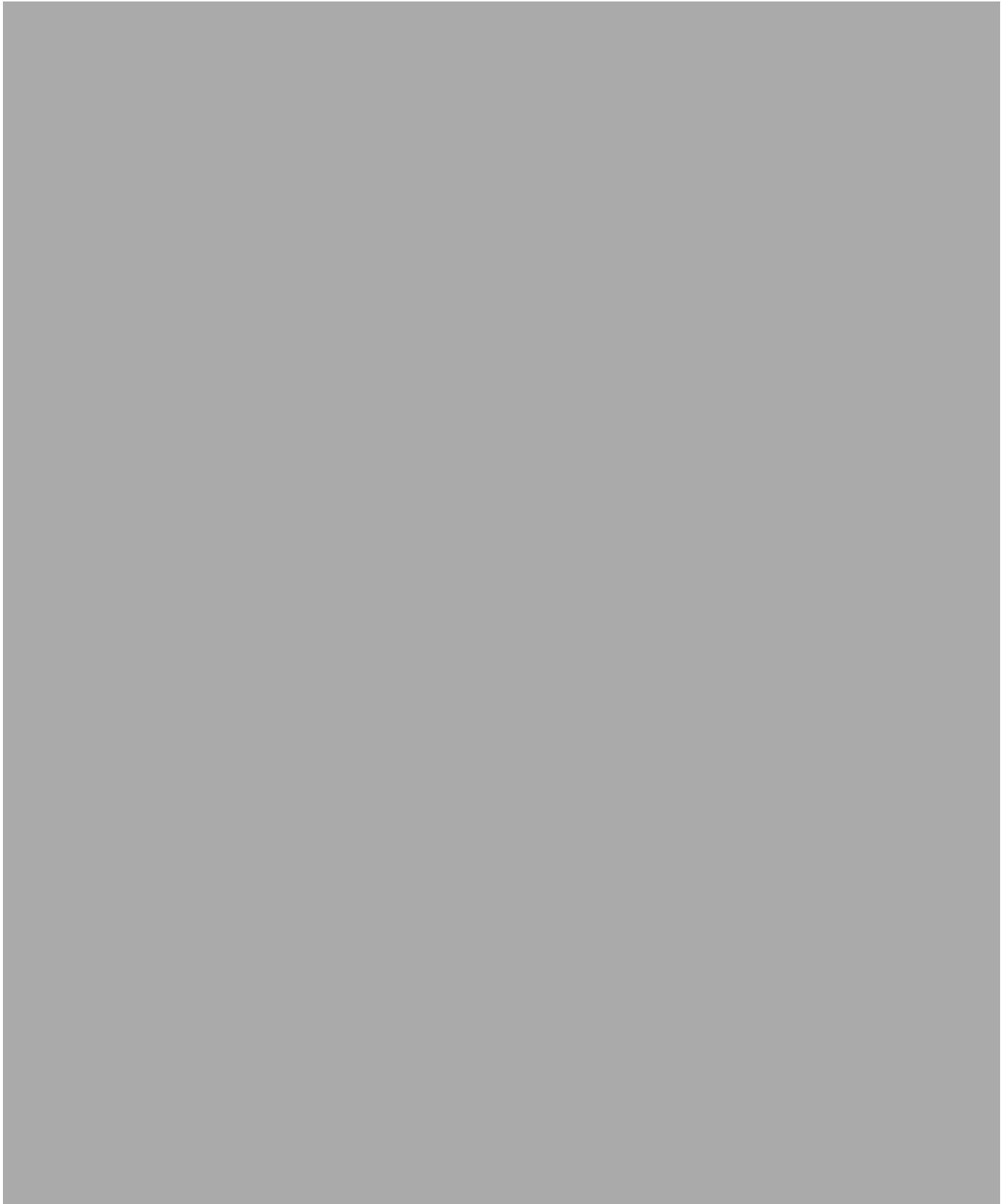
5	<p>Paper selection:</p> <p>Paper selection will be based on the inclusion criteria. The initial selection will be made based on the title and abstract. The full-text version will be obtained for papers that appear to meet the inclusion criteria or avoid the exclusion criteria. This also applies where a decision could not be made on the basis of the information available in the title and abstract.</p>
6	<p>Data extraction:</p> <p>Publication information, study characteristics and findings related to the research question will be recorded in a data extraction form.</p>
7	<p>Synthesis of results:</p> <p>Narrative synthesis will be used.</p>
8	<p>Exclusion criteria:</p> <p>Non-English language papers.</p>
	Miscellaneous
9	<p>Documenting the searches:</p> <p>The electronic database searches are documented in Appendix 4.</p>
10	<p>Protocol modifications:</p> <p>Any modifications to the protocol will be recorded in the report.</p>

Appendix 7: Ethical Approval for phase one of the research



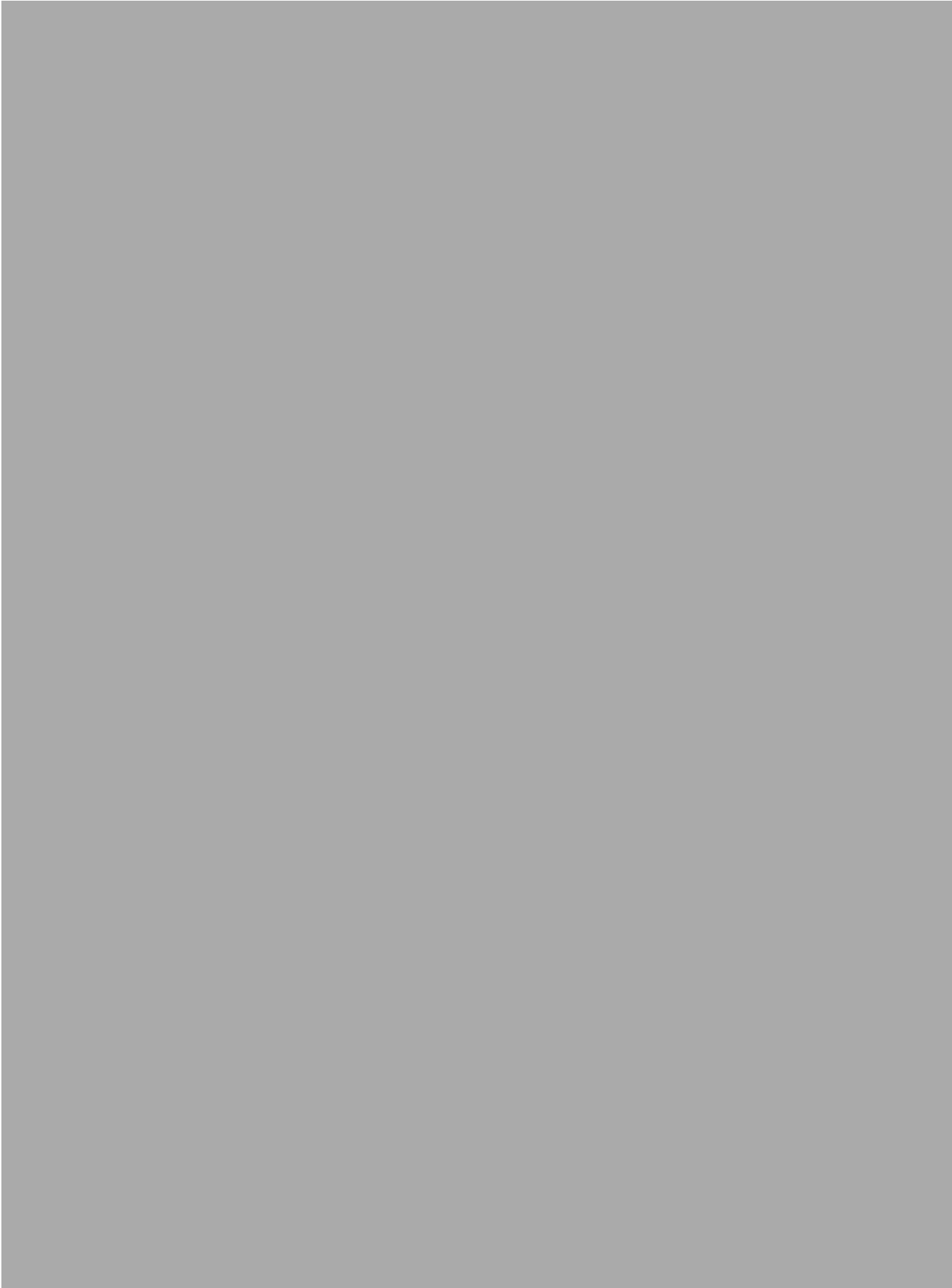


Appendix 8: Ethical Approval for phase two of the research





Appendix 9: Ethical approval for the Canadian phase of the research



Appendix 10: Sample invitation letter to modellers – phase one

UNIVERSITY OF
BIRMINGHAM

Health Economics Unit
University of Birmingham
Edgbaston
Birmingham
B15 2TT

<Address>

<Address>

<Address>

<Address>

Dear <Name>

I am a PhD student at the University of Birmingham conducting research on best practice in decision-analytic model building. As part of my research I am carrying out a number of short interviews with expert modellers to gain insight and opinion on the current process of model development.

I would be very interested in speaking to you given your experience of developing decision-analytic models. I am looking to gain a greater understanding of the methods used to develop the structure of a model and your opinion on which aspects of the process require improvement and further investigation. This is with an overall objective to develop best practice guidance.

If you agree to take part in the interview it will last approximately one hour and can be conducted at a time and in a location most convenient to you [most likely at your place of work]. I have attached an information sheet which will provide you with further information about the research.

I will follow up this letter with an email in 7 days' time to see if you would like to participate.

Thank you for your time.

Yours sincerely,

Sam Husbands
Doctoral Researcher
Health Economics Unit
University of Birmingham

Appendix 11: Informant information sheet – phase one



Best Practice in Decision-analytic Model Building Study - Information Sheet

Information for the participant

We would like to invite you to participate in our research study. This information sheet has been designed to help you to decide whether you would like to take part. It will explain the purpose of the research, what will be required of you as a participant, and how the data gathered from you will be managed. Please take a couple of minutes to read through the information below and ask any questions about the research that you may have.

Information about the research

This research study is being carried out and funded by the University of Birmingham. The overall objective of this research is to develop best practice guidance for building the structure of a decision-analytic model, in conjunction with the opinion and work of expert modellers and clinicians.

This study has been approved by the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham.

Why have I been invited to take part?

You have been invited to take part in this research due to your expertise in modelling and experience of developing decision-analytic models for the purpose of economic evaluation. Your views on the model building process will be central to enhancing understanding about where and how it can be improved, and will result in guidance on how to achieve an optimum structure and cost-effectiveness recommendation.

Do I have to take part?

No, you are under no obligation to participate. Your involvement is completely voluntary, but would be very beneficial to this research.

What is the research about?

With an aim to develop best practice for building a decision-analytic model structure, in this initial phase of the research we are asking you to take part in a face-to-face, in-depth interview to learn of your experiences of modelling. We are especially interested in your opinion on current practice in model building and your opinion on aspects of the process which require further investigation and development. This is with the intention of exploring and potentially improving these aspects through case study research with teams of modellers and clinicians.

How will the data from my interview be used?

It will be used to inform the remaining phases of the research as well as the overall objective of developing best practice for building a model structure. The second phase of the research involves case studies with a number of modelling teams and clinicians. The aim of this interview is to highlight issues relevant to the modelling process which can then be observed

and discussed further within the case studies, with the intention of developing suggestions for improving methods of model building.

The findings of this research will form an integral part of a PhD thesis, with key results being presented through seminars, academic journals and conferences.

What will happen if I take part?

Should you agree to take part, the interview will last up to one hour. During this time you will be asked to discuss a series of topics relevant to the model building process. The questions will ask you to draw upon your own modelling experiences and your knowledge of the field. Although the interview will follow a broad topical structure, please feel free to contribute any information which you believe to be relevant to the research objective.

Can I withdraw from the research?

