

Bangor University

DOCTOR OF PHILOSOPHY

Quantitative evidence synthesis methods for the assessment of the effectiveness of treatment sequences for clinical and economic decision-making

Lewis, Ruth

Award date:
2019

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**QUANTITATIVE EVIDENCE SYNTHESIS METHODS FOR THE ASSESSMENT OF THE
EFFECTIVENESS OF TREATMENT SEQUENCES FOR CLINICAL AND ECONOMIC DECISION-
MAKING**

Ruth Ann Lewis

A thesis submitted to Bangor University in fulfilment of the requirement for a degree of Doctor of
Philosophy

North Wales Centre for Primary Care Research, School of Health Sciences

Bangor Institute for Health and Medical Research

Bangor University

Submitted 2019

DECLARATION

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

ABSTRACT

The sequential use of alternative treatments for chronic conditions represents a complex intervention; previous treatment, evolving disease, and patient characteristics affect both the choice and effectiveness of subsequent treatments. This thesis develops a new framework for conducting quantitative evidence synthesis of the effectiveness of sequential treatment options within a health technology assessment (HTA) or similar process. It covers methods for developing summary estimates of clinical effectiveness or the clinical inputs for the cost-effectiveness assessment, and can encompass any disease condition. The framework was developed through in-depth evaluation of current approaches using integrated literature reviews.

Key challenges of developing summary effect estimates of interventions conditional on previous treatments were first identified using a HTA of sciatica treatments. Network meta-analyses allowed comparison of multiple treatments, but the limitations of the evidence base, and poor reporting of previous treatments precluded the evaluation of treatment sequences.

A review of NICE guidance identified the type of challenges faced by policy makers and showed that treatment sequencing is pertinent for a wide range of clinical conditions. It also indicated that treatment sequencing was often considered as part of the economic evaluation only, and not the clinical evaluation.

A comprehensive review of quantitative evidence synthesis methods considered:

- i. Meta-analytic methods for assessing the clinical effectiveness of treatment sequences
- ii. Simplifying assumptions made by decision analytic modelling studies in the absence of an adequate evidence base to inform treatment effect estimates conditional on positioning in the sequence
- iii. Decision analytic modelling approaches used to evaluate the effectiveness of treatment sequences

The findings of the review demonstrated that estimating the effectiveness of a sequence of treatments is not straightforward or trivial, and is severely hampered by the limitations of the evidence base. There is no single best way to evaluate treatment sequences, however some approaches could be re-used or adapted, sharing ideas across different disease conditions. Each has advantages and disadvantages, and is influenced by the evidence available,

extent of treatment sequences, and complexity in the decision problem. Due to the scarcity of data, modelling studies applied simplifying assumptions to data on discrete treatments. A coding scheme for all possible assumptions was developed, providing a unique resource to aid the critique of existing models.

The thesis illuminates a significant gap in the methods development. It also demonstrates important limitations in the primary studies, which tends to focus on the evaluation of single treatments with poor reporting of any previous or subsequent treatments. The increasing use of network meta-analysis in HTA demonstrates the acknowledgment that clinical and policy decision making needs to account for the multiple treatments available for many chronic conditions. However, the sequential use of these treatments has yet to be accounted for within the clinical evaluation, with most meta-analysis being conducted of single treatments that may or may not be stratified by line of therapy. The economic modelling exposes the need to consider treatment sequences, but this is often based on the simplifying assumption of treatment independence. The use of simplifying assumptions leads to uncertainty and potential bias in estimating the effectiveness and cost effectiveness of treatments, and can lead to the wrong decision.

In summary, there has been no co-ordinated approach to the important issue of evaluating the effectiveness and cost-effectiveness of treatment sequencing. This is a major shortfall at a time when the cohort of people with complex chronic conditions, requiring sequential treatments, is increasing. The findings of the thesis will help policy makers and researchers gain traction in answering questions about the effectiveness of different treatment sequences.

ACKNOWLEDGEMENTS

I would like to extend my thanks and gratitude to the many people who have helped me along the way in completing this thesis.

My sincere thanks go to my lead supervisors Professor Clare Wilkinson and Professor Dyfrig Hughes for their ongoing academic support and mentorship. I am particularly indebted to Clare for her ongoing enthusiasm, and the opportunities she has provided me throughout my career. My sincere thanks also go to my extended supervisory team, Professor Alex Sutton, Dr Nerys Woolacott, Professor Nefyn Williams, Professor Ceri Phillips, and Francis Ruiz for their continued support and useful comments throughout my PhD. I am especially grateful to Nerys who was also my mentor and Alex for his valuable help in performing the network meta-analysis.