Yes, at any time. You are not obliged to provide a reason for your decision to withdraw. Upon withdrawal you will have the option to request that any existing data you have given be destroyed. If you decide to withdraw we ask that this request is given within one week of the interview's completion, to avoid the data being analysed and used to inform future interviews and phases of the research.

Is the research confidential?

Yes, the research is confidential. Any data you give will be entirely anonymised and assigned a unique identity code to enable us to store your data safely and keep your personal information and digital recordings separate. All personal and interview information will be kept on a secured computer network which only the lead researcher will have access to. Data will be stored at the University of Birmingham for 10 years. Any writing or publications relevant to this research will not include any information that can identify you.

Where can I get more information about the study?

If you require any further information about the research please contact Sam Husbands on [REDACTED]

What if I have a complaint about the research?

If you want to raise any concerns about the research please contact Professor Joanna Coast at [REDACTED]

What now?

Please take your time to decide whether you would like to take part in this study and ensure that you ask any other questions that you may have.

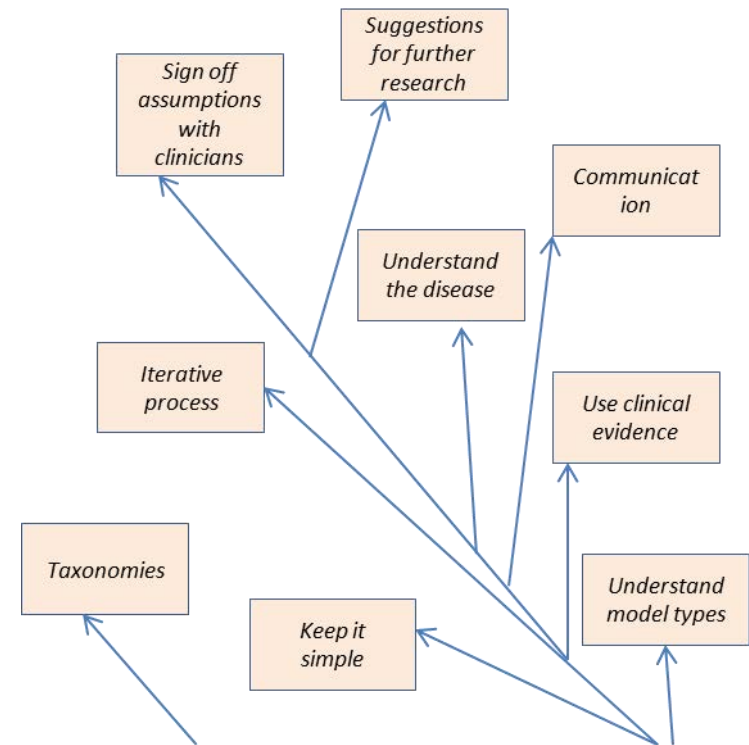
If you would like to take part in the research please sign the consent form.

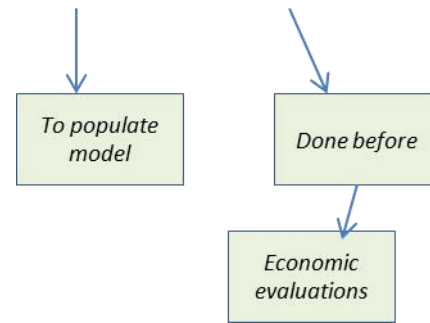
Thank you.

Appendix 12: Interview guide – phase one

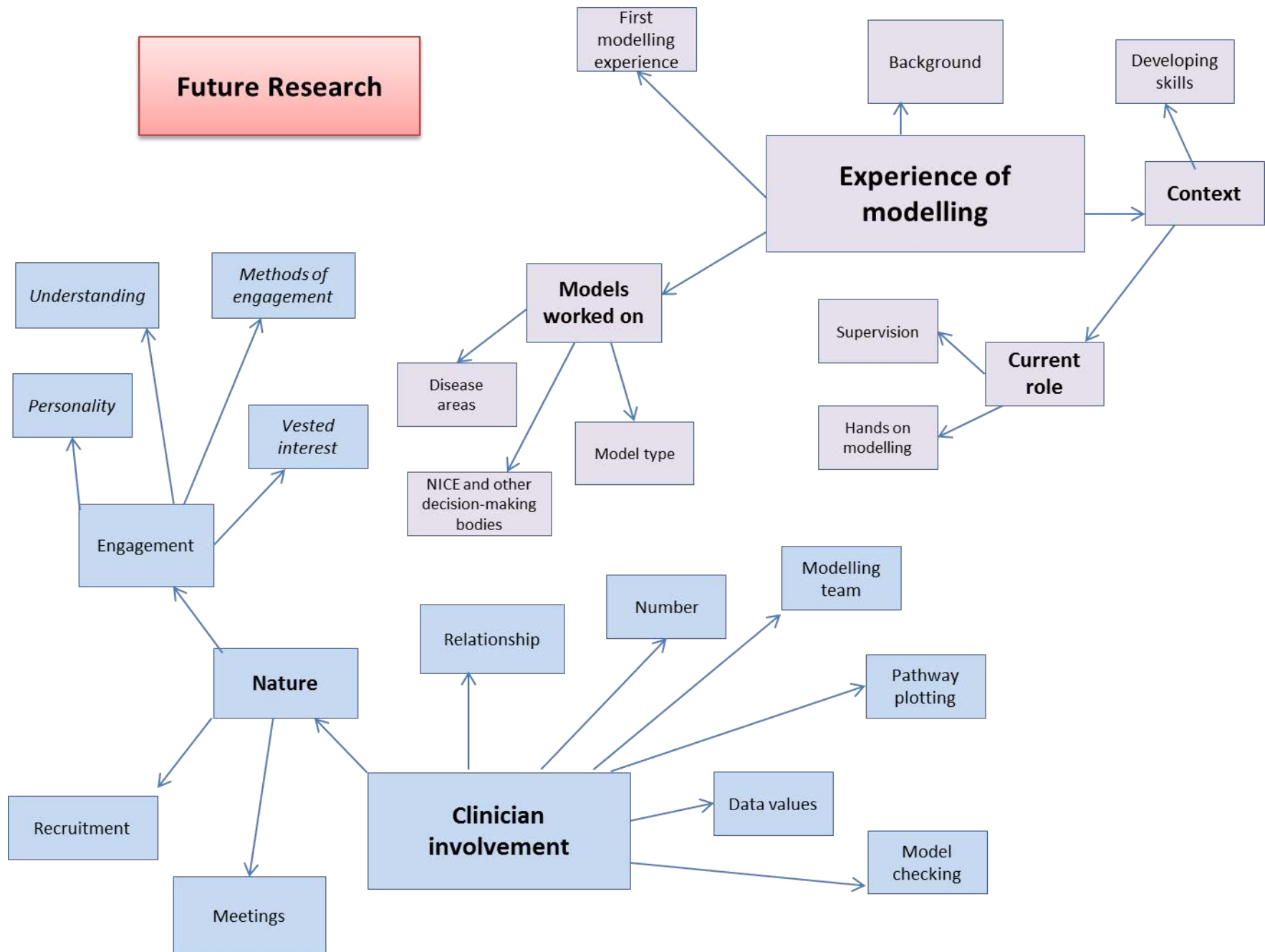
<p>CHECKLIST</p> <ul style="list-style-type: none"> • REPEAT PURPOSE OF THE RESEARCH • INTRODUCE TAPE RECORDER • CONFIDENTIALITY AND ANONYMITY • RIGHT TO WITHDRAW • TRANSCRIPTS • CONSENT FORM 	<p>Examples of modelling</p>	<p>-Can you think of an example of a model which you have worked on, where its development process was done particularly well?</p> <ul style="list-style-type: none"> • <i>Can you tell me about the process used?</i> • <i>Why was it good?</i> <p>-Can you think of any examples of modelling processes which contrast with this? [<i>i.e. examples of bad modelling processes</i>]</p> <p>-Have you found the process easier/harder depending on the particular disease area that you are working in?</p> <p>-Have you found the process easier/harder depending on the clinician that you speak to?</p> <ul style="list-style-type: none"> • <i>The number of clinicians that you speak to?</i> <p>-Do you think that there is an ideal way to model? [<i>If you had unlimited resources- time, money, people?</i>]</p>	
<p>Background: Modelling experience</p>	<p>-What is your experience of modelling?</p> <ul style="list-style-type: none"> • <i>How did you begin?</i> • <i>What is your current role? What does it involve?</i> • <i>Have you worked on many models?</i> • <i>What type of modelling have you done previously? Disease areas? Model types?</i> • <i>Have you ever done work for NICE? What?</i> 	<p>Model building guidance</p>	<p>-Do you or have you ever used modelling guidance to assist you in model building?</p> <ul style="list-style-type: none"> • <i>Which guidance?</i> • <i>How did you use it?</i> <p>-What do you think to the published modelling guidance as a whole?</p> <ul style="list-style-type: none"> • <i>What does it do? What does it not do?</i> • <i>In terms of structure?</i> • <i>Are there any that are particularly good? Particularly bad?</i> <p>-Do you think that it is possible to develop “a one size fits all” guidance for the process of developing the structure of a model? <i>Why?</i></p> <p>- Have you read the newest version of the ISPOR-SMDM Task Force modelling guidance? What do you think?</p> <ul style="list-style-type: none"> • <i>Do you believe that it is sufficient to assist modellers in developing a model structure?</i> • <i>Where is it sufficient/good? Where is it lacking?</i> <p>- What do you think can be done to assist modellers in the model building process?</p>
<p>Model building process</p>	<p>-Can you talk me through the process by which you usually develop the structure of a decision-analytic model?</p> <ul style="list-style-type: none"> • <i>Why do you include this particular stage? Are there any other stages which you could include?</i> <p>-Do you speak to clinicians?</p> <ul style="list-style-type: none"> • <i>How many? At what stage(s)? How do you ‘recruit’ them?</i> <p>-What model checking activities are carried out?</p> <ul style="list-style-type: none"> • <i>In terms of the model structure? Who is involved in this?</i> <p>-Have you worked in modelling teams where this process is done differently?</p>	<p>Future research</p>	<p>REITERATE THE OBJECTIVE OF THE RESEARCH PROJECT</p> <p>-Which aspects of the model building process do you think require further investigation [for the purpose of developing best practice guidance]?</p> <p>-Can you think of teams of modellers who would provide interesting case studies for the purpose of this research?</p>
<p>General modelling questions</p>	<p>-What would you define as the structure of a model?</p> <ul style="list-style-type: none"> • <i>What are the boundaries?</i> <p>-What would you define as a structural model error?</p> <p>-What would you define as a ‘good’ modelling outcome?</p>	<p>Any other points?</p>	<p>Is there anything else you would like to add which could contribute to this research?</p>

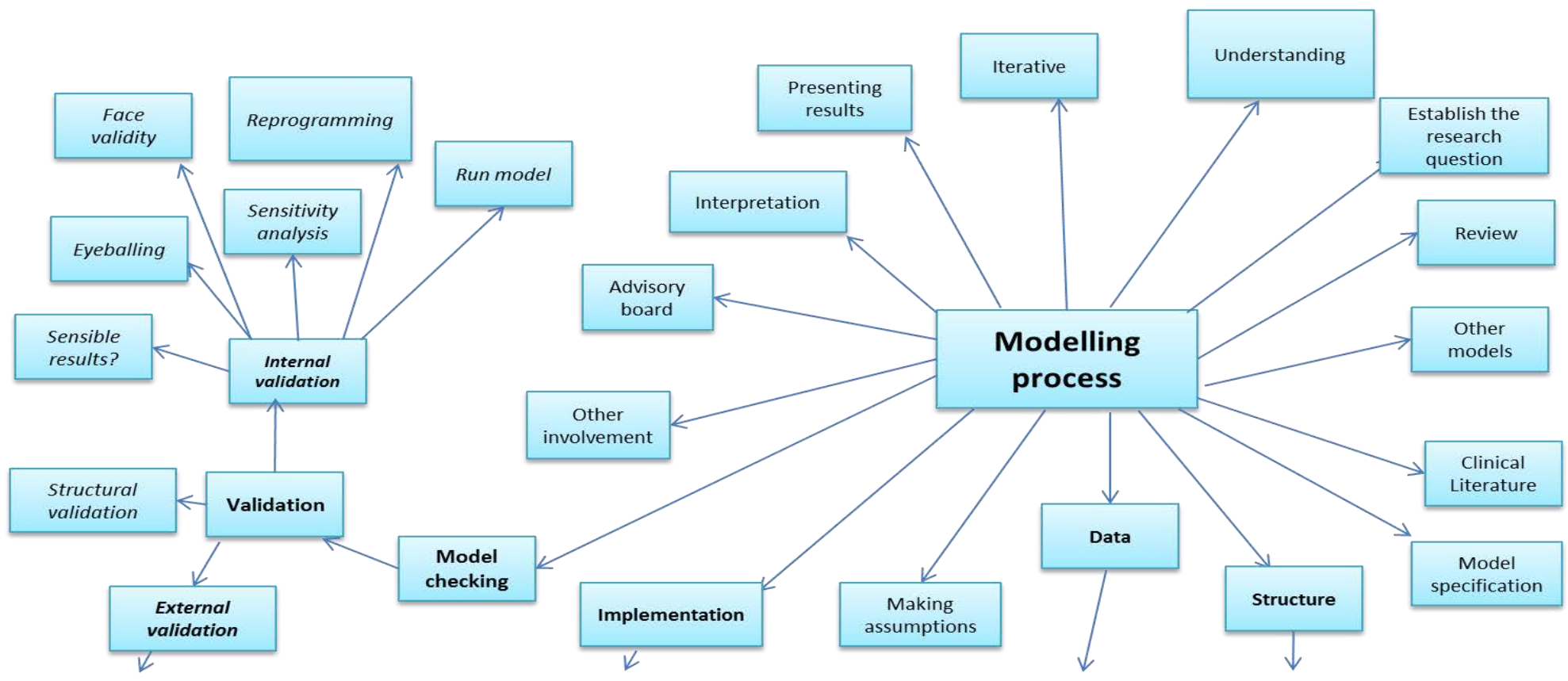
Appendix 13: Initial version of coding structure – phase one

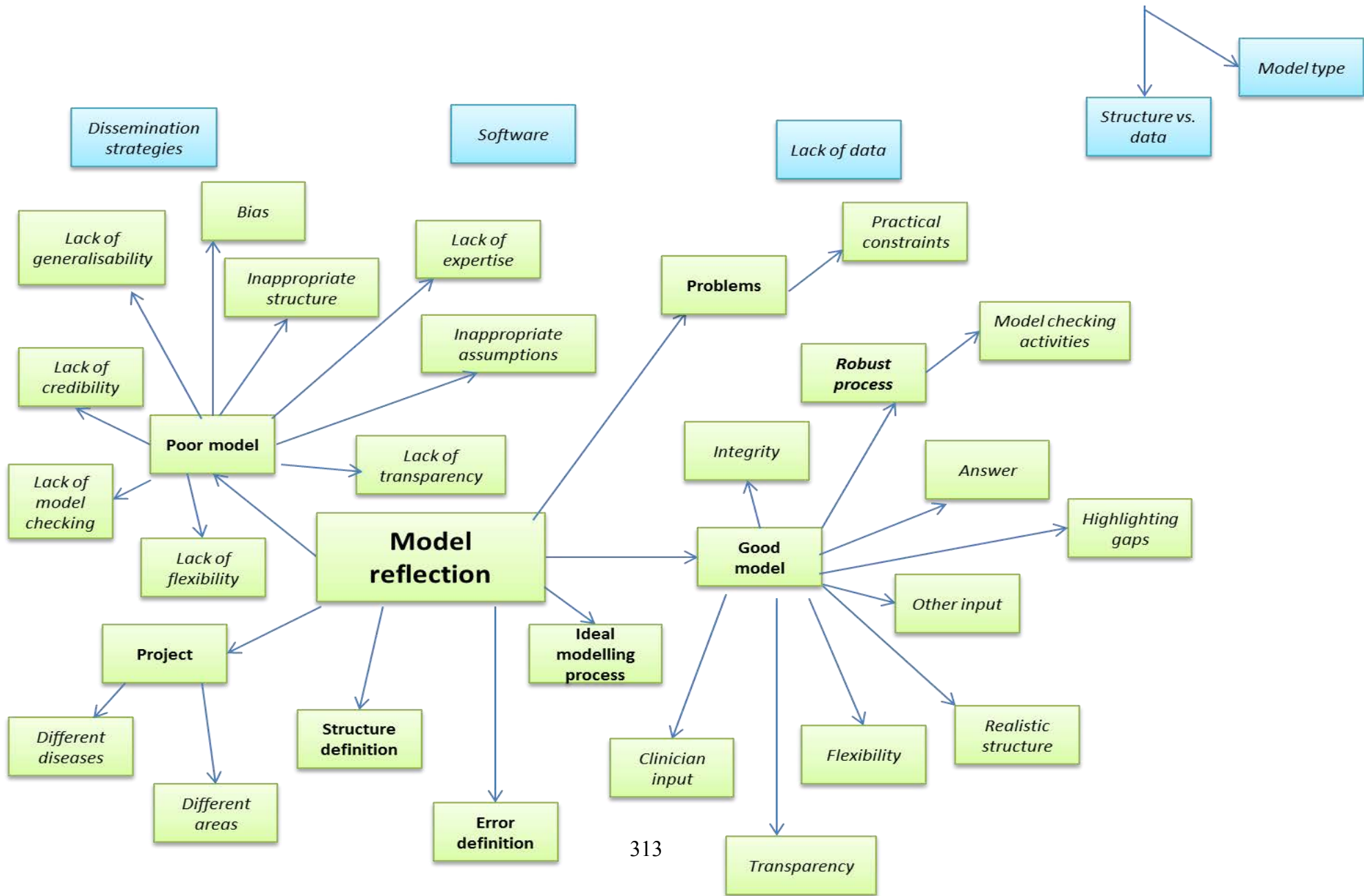




Appendix 14: Final version of coding structure – phase one







Appendix 15: Final version of coding structure definitions

1. Experience of modelling – In reference to the informant’s background in modelling and current skills and experience.

1.1 First modelling experience – In reference to the very first experience that the informant had of building a model

1.2 Context – Referring to the circumstances around the informant becoming a modeller and their past and current modelling projects

1.2.1 Background – Reference to the nature of an informant’s background, which may affect the way in which they model

1.2.2 Developing skills – Referring to how the informant built up modelling abilities

1.2.3 Current role – Informant is referring to the context of their current role and what it involves

1.2.3.1 Hands on Modelling – Where the informant refers to their role as one of physically building a model

1.2.3.2 Supervision – Where the informant refers to their role as a supervisor of the building of a model.

1.2.3.3 Seniority – How senior a modeller is or appears in relation to others working within their organisation or in the context of the modelling field

1.2.4 Models worked on/previous models – Reference to past models that the informants have worked on

1.2.4.1 – Disease area – Reference to the disease areas that the informants have built models in previously

1.2.4.2 – Model type – Reference to the types of models that the informants have built previously

1.3 Commissioning body – Informants discuss the requirements of different commissioning bodies for the models that are submitted to them

1.3.1 NICE - The informant's discussing any modelling work that they have specifically done for NICE

1.3.2 Other decision-making bodies - The informant's discussing any modelling work that they have specifically done for other decision-making bodies (outside of NICE), such as CADTH

1.4 Training and guidance – Where informants discuss any training they have given or guidance they have produced themselves specific to modelling

1.5 Model reviewing – Any reference to model reviewing work that the informants may have carried out

2. Modelling process – The informant makes reference to the stages that are involved in developing a model

2.1 Understanding – A stage of the modelling process of understanding the intervention/problem to be modelled

2.2 Establish the research question – A stage of the modelling process as determining the research question

2.3 Other models - Reference to the use of other people's models to inspire a new model being built

2.3.1 Structure – how informants use other models to influence the structure of their own models

2.3.2 Data – how informants use other models to influence the data used in their own models

2.3.3 Use – Informants' comments on how far they will use other people's models to inform their own

2.3.4 Identifying other models – Discussion of the methods used to identify existing models

2.4 Literature – Reference to using the clinical and/or economic literature to determine the structure of a model

2.5 Review - Reference to conducting some sort of review prior to the development of a model structure

2.6 Model specification/protocol - Reference to the development of a model specification document prior to or alongside the development of a model structure

2.7. Structure – Direct reference to establishing/building the structure of the model

2.7.1 Data vs structure – Discussion on how far available data should influence the development of a structure

2.7.2. Model type – Comments on the types of models used and how a particular model is selected to represent a problem

2.7.3. Simplicity vs complexity – Comments on how simple/complex a model structure needs to be to address a particular decision problem

2.8 Data – reference to the use of data in a model

2.8.1 Lack of data - Discussion of what happens when there is a lack of data available to populate a model

2.8.2 Quality of data – Discussion of what happens when the data available is of low quality

2.9 Making assumptions - Reference to the stage of creating model assumptions where data or information is not available in the literature

2.10 Implementation – Reference to building the model in a software platform

2.10.1 Software – Reference to the use of a particular software in the development of a model

2.11 Iterative - Reference to the modelling process as being one that is iterative (involves backward and forward movement between stages in the modelling process)

2.12 Advisory board – Informants’ reference to a stage in the process involving the undertaking of an advisory board

2.13 Model checking activities – Reference to activities that are carried out to ensure that a model is as it should be

2.13.1 Validation – Reference to the activities carried out under the umbrella term of validation – involving making sure that a model is a true representation of the disease/problem/intervention

2.13.1.1 Internal validation - Reference to the activities carried out under the umbrella term internal validation- ensuring that the internal workings of a model are performing as they should be

2.13.1.1.1 Run Model - Where informants discussing running a model to ensure it is working correctly and producing sensible and intuitive results

2.13.1.1.2 Eyeballing - Checking the mathematical aspects of the modelling by ensuring that the correct values have been entered into the software/spreadsheet

2.13.1.1.3 Face validity – Checking whether a model appears valid in the face of what is already known about the nature of a particular disease

2.13.1.1.4 Reprogramming (software) - Building the model in an alternative software program or in the same software again to check if same model result is achieved

2.13.1.1.5 Sensible results? - Checking the model results by ensuring that they stand up against existing similar models and also intuition

2.13.1.1.6 Log changes – where informants discuss formally noting all the changes that they make to the model as a means of checking that these have occurred

2.13.1.2 External validation – Reference to activities which involve checking a model and its finding against sources that are external to it

2.13.2.1 Dissemination strategies – Reference to a means of checking the model which involves presenting the findings to outside parties (aside from clinicians) for their feedback

2.13.1.3 Structural validation – Reference to checks that are carried out on the structure of a model

2.13.2 Sensitivity analysis - Reference to checking the model results through the use of sensitivity analysis i.e. changing a parameter(s) to study its impact on the results of a model

2.14 – Other interaction – Reference to getting other people outside of the modelling team involved in the modelling process

2.15 Interpretation – Reference to an explicit stage in the process in which an informant translates the model results into overall findings

2.16 Presenting results - In reference to the way in which the model results are handed over to the decision-makers, project commissioners or disseminated to others

3. Model reflection – In reference to an informant’s evaluative comments on aspects of the modelling process.

3.1 Project - Where informants' discuss the modelling process being different according to the particular project that they are working on

3.1.1 Different areas - Informants' suggesting that model structures may be developed differently depending on the area of the country in which the study is undertaken

3.1.2 Different diseases - Reference to the informants' belief that some diseases are more difficult to model than others and require different modelling techniques and pathways

3.1.3 Individual modellers - Individual modellers and their aptitude for modelling may affect the way in which they build models

3.2 Structure definition – reference to how the informants define the structure of a model

3.3 Error definition – reference to how the informants define error in the structure of the model

3.4 Good model [Process] – Reference to what the informant believes constitutes a good model and outcome

3.4.1 Clinical input - A good model is one that has had input from people with clinical expertise

3.4.2 Other input – A good model is one that has had input from people with relevant expertise, aside from clinicians

3.4.3 Model specification – A good modelling process is one that includes the development of a model specification document