I am especially grateful to my colleagues Maggie Hendry and Alun Surgey, for proof reading my thesis. I am also hugely appreciative to Maggie for her unwavering enthusiasm and support for me to finish my PhD, and for being a good friend throughout.

A special mention goes to Annie Hendry, Barbra France, Emma Jones, Richard Evans, Stefanie Disbeschl and Mathew Jones for their valued help in undertaking various tasks when I needed it, such as finding studies, sorting out references, developing diagrams, and developing the word count.

I would also like to thank the Welsh Assembly Government who funded my NIHR Doctorate Fellowship through Health and Care Research Wales.

Finally, and by no means least, my heartfelt thanks go to Chris, my family and my friends for their ongoing support and immense patience.

TABLE OF CONTENTS

Title page	1
Declaration and consent	2
Abstract	3
Acknowledgments	5
Table of contents	6
List of figures	11
List of tables	13
List of boxes	14
Abbreviations	15
Chapter 1: Aims and background for the thesis	17
1.1 Introduction and chapter overview	17
1.2 Treatment sequencing	18
1.2.1 Treatment sequencing and implications for health care decision-making	18
1.2.2 Factors that influence the effectiveness of individual treatments when used as part of a sequence	19
1.3 The potential limitations of the evidence base for evaluating treatment sequences	22
1.3.1 Randomised controlled trials for informing treatment sequences	22
1.3.2 Randomised controlled trials to inform the introduction of new technologies	23
1.3.3 Non-randomised data as evidence for treatment sequences	24
1.4 Systematic review and quantitative evidence synthesis	25
1.4.1 Meta-analysis	25
1.4.2 Decision analytic modelling	27
1.5 The challenge of applying meta-analysis of the best available evidence to inform treatment sequencing	29
1.5.1 Meta-analysis of treatment sequencing as a complex intervention	29
1.5.2 Applying meta-analysis of discrete treatments to inform treatment sequences	30
1.6 Health technology assessment and clinical guidelines	35
1.7 Personalised medicine	36
1.8 Summary of the problem relating to sequential treatments	38
1.9 Aims and objectives of the thesis	38
1.10 Structure of thesis and summary of contribution	39
1.10.1 Structure of the thesis	39
1.10.2 Summary of contribution	40
1.10.3 My research experience and novel contribution of the thesis	41
1.10.4 Dissemination	42
Chapter 2: A systematic review and network meta-analyses of the clinical effectiveness of different management strategies for sciatica – an example of a health technology assessment	43
2.1 Chapter overview	43
2.2 Background to the health technology assessment	43
2.3 Introduction to the clinical evaluation of sciatica treatments	45
2.4 Methods	46
2.4.1 Review methods	46
2.4.2 Search strategy	46
2.4.3 Study selection and data extraction	46
2.4.4 Outcome measures	48
2.4.5 Statistical analysis	48

2.5	Results	49
	2.5.1 Included studies	49
	2.5.2 Overall response	55
	2.5.3 Pain intensity	56
	2.5.4 Between study heterogeneity, model fit and comparison with standard pairwise meta-analyses	59
2.6	Discussion	59
	2.6.1 Summary of results	59
	2.6.2 Findings of previous reviews	60
	2.6.3 Strengths and weaknesses	60
2.7	Conclusions	62
2.8	Implications of the network meta-analysis of sciatica treatments for methods development in evaluating treatment sequences	63
	2.8.1 Implications of the findings in terms of assessing heterogeneity to inform the need to consider treatment sequencing	63
	2.8.2 Implications relating to the meta-analytical methods used	64
2.9	The next step	67
Chapter 3: A review of quantitative synthesis methods for assessing the effects of treatment sequences		69
3.1	Chapter overview	69
3.2	Introduction	69
3.3	Aim and objectives	69
3.4	Methods	70
	3.4.1 Literature search	70
	3.4.2 Sources searched	71
	3.4.3 Assessing study relevancy	73
	3.4.4 Study classification and data extraction	74
3.5	Quantity of literature included in the review	75
3.6	Chapter summary and next stage	78
Chapter 4: Clinical and economic decision-making regarding treatment sequencing within NICE's health technology appraisal and clinical guideline process		79
4.1	Chapter overview	79
4.2	Summary of the NICE process and how it serves the UK NHS	79
4.3	NICE Methodological guidance on evaluating treatment sequencing	80
4.4	A review of NICE guidance documents that recorded treatment sequencing as a pertinent issue	81
	4.4.1 Objectives	81
	4.4.2 Searches	81
	4.4.3 Overall findings	82
	4.4.4 Clinical conditions for which treatment sequencing was an important consideration for NICE decision making	85
	4.4.5 The evaluation of treatments sequences within clinical guidelines	86
	4.4.6 The evaluation of treatments sequences within the technology appraisals guidance	86
4.5	The omission of NICE single technology appraisals as a potential limitation of the review of modelling studies	87
4.6	Chapter summary and what next	88
Chapter 5: Meta-analytic methods used for evaluating treatment sequences		89
5.1	Chapter overview	89