3.4.4 Realistic – A good model is one that has a realistic structure in relation to the disease that it is representing

- 3.4.5 *Confirmation* - A good model (outcome) is one that has been validated by the relevant people or information
- 3.4.6 *Model checking activities* – A good model process is one that contains sufficient model checking activities
- 3.4.7 *Answer* – A good model is one that is able to provide an answer to the decision problem
- 3.4.8 *Highlighting gaps* – A good model is one that can be used to highlight gaps in the literature and/or existing knowledge
- 3.4.9 *Providing information* – A good model is one that can be used to generate information on a particular issue or problem
- 3.4.10 *Flexible* – A good model is one that is flexible – allowing a number of different decision problems to be represented by a model structure
- 3.4.11 – *Transparency* – A good model and model process is one that is transparent in terms of how it has been developed
- 3.4.12 *Integrity* - A good model is one that has been developed with integrity by the modeller i.e. built objectively and not with bias towards a particular outcome
- 3.4.13 *Robust* - A good model is one that has followed a rigorous process of development
- 3.4.14 *Confidence in* – A good model is one that the modeller has confidence in, in terms of its development and result
- 3.4.15 *Credible*- A good model is one that is credible in terms of how it has been built and what information has been used to develop it
- 3.4.16 *Good modeller* – A good model is one that has been developed by a ‘good’ modeller
- 3.4.17 *Difficult to know* – The informant(s) struggles to define what a good model is
- 3.4.18 - *'Ideal model [process]'* – Informants’ discussion of what they would define as being an ideal model or an ideal modelling process
- 3.5 *Poor model* - What the informant defines as a poor model [process]
- 3.5.1 *Wrong mode/inappropriate structure* - A model is poor due to an incorrect model type or structure being used
- 3.5.2 *Inappropriate assumptions* - A model is poor due to it having been based on unrealistic or inaccurate assumptions
- 3.5.3 *No input* - A poor model is one that has had no or little input from a wider modelling team/other experts

3.5.4 Lacks transparency - A poor model is one that is not transparent in terms of the methods that have been used to develop it

3.5.5 Inconsistency - Relating a poor model outcome to one that does not address the original decision problem

3.5.6 Missing stage – A poor model is one where essential stages of the modelling process have been missed

3.5.6.1 Lack of model checking activities – A model is of poor quality because it has not undergone enough model checking activities

3.5.7 Technical errors – The standard of a model is poor due to technical errors, in reference to the way that the model has been programmed into software

3.5.8 Lack of modelling expertise – The person building the model does not have the appropriate expertise, leading to a poor model (outcome)

3.5.9 Data inputs - The model is of a poor standard because of the data that has been used to develop and/or parameterize it

3.5.10 Lack of generalisability – A poor model is one that’s design means that its results are not generalizable

3.5.11 Lack of credibility – A poor model is one that lacks credibility due to either the methods or inputs that have been used to develop it

3.5.12 Lack of flexibility – A poor model is one that cannot be used outside the context of the decision problem that it was originally designed to address

3.5.13 Biased – A poor model is one that is biased due to the way in which it has been developed

3.6 Problems- Reference to any problems or issues that the informant(s) may have encountered in the model building process

3.6.1. Practical issues - Where informants have cited practical issues that have occurred whilst building a model

3.6.2 Miscommunication – Discussion of problems that have arisen in model development due to miscommunication between the parties involved

3.7 Improvements -Informants reflect on any improvements that they feel could be made to their own modelling process

3.8 Timeframe –Discussion of the timeframes that the informants follow when building a model

3.9 Client involvement – Discussion and reflection on the nature of client involvement in the model building process

3.10 Patient involvement – Discussion and reflection on the nature of patient involvement in the model building process

3.11 Model reviewing – Reflection on the process an informant has used or is using to review other models

4. Clinician involvement – Reference to developing a model in conjunction with clinical experts

4.1 Pathway plotting - Reference to the process of planning and developing the model pathways with clinicians

4.2 Data values – Reference to use of clinician opinion in identifying and obtaining data values to be used in a model

4.3 Checking – Reference to the clinicians’ involvement in checking aspects of the model

4.4 Model focus – Reference to the clinicians’ involvement in decided what the model should look at and question, what should be included in the model

4.5 Number – Reference to the number of clinicians that are involved in model development

4.6. Relationship – Reference to the nature of the professional relationship that the informants have with the clinicians who they are working with

4.7 Nature - Reference to the nature of a clinician’s involvement on a project

4.7.1 Recruitment – Discussion on how clinicians are recruited to modelling projects

4.7.2 Meetings – Reference to how the meetings between modellers and clinicians might take place

4.7.3 Engagement – Discussion on the ability of clinicians to engage in a model’s development process

4.7.3.1 Understanding – Reference to the fact that clinicians are able/not able to engage in a model development process because of their understanding/lack of understanding of health economic modelling

4.7.3.2 Personality – Reference to the fact that clinicians may engage more/not engage in model development due to their personal preferences and attitude towards collaboration and modelling

4.7.3.3 Vested interest – Reference to the fact that a clinician having a ‘vested interest’ in a particular model result may have an effect on how engaged they are in the model development process

4.7.3.4 Methods of engagement – Informants’ comments on methods that they use or can be used to engage clinicians in model development

4.8 Modelling team – Reference to the team who are involved in the model development process (outside of the modeller (informant) and the clinicians)

4.9 Agreement – Referring to agreement between modeller and clinician on certain aspects of the modelling process

4.10 Disagreement - Reference to issues or disagreements between the modeller (informant) and clinicians or between the clinicians themselves

5. Modelling guidance - Reference to types of modelling guidance and the informants’ use or opinion of it, or guidance given by the informant

5.1 Type of guidance – The informant's reference to the type of modelling guidance that is generally available (or that they are aware of as being available)

5.2 Use of guidance – In reference to the informant’s use of modelling guidance

5.3 Attitude towards guidance – In reference to the informant’s attitude towards particular modelling guidance or the concept of modelling guidance in general

5.3.1 One size- fits-all - Attitude of the informants towards the possibility of developing or having a ‘one-size-fits-all’ guidance

5.4 Informant’s guidance - Guidance given by the informant as to how a model should be developed (reference to the particular actions that a modeller should carry out i.e. their ‘top tips’)

6. Future research - Suggestions from the informants concerning which aspects of the modelling process should be investigated further for the purpose of developing modelling guidance.

6.1 Area of focus – Discussion of which aspects of model development particularly require further investigation and potentially the development of guidance

6.2 As a whole – Discussion of what the developed guidance should focus on as a whole

6.3 Format – Discussion on how the developed guidance should be organised and formatted

6.4 Case studies – Specific suggestions related to the subject of case studies for phase two

Appendix 16: Excerpt from matrices
comparing responses of interview
informants – phase one

MATRIX	MEDELLING PROCESS		Quality control under validation		Quality control under validation		Quality control under validation		Quality control under validation		Quality control under validation	
	DATA	MODEL CHECKING	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity	Formal validity
I11	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I12	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I13	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I14	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I15	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I16	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I17	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I18	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I19	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I20	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I21	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I22	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I23	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
I24	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Appendix 17: Consent form for phase one of the research

Best Practice in Decision-analytic Model Building Study – Consent Form

Participant Identification Code:

Name of Researcher: Samantha Husbands

To take part in the study you must agree to the following:

Please Initial Boxes and
Sign Below to Consent

1. I confirm that I have read and understood the information sheet for the research entitled ‘best practice in decision-analytic model building study’. I confirm that I have had the opportunity to consider the information provided, ask any questions and have those questions answered.
2. I understand that my participation in the study is entirely voluntary and that I am free to withdraw from the study at any time without giving any reason or explanation. I understand that any personal information collected within this project will be anonymised and will remain strictly confidential.
3. I agree to participate in the above study and consent to the interview being audio recorded. I understand that the data may be looked at by the Research Team at the University of Birmingham and anonymised quotes from this interview may be included in research reports.

I hereby consent to participate in this research project and confirm that I understand and agree with the statements listed above.

Signature of Participant:

Date:

Signature of Interviewer:

Date:

Should you have any further queries or concerns please contact:

Sam Husbands
Health Economics Unit
Public Health Building
University of Birmingham
Birmingham
B15 2TT



Appendix 18: Informant information sheet - phase two



Best Practice in Decision-analytic Model-Building Case Study - Information Sheet

Information for the participant

We would like to invite you to participate in our research study. This information sheet has been designed to help you to decide whether you would like to take part. It will explain the purpose of the research, what will be required of you as a participant, and how the data gathered from you will be managed. Please take a couple of minutes to read through the information below and ask any questions about the research that you may have.

Information about the research

This research study is being carried out and funded by the University of Birmingham. The overall objective of this research is to develop best practice guidance for building the structure of a decision-analytic model, in conjunction with the opinion and work of expert modellers and clinicians.

This study has been approved by the Science, Technology, Engineering and Mathematics Ethical Review Committee at the University of Birmingham.

Why have I been invited to take part?

You have been invited to take part in this research due to your involvement in the model development process at this modelling centre and due to your clinical or modelling expertise. This research will be in the format of a case study which will observe the development of a model from beginning to end with the intention of reporting on best practice guidance that can be used by those involved in the development of decision-analytic models nationally. It is anticipated that developing guidance that can be used to improve modelling processes will lead to better quality models generally which can be used to inform important funding and policy-making decisions. The guidance will also be designed to help those that are new to the modelling process, making it useful for training purposes. The case studies will involve the researcher (Samantha Husbands) being present to observe modelling meetings and being included in email and telephone correspondence between the modelling and clinical team. Samantha will also be present to conduct interviews with any clinical or modelling experts who are willing to discuss their thoughts and opinions on the observed modelling process. Samantha's involvement in the modelling process will be entirely non-participatory and so is not expected to disrupt the flow of work within the centre.

Do I have to take part?

No, you are under no obligation to participate. Your involvement is completely voluntary, but would be very beneficial to this research.

What will happen if I take part?

Should you agree to take part you will be asked for written consent to be involved in the research and for your involvement to be audio-recorded (i.e. discussion in modelling meetings). You may also be asked to take part in a short interview on the modelling process however, you are under no obligation to agree to this.

Can I withdraw from the research?

Yes, at any time. You are not obliged to provide a reason for your decision to withdraw. Upon withdrawal, we will ask you whether you give permission for us to continue to use your data. If you do not give permission we will automatically remove it from the study and destroy it. If you decide to withdraw we ask that this request is given within one week of your involvement in the research to avoid the data being analysed and used to inform future observation and interviews.

How will the data from my involvement in the case study be used?

It will be used to inform the content and format of the best practice modelling guidance which will be the final outcome of this research. The aim of the case study is to observe and discuss what constitutes best practice in the modelling process, where guidance is needed in model building and what the final guidance should include. The interviews will be used to follow-up on what Samantha has observed in the meetings, email correspondence and the model development process generally. They will focus on the opinions of modellers and clinicians on the current process of model development in relation to best practice, and what should be documented in the final guidance.

The findings of this research will form an integral part of a PhD thesis, with key results being presented through seminars, academic journals and conferences.

Is the research confidential?

Yes, the research is confidential. Any data you give will be anonymised and assigned a unique identity code to enable us to store your data safely and keep your personal information and digital recordings separate. All personal and interview information will be kept on a secured computer network which only the lead researcher (Samantha) will have access to. Data will be stored at the University of Birmingham for 10 years. Any writing or publications relevant to this research will not include any information that can identify you.

Where can I get more information about the study?

If you require any further information about the research please contact Samantha Husbands on [REDACTED]

What if I have a complaint about the research?

If you want to raise any concerns about the research please contact Professor Joanna Coast at j.coast@bham.ac.uk.

What now?

Please take your time to decide whether you would like to take part in this study and ensure that you ask any other questions that you may have.

If you would like to take part in the research please sign the consent form.

Thank you.

Appendix 19: Interview guide for phase two

CHECKLIST	REPEAT PURPOSE OF THE RESEARCH INTRODUCE TAPE RECORDER	CONFIDENTIALITY AND ANONYMITY RIGHT TO WITHDRAW	CONSENT FORM TRANSCRIPT
1.0 The current model (QUESTIONS SPECIFICALLY ABOUT THE CURRENT MODEL BEING WORKED ON)	<p>Can you tell me about your role in the development of the current model?</p> <ul style="list-style-type: none"> • What does it involve? • At what stages are you involved? • Who else do you work with on a model? • What is their involvement? <p>Can you talk me through the development process so far of the model that you are working on?</p> <p>How do you feel that the model's development is going?</p> <ul style="list-style-type: none"> • Is there anything that you feel has gone particularly well? • Is there anything that you feel has not gone as well? 		<p>What is your opinion of the structure of the model that has been developed so far?</p> <ul style="list-style-type: none"> • What do you think is good about it? • What do you think is not as good? • Where do you think the greatest uncertainty lies within the model? <p>How do you feel that the communication is going between everyone involved in the process (between the modellers and clinicians)?</p> <ul style="list-style-type: none"> • Can you tell me more about how this communication is managed? • Do you think that there is any way in which this communication could be improved? <p>What are your thoughts on the model development process that is used?</p> <ul style="list-style-type: none"> • What are its strengths? • What are its weaknesses? • Where and how could it be improved? <p>What is the next stage of your involvement in the modelling process?</p>
2.0 Background	<p>Have you worked on many models?</p> <ul style="list-style-type: none"> • Here? Elsewhere? • Has your role always been the same on every model (even if elsewhere)? 		
3.0 General modelling questions	<p>When you've been involved in the development of other models, was there anything different about the process used?</p> <ul style="list-style-type: none"> • Was it better or not as good as the current process? • In terms of communication? <p>Are you given any guidance on how to conduct your role in the modelling process?</p>		<p>Which aspects of your role do you find particularly difficult?</p> <p>Do you think that there is anything that can be done to make your role in the modelling process easier?</p> <ul style="list-style-type: none"> • In terms of what the clinicians/modellers involved could do?
4.0 Outcome of the research	<p>[QUESTIONS FOR MODELLERS] I am gauging from the interviews and the previous research that guidance on 'X' area of the modelling process might be important, what do you think?</p> <ul style="list-style-type: none"> • What do you think the content of this guidance should be? • Are there any other areas which you think would be important? 		
5.0 Ad-hoc questions	<p>Ask any additional questions which have been generated through observation or which are relevant to the particular stage of model development.</p>	6.0 Any other points?	<p>Is there anything else you would like to add which could contribute to this research?</p>

Appendix 20: Informant interview information sheet – phase two



Best Practice in Decision-analytic Model-Building Case Study – Interview Information Sheet

Additional information

You have already been given an information sheet which outlines the purpose and nature of the case study that you have agreed to be involved in. Another copy of this is attached for your information. This information sheet has been designed to provide additional information on the interview element to help you to decide whether you would like to take part. It will outline the purpose of the interview, what is required of you as a participant and how the data gathered from you will be managed. Please take a couple of moments to read through the information below and ask any questions about the interview that you may have.

What is the purpose of the interview?

The interview element has been designed to allow the researcher (Samantha Husbands) to follow-up and ask questions on the model building practice that she has observed during the case study. The aim of this is to decide through discussion with modellers and clinicians involved in the modelling process, what constitutes best practice in a model's development.

Why have I been invited to take part?

Having already agreed to take part in the observation element of this case study we are now asking you to take part in a face-to-face, semi-structured interview which will allow Samantha to learn in more detail about the modelling process, your thoughts and opinions on best practice, and the modelling guidance that will be produced as the outcome of this study.

Do I have to take part?

No, you are under no obligation to participate. Your involvement is completely voluntary, but would be very beneficial to this research.

What will happen if I take part?

If you agree to take part the interview will last up to one hour. During this time you will be asked to discuss your opinion on the current model that you are working on and how the methods used relate to what you consider to be best practice in model development. You will also be asked for your opinion on what the final guidance produced through this research should focus on and include. Your thoughts and opinions will inform what is reported as best practice in the model building guidance that will be developed through this research.

Can I withdraw from the research?

Yes, at any time. You are not obliged to provide a reason for your decision to withdraw. Upon withdrawal, we will ask you whether you give permission for us to continue to use your data. If you do not give permission we will automatically remove it from the study and destroy it. If you decide to withdraw we ask that this request is given within one week of the interview's completion to avoid the data being analysed and used to inform future observation and interviews.

How will the data from my interview be used?

It will be used to inform the content and format of the best practice modelling guidance which will be the final outcome of this research. The data gathered from the observations and the interviews will be used in conjunction to allow Samantha to understand what should be considered and documented as best practice in model building.

The findings of this research will form an integral part of a PhD thesis, with key results being presented through seminars, academic journals and conferences.

Is the research confidential?

Yes, the research is confidential. Any data you give will be anonymised and assigned a unique identity code to enable us to store your data safely and keep your personal information and digital recordings separate. All personal and interview information will be kept on a secured computer network which only the lead researcher will have access to. Data will be stored at the University of Birmingham for 10 years. Any writing or publications relevant to this research will not include any information that can identify you.

Where can I get more information about the study?

If you require any further information about the research please contact Samantha Husbands on



What if I have a complaint about the research?

If you want to raise any concerns about the research please contact Professor Joanna Coast at j.coast@bham.ac.uk.

What now?

Please take your time to decide whether you would like to take part in this study and ensure that you ask any other questions that you may have.

If you would like to take part in the research please sign the consent form.

Thank you.

Appendix 21: Consent form for phase two of the research (consent for interview)

Best Practice in Decision-analytic Model Building – Study Consent Form

Additional Consent for Interview

Participant Identification Code:
Name of Researcher: Samantha Husbands

To take part in the study you must agree to the following:

Please Initial Boxes and
Sign Below to Consent

1. I confirm that I have read and understood the interview information sheet for the research entitled ‘best practice in decision-analytic model building study’. I confirm that I have had the opportunity to consider the information provided, ask any questions and have those questions answered.
2. I understand that my participation in the interview is entirely voluntary and that I am free to withdraw from the interview at any time without giving any reason or explanation. I understand that any personal information collected within this project will be anonymised and will remain strictly confidential.
3. I agree to participate in the above interview and consent to it being audio recorded. I understand that the data may be looked at by the Research Team at the University of Birmingham and anonymised quotes from this interview may be included in research reports.

I hereby consent to participate in this research project and confirm that I understand and agree with the statements listed above.

Signature of Participant:

Date:

Signature of Interviewer:

Date:

Should you have any further queries or concerns please contact:

Sam Husbands
Health Economics Unit
Public Health Building
University of Birmingham
Birmingham
B15 2TT



Appendix 22: Thematic framework for Case Study A

1.1 Background

1.1.1 Objective of the model

1.1.2 Modelling team

1.2 Structural development

1.2.1 Patient population

1.2.2 Learning about condition

1.2.3 Pathway plotting

1.2.3.1 Beginning the pathway

1.2.3.2 Where patients go and why

1.2.3.3 Decisions on how to model pathways

1.2.3.4 Outcome of pathways

1.2.3.5 Making structural assumptions

1.2.3.5.1 Process for making assumptions

1.2.3.5.2 Assumptions made

1.2.3.6 Determining model boundaries

1.2.3.6.1 Time horizon

1.2.3.6.2 Excluding patients

1.2.3.7 Basis of a structure

1.2.3.7.1 NICE guidelines

1.2.3.7.2 Use of clinical literature

1.2.3.7.3 Use of clinical effectiveness data

1.2.3.8 Comparator

1.2.3.8.1 Allowing for misdiagnosis

1.2.3.9 Location of treatment pathway (generalisability)

1.2.4 Communication in structural development

1.2.4.1 Clarification of health economic terms

1.2.4.2 Miscommunication/misunderstanding

1.2.5 Clinical collaborators

- 1.2.6 Iterative
- 1.2.7 Implementation in software

1.3 Data: populating the model

- 1.3.1 Clinical data
 - 1.3.1.1 Probability data
 - 1.3.1.2 Clinical-effectiveness data
- 1.3.2 Economic data
 - 1.3.2.1 Cost data
 - 1.3.2.2 Utility data
- 1.3.3 Clinician involvement
- 1.3.4 Problems with data

1.4 Results

- 1.4.1 Results
 - 1.4.1.1 Questioning results
 - 1.4.1.2 Questioning assumptions
 - 1.4.1.3 Miscommunication over results
 - 1.4.1.4 Presentation of results

1.5 Model checking

- 1.5.1 Model checking
 - 1.5.1.1 Secondary analysis
 - 1.5.1.2 Internal validation
 - 1.5.1.2.1 Sensitivity analysis
 - 1.5.1.3 External validation
 - 1.5.1.4 Structural validation
 - 1.5.1.5 Clinician involvement

1.6 Clinician involvement

- 1.6.1 Number of clinicians
- 1.6.2 Meetings with clinicians
- 1.6.3 Clinician understanding of health economics and modelling
- 1.6.4 Guidance for clinicians

1.7 Informant reflections

1.7.1 Reflections on model structure

1.7.2 Reflections on roles in model development

1.7.3 Reflections on communication

1.7.4 Issues and problems within model development

Appendix 23: Thematic framework for Case Study B

1.1 Background

- 1.1.1 Objective of the model
- 1.1.2 Modelling team
- 1.1.3 Clinician training

1.2 Structural development

- 1.2.1 Patient population
- 1.2.2 Learning about condition
- 1.2.3 Pathway plotting
 - 1.2.3.1 Beginning the pathway
 - 1.2.3.2 Where patients go and why
 - 1.2.3.3 Decisions on how to model pathways
 - 1.2.3.4 Outcome of pathways
 - 1.2.3.5 Making structural assumptions
 - 1.2.3.5.1 Process for making assumptions
 - 1.2.3.5.2 Assumptions made
 - 1.2.3.6 Determining model boundaries
 - 1.2.3.6.1 Time horizon
 - 1.2.3.6.2 Excluding patients
 - 1.2.3.7 Basis of a structure
 - 1.2.3.7.1 NICE guidelines
 - 1.2.3.7.2 Use of clinical literature
 - 1.2.3.7.3 Use of clinical effectiveness data
 - 1.2.3.8 Comparator
 - 1.2.3.8.1 Allowing for misdiagnosis
 - 1.2.3.9 Location of treatment pathway (generalisability)
- 1.2.4 Communication in structural development
 - 1.2.4.1 Clarification of health economic terms
 - 1.2.4.2 Miscommunication/misunderstanding

- 1.2.5 Clinical collaborators
- 1.2.6 Iterative
- 1.2.7 Implementation in software

1.3 Data: populating the model

- 1.3.1 Clinical data
 - 1.3.1.1 Probability data
 - 1.3.1.2 Clinical-effectiveness data
- 1.3.2 Economic data
 - 1.3.2.1 Cost data
 - 1.3.2.2 Utility data
- 1.3.3 Clinician involvement
- 1.3.4 Problems with data

1.4 Results

- 1.4.1 Results
 - 1.4.1.1 Questioning results
 - 1.4.1.2 Questioning assumptions
 - 1.4.1.3 Miscommunication over results
 - 1.4.1.4 Presentation of results

1.5 Model checking

- 1.5.1 Model checking
 - 1.5.1.1 Secondary analysis
 - 1.5.1.2 Internal validation
 - 1.5.1.2.1.1 Sensitivity analysis
 - 1.5.1.3 External validation
 - 1.5.1.4 Structural validation
 - 1.5.1.5 Clinician involvement

1.6 Clinician involvement

- 1.6.1 Number of clinicians
- 1.6.2 Meetings with clinicians
- 1.6.3 Clinician understanding of health economics and modelling

1.6.4 Guidance for clinicians

1.7 Informant reflections

1.7.1 Reflections on model structure

1.7.2 Reflections on roles in model development

1.7.3 Reflections on communication

1.7.4 Issues and problems within model development

REFERENCES

- Andronis, L., Barton, P., Bryan, S., 2009. Sensitivity analysis in economic evaluation: an audit of NICE current practice and a review of its use and value in decision-making. *Health Technol. Assess.* 13, ix–61.
- Arrow, K.J., 1963. Uncertainty and the Welfare Economics of Medical Care. *Am. Econ. Rev.* LIII, 941–973.
- Baker, R., Robinson, A., 2004. Responses to standard gambles: are preferences “well constructed”? *Health Econ.* 13, 37–48.
- Balshem, H., Helfand, M., Schünemann, H.J., Oxman, A.D., Kunz, R., Brozek, J., Vist, G.E., Falck-Ytter, Y., Meerpohl, J., Norris, S., Guyatt, G.H., 2011. GRADE guidelines: 3. Rating the quality of evidence. *J. Clin. Epidemiol.* 64, 401–406.
- Barton, P., Bryan, S., Robinson, S., 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J. Heal. Serv. Res. Policy* 9, 110–118.
- Baxter, P., Jack, S., 2008. Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *Qual. Rep.* 13, 544–559.
- Birch, S., Donaldson, C., 2003. Valuing the benefits and costs of health care programmes: where’s the “extra” in extra-welfarism? *Soc. Sci. Med.* 56, 1121–1133.
- Black, J., Hashimzade, N., Myles, G., 2009. *A Dictionary of Economics*, 3rd ed. Oxford University Press, Oxford.
- BMJ, 2015. Editor’s checklists: Qualitative research checklist [WWW Document]. *BMJ*. URL <http://www.bmj.com/about-bmj/resources-authors/article-types/research/editors-checklists> (accessed 5.24.15).
- Boadway, R., Bruce, N., 1984. *Welfare economics*. Blackwell, Oxford.
- Boren, T., Ramey, J., 2000. Thinking aloud: reconciling theory and practice. *IEEE Trans. Prof. Commun.* 43, 261–278.
- Bowen, G.A., 2008. Naturalistic inquiry and the saturation concept: a research note. *Qual. Res.* 8, 137–152.
- Brennan, A., Chick, S.E., Davies, R., 2006. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ.* 15, 1295–1310.
- Briggs, A., Claxton, K., Sculpher, M., 2006. *Decision modelling for health economic evaluation*. Oxford University Press, New York.

- Briggs, A.H., 2001. Handling uncertainty in economic evaluation and presenting the results, in: Drummond, M., McGuire, A. (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, Oxford, pp. 172–214.
- Briggs, A.H., 2000. Handling Uncertainty in Cost-Effectiveness Models. *Pharmacoeconomics* 17, 479–500.
- Briggs, A.H., Weinstein, M.C., Fenwick, E. a L., Karnon, J., Sculpher, M.J., Paltiel, A.D., 2012. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Heal.* 15, 835–842.
- Brouwer, W.B.F., Culyer, A.J., van Exel, N.J. a., Rutten, F.F.H., 2008. Welfarism vs. extra-welfarism. *J. Health Econ.* 27, 325–338.
- Bryman, A., 2008. *Social Research Methods*, 3rd ed. Oxford University Press, Oxford.
- Buxton, M.J., Drummond, M.F., Van Hout, B. a., Prince, R.L., Sheldon, T. a., Szucs, T., Vray, M., 1997. Modelling in Economic Evaluation: An Unavoidable Fact of Life. *Health Econ.* 6, 217–227.
- Cairns, J., 2001. Discounting in economic evaluation, in: Drummond, M., McGuire, A. (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, Oxford, pp. 236–255.
- Canadian Agency for Drugs and Technologies in Health, 2015. About CADTH: What Does the Evidence Say? [WWW Document]. CADTH. URL <https://www.cadth.ca/about-cadth> (accessed 10.19.15).
- Canadian Agency for Drugs and Technologies in Health, 2014. Submission Guidelines for the CADTH Common Drug Review [WWW Document]. CADTH Common Drug Rev. URL https://www.cadth.ca/media/cdr/process/CDR_Submission_Guidelines.pdf (accessed 12.10.15).
- Canadian Agency for Drugs and Technologies in Health, 2006. Guidelines for the economic evaluation of health technologies: Canada [WWW Document]. URL http://cadth.ca/media/pdf/186_EconomicGuidelines_e.pdf (accessed 4.25.14).
- Caro, J.J., Briggs, A.H., Siebert, U., Kuntz, K.M., 2012. Modeling Good Research Practices - Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force - 1. *Value Heal.* 15, 796–803.
- Centre for Reviews and Dissemination, 2008. Systematic Reviews: CRD's guidance for undertaking reviews in health care [WWW Document]. URL http://www.york.ac.uk/media/crd/Systematic_Reviews.pdf (accessed 1.30.14).
- Chen, Y.F., Jobanputra, P., Barton, P., Jowett, S., Bryan, S., Clark, W., Fry-Smith, A., Burls, A., 2006. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Heal. Technol Assess.*

- Chilcott, J., Tappenden, P., Rawdin, A., Johnson, M., Kaltenthaler, E., Paisley, S., Papaioannou, D., Shippam, A., 2010. Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review. *Heal. Technol Assess.*
- Coast, J., 2009. Maximisation in extra-welfarism: A critique of the current position in health economics. *Soc. Sci. Med.* 69, 786–792.
- Coast, J., 2004. Is economic evaluation in touch with society's health values? *BMJ* 329, 1233–1236.
- Coast, J., 1999. The appropriate uses of qualitative methods in health economics. *Health Econ.* 8, 345–353.
- Coast, J., Horrocks, S., 2007. Developing attributes and levels for discrete choice experiments using qualitative methods. *J. Health Serv. Res. Policy* 12, 25–30.
- Coast, J., McDonald, R., Baker, R., 2004. Issues arising from the use of qualitative methods in health economics. *J. Health Serv. Res. Policy* 9, 171–176.
- Coast, J., Smith, R.D., Lorgelly, P., 2008. Welfarism, extra-welfarism and capability: The spread of ideas in health economics. *Soc. Sci. Med.* 67, 1190–1198.
- Coffey, A., 1999. *The Ethnographic Self: Fieldwork and the Representation of Identity.* SAGE Publications, London.
- Cooper, K., Brailsford, S., Davies, R., 2007. Choice of modelling technique for evaluating health care interventions. *J. Oper. Res. Soc.* 58, 168–176.
- Cooper, N.J., Sutton, A.J., Abrams, K.R., Turner, D., Wailoo, A., 2004. Comprehensive decision analytical modelling in economic evaluation: A Bayesian approach. *Health Econ.* 13, 203–226.
- Creswell, J.W., 2013. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, 3rd ed. SAGE Publications, Thousand Oaks, California.
- Crotty, M., 1998. *The Foundations of Social Research: Meaning and Perspective in the Research Process.* SAGE Publications, London.
- Culyer, A.J., 1989. The Normative Economics of Health Care Finance and Provision. *Oxford Rev. Econ. Policy* 5, 34–58.
- Dancy, J., 2001. *Introduction to contemporary epistemology.* Blackwell, Oxford.
- Das Nair, R., Orr, K.S., Vedhara, K., Kendrick, D., 2014. Exploring recruitment barriers and facilitators in early cancer detection trials: the use of pre-trial focus groups. *Trials* 15, 98.
- Denzin, N.K., Lincoln, Y.S., 2000. Introduction: The Discipline and Practice of Qualitative Research, in: Denzin, N.K., Lincoln, Y.S. (Eds.), *Handbook of Qualitative Research.* SAGE Publications, Thousand Oaks, California, pp. 1–28.

- Denzin, N.K., Lincoln, Y.S., 1998. Introduction: Entering the Field of Qualitative Research, in: Denzin, N.K., Lincoln, Y.S. (Eds.), *The Landscape of Qualitative Research: Theories and Issues*. SAGE Publications, Thousand Oaks, California, pp. 1–34.
- Donaldson, C., 1998. The (near) equivalence of cost-effectiveness and cost-benefit analyses. Fact or fallacy? *Pharmacoeconomics* 13, 389–396.
- Donovan, J., Mills, N., Smith, M., Brindle, L., Jacoby, A., Peters, T., Frankel, S., Neal, D., Hamdy, F., 2002. Improving design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *BMJ* 325, 766–770.
- Donovan, J.L., de Salis, I., Toerien, M., Paramasivan, S., Hamdy, F.C., Blazeby, J.M., 2014a. The intellectual challenges and emotional consequences of equipoise contributed to the fragility of recruitment in six randomized controlled trials. *J. Clin. Epidemiol.* 67, 912–20.
- Donovan, J.L., Paramasivan, S., de Salis, I., Toerien, M., 2014b. Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials. *Trials* 15, 5.
- Doubilet, P., Begg, C.B., Weinstein, M.C., Braun, P., McNeil, B.J., 1985. Probabilistic Sensitivity Analysis Using Monte Carlo Simulation. *Med Decis Mak.* 5, 157–177.
- Drummond, M., Sculpher, M., Torrance, G., O’Brien, B., Stoddart, G., 2005. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, Oxford.
- Eborall, H.C., Stewart, M.C.W., Cunningham-Burley, S., Price, J.F., Fowkes, F.G.R., 2011. Accrual and drop out in a primary prevention randomised controlled trial: qualitative study. *Trials* 12, 7.
- Eddy, D.M., Hollingworth, W., Caro, J.J., Tsevat, J., McDonald, K.M., Wong, J.B., 2012. Model Transparency and Validation : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Value Heal.* 15, 843–850.
- Ericsson, K.A., Simon, H.A., 1984. *Protocol analysis: Verbal reports as data*. MIT Press, Cambridge.
- Evans, D., 2003. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J. Clin. Nurs.* 12, 77–84.
- Featherstone, K., Donovan, J.L., 1998. Random allocation or allocation at random? Patients’ perspectives of participation in a randomised controlled trial. *BMJ* 317, 1177–80.
- Fegert, J.M., Slawik, L., Wermelskirchen, D., Nübling, M., Mühlbacher, A., 2011. Assessment of parents’ preferences for the treatment of school-age children with ADHD: a discrete choice experiment. *Expert Rev. Pharmacoeconomics Outcomes Res.* 11, 245–252.

- Firestone, W. a., 1993. Alternative Arguments for Generalizing From Data as Applied to Qualitative Research. *Educ. Res.* 22, 16–23.
- Flick, U., 2009. *An Introduction to Qualitative Research*, 4th ed. SAGE Publications, London.
- Flick, U., von Kardorff, E., Steinke, I., 2004. What is Qualitative Research? An Introduction to the Field, in: Flick, U., von Kardorff, E., Steinke, I. (Eds.), *A Companion to Qualitative Research*. SAGE Publications, London, pp. 3–11.
- Flyvbjerg, B., 2006. Five Misunderstandings About Case-Study Research. *Qual. Inq.* 12, 219–245.
- Fonteyn, M.E., Kuipers, B., Grobe, S.J., 1993. A Description of Think Aloud Method and Protocol Analysis. *Qual. Health Res.* 3, 430–441.
- Gagliardi, A.R., Dobrow, M.J., 2011. Paucity of qualitative research in general medical and health services and policy research journals: analysis of publication rates. *BMC Health Serv. Res.* 11, 268.
- Gale, N.K., Heath, G., Cameron, E., Rashid, S., Redwood, S., 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med. Res. Methodol.* 13, 117.
- Garber, A., Weinstein, M.C., Torrance, G., Kamlet, M., 1996. Theoretical Foundations of Cost-Effectiveness Analysis, in: Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 25–53.
- Garrison, L., 2003. The ISPOR Good Practice Modeling Principles - A Sensible Approach: Be Transparent, Be Reasonable. *Value Heal.* 6, 6–8.
- Geneau, R., Massae, P., Courtright, P., Lewallen, S., 2008. Using qualitative methods to understand the determinants of patients' willingness to pay for cataract surgery: a study in Tanzania. *Soc. Sci. Med.* 66, 558–68.
- Glaser, B.G., 1965. The Constant Comparative Method of Qualitative Analysis. *Soc. Probl.* 12, 436–445.
- Glaser, B.G., Strauss, A.L., 1968. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Weidenfeld and Nicolson, London.
- Glesne, C., 2006. *Becoming Qualitative Researchers: An Introduction*, 3rd ed. Pearson, New York.
- Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C., 1996. *Cost-Effectiveness in Health and Medicine*. Oxford University Press, Oxford.
- Gopinath, D., Smith, A.R.B., Holland, C., Reid, F.M., 2013. Why don't women participate? A qualitative study on non-participation in a surgical randomised controlled trial. *Int. Urogynecol. J.* 24, 969–75.

- Gray, A.M., Clarke, P.M., Wolstenholme, J.L., Wordsworth, S., 2011. *Applied Methods of Cost-effectiveness Analysis in Health care*. Oxford University Press, Oxford.
- Green, J., Thorogood, N., 2009. *Qualitative Methods for Health Research*, 2nd ed. SAGE Publications, London.
- Guba, E.G., 1990. The Alternative Paradigm Dialog, in: Guba, E.G. (Ed.), *The Paradigm Dialog*. SAGE Publications, Newbury Park, California, pp. 17–30.
- Guba, E.G., Lincoln, Y.S., 1998. Competing Paradigms in Qualitative Research, in: Denzin, N.K., Lincoln, Y.S. (Eds.), *The Landscape of Qualitative Research: Theories and Issues*. SAGE Publications, Thousand Oaks, California, pp. 195–220.
- Halpern, M.T., Luce, B.R., Brown, R.E., Geneste, B., 1998. Health and Economic Outcomes Modeling Practices: A Suggested Framework. *Value Heal.* 1, 131–147.
- Hamilton, D.W., de Salis, I., Donovan, J.L., Birchall, M., 2013. The recruitment of patients to trials in head and neck cancer: a qualitative study of the EaStER trial of treatments for early laryngeal cancer. *Eur. Archives Otorhinolaryngol. Head Neck Surg.* 270, 2333–7.
- Hammersley, M., 1992. *What’s wrong with ethnography?* Routledge, London.
- Hart, E., Lymbery, M., Gladman, J.R., 2005. Away from home: an ethnographic study of a transitional rehabilitation scheme for older people in the UK. *Soc. Sci. Med.* 60, 1241–1250.
- Haynes, K., 2012. Reflexivity in Qualitative Research, in: Symon, G., Cassel, C. (Eds.), *Qualitative Organizational Research: Core Methods and Current Challenges*. SAGE Publications, London, pp. 72–89.
- Hicks, J., 1939. The Foundations of Welfare Economics. *Econ. J.* 49, 694–712.
- Hopf, C., 2004. Research Ethics and Qualitative Research, in: Flick, U., Kardorff, E. v., Steinke, I. (Eds.), *A Companion to Qualitative Research*. SAGE Publications, London, pp. 334–339.
- Hurley, J., 1998. Welfarism, Extra-Welfarism and Evaluative Economic Analysis in the Health Sector, in: Barer, M.L., Getzen, T.E., Stoddart, G.L. (Eds.), *Health, Health Care and Health Economics*. John Wiley & Sons, Chichester, pp. 373–395.
- INVOLVE, 2015. What is public involvement in research? [WWW Document]. URL <http://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/> (accessed 11.20.15).
- Jain, R., Grabner, M., Onukwugha, E., 2011. Sensitivity Analysis in Cost-Effectiveness studies: From Guidelines to Practice. *Pharmacoeconomics* 29, 297–314.
- Kaldor, N., 1939. Welfare Propositions of Economics and Interpersonal Comparisons of Utility. *Econ. J.* 49, 549–552.

- Kaltenthaler, E., Essat, M., Tappenden, P., Paisley, S., 2014. Identification and review of cost-effectiveness model parameters: a qualitative study. *Int. J. Technol. Assess. Health Care* 30, 333–40.
- Kaltenthaler, E., Tappenden, P., Paisley, S., 2013. Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments. *Value Heal.* 16, 830–6.
- Kaltenthaler, E., Tappenden, P., Paisley, S., Squires, H., 2011. NICE DSU Technical Support Document 13: Identifying And Reviewing Evidence To Inform The Conceptualisation And Population Of Cost-Effectiveness Models [WWW Document]. NICE Decis. Support Unit. URL http://www.nicedsu.org.uk/TSD_13_model_parameters.pdf (accessed 3.11.15).
- Kaltenthaler, E.C., Dickson, R., Boland, A., Carroll, C., Fitzgerald, P., Papaioannou, D., Akehurst, R., 2012. A qualitative study of manufacturers' submissions to the UK NICE single technology appraisal process. *BMJ Open* 2, e000562.
- Karnieli-Miller, O., Strier, R., Pessach, L., 2009. Power relations in qualitative research. *Qual. Health Res.* 19, 279–89.
- Karnon, J., Brown, J., 1998. Selecting a decision model for economic evaluation: a case study and review. *Health Care Manag. Sci.* 1, 133–140.
- Karnon, J., Goyder, E., Tappenden, P., McPhie, S., Towers, I., Brazier, J., Madan, J., 2007. A review and critique of modelling in prioritising and designing screening programmes. *Heal. Technol Assess* 11.
- Karnon, J., Stahl, J., Brennan, A., Caro, J.J., Mar, J., Möller, J., 2012. Modeling using Discrete Event Simulation : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value Heal.* 15, 821–827.
- Ke, K.M., Mackichan, F., Sandy, J.R., Ness, A.R., Hollingworth, W., 2013. Parents' perspectives on centralized cleft services for children: the development of a DCE questionnaire. *Oral Dis.* 19, 185–92.
- Kolb, S.M., 2012. Grounded Theory and the Constant Comparative Method : Valid Research Strategies for Educators. *J. Emerg. Trends Educ. Res. Policy Stud.* 3, 83–86.
- Krefting, L., 1991. Rigor in qualitative research: the assessment of trustworthiness. *Am. J. Occup. Ther.* 45, 214–222.
- Kuntz, K.M., Weinstein, M.C., 2001. Modelling in economic evaluation, in: Drummond, M., McGuire, A. (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, New York, pp. 141–171.
- Kuper, A., Lingard, L., Levinson, W., 2008. Critically appraising qualitative research. *BMJ* 337, a1035.

- Lefebvre, C., Manheimer, E., Glanville, J., 2008. Searching for studies, in: Higgins, J.P.T., Green, S. (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley-Blackwell, Chichester, pp. 95–150.
- Legard, R., Keegan, J., Ward, K., 2003. Qualitative Research Practice: A Guide for Social Science Students and Researchers, in: Richie, J., Lewis, J. (Eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE Publications, London, pp. 138–169.
- Lipscomb, J., Weinstein, M.C., Torrance, G., 1996. Time Preference, in: Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 214–246.
- Malterud, K., 2001. Qualitative research: standards, challenges, and guidelines. *Lancet* 358, 483–8.
- Mandelblatt, J., Fryback, D., Weinstein, M.C., Russel, L.B., Gold, M.R., Hadorn, D., 1996. Assessing the Effectiveness of Health Interventions, in: Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 135–175.
- Manning, W., Fryback, D., Weinstein, M.C., 1996. Reflecting Uncertainty in Cost-Effectiveness Analysis, in: Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 247–275.
- Mays, N., Pope, C., 2000. Qualitative research in health care: Assessing quality in qualitative research. *BMJ* 320, 50–2.
- McCabe, C., Claxton, K., Culyer, A.J., 2008. The NICE cost-effectiveness threshold: What it is and what that means. *Pharmacoeconomics* 26, 733–744.
- McCabe, C., Dixon, S., 2000. Testing the Validity of Cost-Effectiveness Models. *Pharmacoeconomics* 17, 501–513.
- McGuire, A., 2001. Theoretical concepts in the economic evaluation of health care, in: Drummond, M., McGuire, A. (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, New York, pp. 1–21.
- McNamara, C., 2009. General Guidelines for Conducting Research Interviews [WWW Document]. Free Manag. Libr. URL <http://managementhelp.org/businessresearch/interviews.htm> (accessed 4.3.14).
- McPake, B., Normand, C., 2008. *Health Economics: An International Perspective*, Second ed. Routledge, New York.
- Merkens, H., 2004. Selection Procedures, Sampling, Case Construction, in: Flick, U., von Kardorff, E., Steinke, I. (Eds.), *A Companion to Qualitative Research*. SAGE Publications, London, pp. 165–171.

- Miguel, F.S., Ryan, M., Amaya-Amaya, M., 2005. "Irrational" stated preferences: a quantitative and qualitative investigation. *Health Econ.* 14, 307–22.
- Miles, M.B., Huberman, M., 1984. Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft. *Educ. Res.* 13, 20–30.
- Miller, W.L., Crabtree, B.F., 1999. Clinical Research: A Multimethod Typology and Qualitative Roundmap, in: Miller, W.L., Crabtree, B.F. (Eds.), *Doing Qualitative Research*. SAGE Publications, Thousand Oaks, California, pp. 3–26.
- Mills, N., Donovan, J.L., Wade, J., Hamdy, F.C., Neal, D.E., Lane, J.A., 2011. Exploring treatment preferences facilitated recruitment to randomized controlled trials. *J. Clin. Epidemiol.* 64, 1127–36.
- Miners, A., Llewellyn, C., Pollard, A., Lagarde, M., Richardson, D., Cairns, J., Fisher, M., Smith, H., 2012. Assessing user preferences for sexually transmitted infection testing services: a discrete choice experiment. *Sex. Transm. Infect.* 88, 510–6.
- Miners, A., Garau, M., Fidan, D., Fischer, A., 2004. Comparing estimates of cost effectiveness submitted to the National Institute for Clinical Excellence (NICE) by different organisations: retrospective study. *BMJ*.
- Moher, D., Cook, D.J., Eastwood, S., Olkin, I., Rennie, D., Stroup, D.F., 1999. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 354, 1896–1900.
- Mooney, G., Russell, E., 2003. Equity in Health Care: The Need for a New Economics Paradigm?, in: Scott, A., Maynard, A., Elliot, R. (Eds.), *Advances in Health Economics*. John Wiley & Sons, Chichester, pp. 205–222.
- Morris, S., Devlin, N., Parkin, D., 2007. *Economic Analysis in Health Care*. John Wiley & Sons, Ltd, Chichester.
- National Institute for Health and Care Excellence, 2015a. What we do [WWW Document]. NICE. URL <https://www.nice.org.uk/about/what-we-do> (accessed 10.19.15).
- National Institute for Health and Care Excellence, 2015b. NICE technology appraisal guidance [WWW Document]. NICE. URL <https://www.nice.org.uk/About/What-we-do/Our-Programmes/NICE-guidance/NICE-technology-appraisal-guidance> (accessed 10.19.15).
- National Institute for Health and Care Excellence, 2015c. Patient and public involvement policy [WWW Document]. Natl. Inst. Heal. Care Excell. URL <https://www.nice.org.uk/media/default/About/NICE-Communities/Public-involvement/Patient-and-public-involvement-policy/Patient-and-public-involvement-policy-November-2013.pdf> (accessed 10.25.15).
- National Institute for Health and Care Excellence, 2014. Guide to the processes of technology appraisal [WWW Document]. NICE Artic. [PMG19]. URL

- <https://www.nice.org.uk/article/pmg19/chapter/3-The-appraisal-process#mta-initiation-and-evidence-submission> (accessed 12.10.15).
- National Institute for Health and Care Excellence, 2013. Guide to the methods of technology appraisal 2013 [WWW Document]. URL <http://www.nice.org.uk/media/D45/1E/GuideToMethodsTechnologyAppraisal2013.pdf> (accessed 10.1.15).
- National Institute for Health and Care Excellence, 2012. Eribulin for the treatment of locally advanced or metastatic breast cancer [WWW Document]. NICE Technol. Apprais. Guid. 250. URL www.nice.org.uk/Guidance/TA250 (accessed 9.26.15).
- National Institute for Health and Care Excellence, 2011. Ranibizumab for the treatment of diabetic macular oedema [WWW Document]. NICE Technol. Apprais. Guid. 237. URL www.nice.org.uk/ta237 (accessed 9.26.15).
- National Institute for Health and Care Excellence, 2010. Eltrombopag for the treatment of chronic immune (idiopathic) thrombocytopenic purpura [WWW Document]. NICE Technol. Apprais. Guid. 205. URL www.nice.org.uk/guidance/TA205 (accessed 9.26.15).
- National Institute for Health and Care Excellence, 2007. Fludarabine monotherapy for the first-line treatment of chronic lymphocytic leukaemia [WWW Document]. NICE Technol. Apprais. Guid. 119. URL www.nice.org.uk/TA119 (accessed 9.26.15).
- Novick, G., 2008. Is There a Bias Against Telephone Interviews In Qualitative Research? *Res Nurs Heal.* 31, 391–398.
- O'Reilly, K., 2009. *Key Concepts in Ethnography*. SAGE Publications, London.
- Obermann, K., Scheppe, J., Glazinski, B., 2013. More Than Figures? Qualitative Research In Health Economics. *Health Econ.* 22, 253–257.
- Paisley, S., 2012. Identifying evidence to inform decision-analytic models of cost-effectiveness: a qualitative study of information seeking processes and behaviour. The University of Sheffield.
- Paramasivan, S., Huddart, R., Hall, E., Lewis, R., Birtle, A., Donovan, J.L., 2011. Key issues in recruitment to randomised controlled trials with very different interventions: a qualitative investigation of recruitment to the SPARE trial (CRUK/07/011). *Trials* 12, 78.
- Patton, M.Q., 2002. *Qualitative Research & Evaluation Methods*, 3rd ed. SAGE Publications, Thousand Oaks.
- Penaloza Ramos, M.C., Barton, P., Jowett, S., Sutton, A.J., 2015. A Systematic Review of Research Guidelines in Decision-Analytic Modeling. *Value Heal.* 18, 512–529.
- Petrou, S., Gray, A., 2011. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 342, d1766.

- Petticrew, M., Roberts, H., 2006. *Systematic reviews in the social sciences: a practical guide*. Blackwell, Oxford.
- Philips, Z., Bojke, L., Sculpher, M., Claxton, K., Golder, S., 2006. Good Practice Guidelines for Decision-Analytic Modelling in Health Technology Assessment. *Pharmacoeconomics* 24, 355–371.
- Phillips, C.D., 2006. What do you do for a living? Toward a more succinct definition of health services research. *BMC Health Serv. Res.* 6, 117.
- Pitman, R., Fisman, D., Zaric, G.S., Postma, M., Kretzschmar, M., Edmunds, J., Brisson, M., 2012. Dynamic Transmission Modeling : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-5. *Value Heal.* 15, 828–834.
- Polit, D.F., Beck, C.T., 2010. Generalization in quantitative and qualitative research: myths and strategies. *Int. J. Nurs. Stud.* 47, 1451–8.
- Pope, C., Mays, N., 1995. Qualitative Research: Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ* 311, 42–45.
- Richie, J., Lewis, J., Elam, G., 2003a. Qualitative Research Practice: A Guide for Social Science Students and Researchers, in: Richie, J., Lewis, J. (Eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE Publications, London, pp. 77–108.
- Richie, J., Spencer, L., 2002. Qualitative Data Analysis for Applied Policy Research, in: Huberman, M., Miles, M.B. (Eds.), *The Qualitative Researcher's Companion*. SAGE Publications, London, pp. 305–329.
- Richie, J., Spencer, L., 1994. Qualitative data analysis for applied policy research, in: Bryman, A., Burgess, R.G. (Eds.), *Analyzing Qualitative Data*. Routledge, pp. 173–194.
- Richie, J., Spencer, L., O'Connor, W., 2003b. Qualitative Research Practice: A Guide for Social Science Students and Researchers, in: Richie, J., Lewis, J. (Eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. SAGE Publications, London, pp. 219–262.
- Roberts, M., Russell, L.B., Paltiel, A.D., Chambers, M., Mcewan, P., Krahn, M., 2012. Conceptualizing a Model : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Value Heal.* 15, 804–811.
- Roberts, T.E., Robinson, S., Barton, P., Bryan, S., Low, N., 2006. Screening for Chlamydia trachomatis: a systematic review of the economic evaluations and modelling. *Sex. Transm. Infect.* 82, 193–200.
- Robinson, A., Dolan, P., Williams, A., 1997. Valuing health status using VAS and TTO: what lies behind the numbers? *Soc. Sci. Med.* 45, 1289–1297.
- Robinson, R., 1993a. Cost-benefit analysis. *BMJ* 307, 924–926.

- Robinson, R., 1993b. Cost-effectiveness analysis. *BMJ* 307, 793–5.
- Robinson, S., Bryan, S., 2013. Does the process of deliberation change individuals' health state valuations? An exploratory study using the person trade-off technique. *Value Health* 16, 806–13.
- Rubin, H.J., Rubin, I.S., 2005. *Qualitative Interviewing: The Art Of Hearing Data*, 2nd ed. SAGE Publications, Thousand Oaks.
- Russell, L.B., Siegel, J.E., Gold, M.R., Luce, B.R., Mandelblatt, J., 1996. Cost-Effectiveness Analysis as a Guide to Resource Allocation in Health: Roles and Limitations, in: Gold, M.R., Siegel, J.E., Russel, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 3–21.
- Ryan, M., Watson, V., Amaya-Amaya, M., 2003. Methodological issues in the monetary valuation of benefits in healthcare. *Expert Rev. Pharmacoeconomics Outcomes Res* 3, 717–727.
- Ryan, M., Watson, V., Entwistle, V., 2009. Rationalising the “irrational”: a think aloud study of discrete choice experiment responses. *Health Econ.* 18, 321–36.
- Saukko, P., 2003. *Doing research in cultural studies: an introduction to classical and new methodological approaches*. SAGE Publications, London.
- Schofield, J., 2000. Increasing the Generalizability of Qualitative Research, in: Gomm, R., Hammersley, M., Foster, P. (Eds.), *Case Study Method*. SAGE Publications, London, pp. 69–97.
- Schulze, B., Angermeyer, M.C., 2003. Subjective experiences of stigma. A focus group study of schizophrenic patients, their relatives and mental health professionals. *Soc. Sci. Med.* 56, 299–312.
- Sculpher, M., Fenwick, E., Claxton, K., 2000. Assessing Quality in Decision Analytic Cost-Effectiveness Models: A Suggested Framework and Example of Application. *Pharmacoeconomics* 17, 461–477.
- Sculpher, M.J., Claxton, K., Drummond, M., McCabe, C., 2006. Whither trial-based economic evaluation for health care decision making? *Health Econ.* 15, 677–687.
- Sen, A., 1985. *Commodities and Capabilities*. North-Holland, Amsterdam.
- Sen, A., 1980. Equality of what?, in: McMurrin, S. (Ed.), *The Tanner Lectures on Human Values*. Cambridge University Press, Cambridge.
- Sen, A., 1970. *Collective Choice and Social Welfare*. Holden-Day, San Francisco.
- Sendi, P.P., Craig, B. a., Pfluger, D., Gafni, A., Bucher, H.C., 1999. Systematic validation of disease models for pharmacoeconomic evaluations. *J. Eval. Clin. Pract.* 5, 283–295.

- Shiell, A., Donaldson, C., Mitton, C., Currie, G., 2002. Health economic evaluation. *J Epidemiol Community Heal.* 56, 85–88.
- Siebert, U., Alagoz, O., Bayoumi, A.M., Jahn, B., Owens, D.K., Cohen, D.J., Kuntz, K.M., 2012. State-Transition Modeling : A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value Heal.* 15, 812–820.
- Silverman, D., 2000. *Doing Qualitative Research: A Practical Handbook.* SAGE Publications, London.
- Smith, R.D., 2007. The relationship between reliability and size of willingness-to-pay values: a qualitative insight. *Health Econ.* 16, 211–6.
- Snape, D., Spencer, L., 2003. The Foundations of Qualitative Research, in: Richie, J., Lewis, J. (Eds.), *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* SAGE Publications, London, pp. 1–23.
- Sonnenberg, F.A., Beck, J.R., 1993. Markov Models in Medical Decision Making: A Practical Guide. *Med. Decis. Mak.* 13, 322–38.
- Soto, J., 2002. Health Economic Evaluations Using Decision Analytic Modeling: Principles and Practices - Utilization of a Checklist to Their Development and Appraisal. *Int. J. Technol. Assess. Health Care* 18, 94–111.
- Squires, H., 2014. A methodological framework for developing the structure of Public Health economic models. The University of Sheffield.
- Srivastava, A., Thomson, S.B., 2009. Framework Analysis: A Qualitative Methodology for Applied Policy Research. *JOAAG* 4, 72–79.
- Stahl, J.E., 2008. Modelling Methods for Pharmacoeconomics and Health Technology Assessment: An Overview and Guide. *Pharmacoeconomics* 26, 131–148.
- Stake, R., 2003. Strategies of Qualitative Inquiry, in: Denzin, N.K., Lincoln, Y.S. (Eds.), *Strategies of Qualitative Inquiry.* SAGE Publications, Thousand Oaks, pp. 134–164.
- Strauss, A., Corbin, J., 1990. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* SAGE Publications, Newbury Park.
- Sun, X., Faunce, T., 2008. Decision-analytical modelling in health-care economic evaluations. *Eur. J. Heal. Econ.* 9, 313–323.
- Tappenden, P., Chilcott, J., Ward, S., Eggington, S., Hind, D., Hummel, S., 2006. Methodological issues in the economic analysis of cancer treatments. *Eur. J. Cancer* 42, 2867–2875.
- The Research Ethics Guidebook: A resource for social scientists, 2015. Consent [WWW Document]. Online. URL <http://ethicsguidebook.ac.uk/Consent-72> (accessed 6.25.15).

- Torrance, G., Siegel, J.E., Luce, B.R., 1996. Framing and Designing the Cost-Effectiveness Analysis, in: Gold, M.R., Siegel, J.E., Russell, L.B., Weinstein, M.C. (Eds.), *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, pp. 54–79.
- Tsuchiya, A., Williams, A., 2001. Welfare economics and economic evaluation, in: Drummond, M., McGuire, A. (Eds.), *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford University Press, New York, pp. 22–45.
- Van den Haak, M., De Jong, M., Jan Schellens, P., 2003. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behav. Inf. Technol.* 22, 339–351.
- Van Osch, S.M., Stiggelbout, A.M., 2008. The Construction of Standard Gamble Utilities. *Health Econ.* 17, 31–40.
- Van Someran, M.W., Barnard, Y.F., Sandberg, J.A.C., 1994. *The Think Aloud Method: A practical guide to modelling cognitive processes*. Academic press, London.
- Wade, J., Donovan, J.L., Lane, J.A., Neal, D.E., Hamdy, F.C., 2009. It's not just what you say, it's also how you say it: opening the “black box” of informed consent appointments in randomised controlled trials. *Soc. Sci. Med.* 68, 2018–28.
- Weinstein, M.C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C., Luce, B.R., 2003. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices - Modeling Studies. *Value Heal.* 6, 9–17.
- Welte, R., Postma, M., Leidl, R., Kretzschmar, M., 2005. Costs and effects of chlamydial screening: dynamic versus static modeling. *Sex. Transm. Dis.* 32, 474–483.
- Wilkinson, D., Birmingham, P., 2003. *Using Research Instruments: A Guide For Researchers*. RoutledgeFalmer, London.
- Williams, I., Bryan, S., McIver, S., 2007. How should cost-effectiveness analysis be used in health technology coverage decisions? Evidence from the National Institute for Health and Clinical Excellence approach. *J. Heal. Serv. Res. Policy* 12, 73–79.
- Wonderling, D., Reinhold, G., Black, N., 2005. *Introduction to Health Economics*. Open University Press, Maidenhead.
- World Health Organization, 2015. Health topics: Primary health care [WWW Document]. *World Heal. Organ.* URL http://www.who.int/topics/primary_health_care/en/ (accessed 10.13.15).
- Yin, R.K., 2014. *Case Study Research: Design and Methods*, 5th ed. SAGE Publications, Thousand Oaks.
- Zweifel, P., Breyer, F., 1997. *Health Economics*. Oxford University Press, Oxford.