5.2	Overall findings	89
5.3	Restricting inclusion to treatment sequencing studies	95
	5.3.1 Number of studies	95
	5.3.2 Limiting inclusion to studies of predefined sequences	95
	5.3.3 The inclusion of adaptive treatment sequencing trials	99
	5.3.4 Lessons about studying treatment sequences that have emerged from restricting inclusion to sequencing studies	100
	5.3.5 Lessons learned about the use of observational data	101
5.4	Sub-group and stratified analyses	101
	5.4.1 Number of studies	101
	5.4.2 Sub-group and stratified analysis	102
	5.4.3 Lessons about studying treatment sequences that have emerged from studies using subgroup or stratified analysis	106
5.5	The use of meta-regression	110
	5.5.1 Number of studies	110
	5.5.2 Meta-regression used for evaluating treatment sequences	110
	5.5.3 Meta-regression used to account for the number of previous treatments or disease duration	112
	5.5.4 Meta-regression used for evaluating the influence of subsequent treatments on overall survival	116
	5.5.5 Lessons about studying treatment sequences that have emerged from studies using meta-regression	118
5.6	Using a modifying factor to adjust the effect of single treatments	121
	5.6.1 Number of studies	121
	5.6.2 Methods used to develop reduction factor	121
	5.6.3 Lessons about studying treatment sequences that have emerged from studies using a modifying factor	124
5.7	Ranking treatments according to their absolute effect estimates	125
	5.7.1 Number of studies and description of approach	125
	5.7.2 Lessons about studying treatment sequences that have emerged from studies that rank treatments according to absolute effects	126
5.8.	A case study from the GetReal project: using real world data to link first- and second-line treatments included in same network meta-analysis	126
5.9	Discussion	128
	5.9.1 Summary of the findings for the review of meta-analytic techniques	128
	5.9.2 Relying on sequencing trials	129
	5.9.3 The use of subgroup and stratified analysis	130
	5.9.4 The use of meta-regression	130
	5.9.5 The limitations and challenges of using meta-regression to inform treatment sequences	131
	5.9.6 The use of absolute treatment effects to identify optimal sequence	132
	5.9.7 The use of N-of-1 RCT	132
	5.9.8 The use of individual-patient data	133
	5.9.9 The use of data from non-randomised studies	134
	5.9.10 Recent methodological developments in meta-analysis	138
5.10	Chapter summary and next step	140
	5.10.1 Chapter summary	140
	5.10.2 Next step	140
	Chapter 6: Review of simplifying assumptions applied to effect estimates in modelling studies for evaluating treatment sequences	141
6.1	Chapter overview	141

6.2	Introduction	141
	6.2.1 Decision analytic modelling and decision uncertainty	141
	6.2.2 Methods for handling structural uncertainty	142
	6.2.3 The use of simplifying assumptions in decision analytic models of treatment sequences	142
	6.2.4 Decision problem relating to treatment sequencing	143
	6.2.5 Accounting for treatments administered before and after the decision point of interest	143
	6.2.6 Potential evidence gap and displacement effect	145
6.3	Aim and objectives of the review of simplifying assumptions used in decision analytic modelling studies	145
6.4	Methods for the review of simplifying assumptions used in decision analytic modelling studies	146
	6.4.1 Identifying relevant studies	146
	6.4.2 Coding of the decision problem and simplifying assumptions	146
6.5	Results of the review of simplifying assumptions	148
	6.5.1 Simplifying assumptions coding scheme	148
	6.5.2 Description of included studies	150
	6.5.3 Summary of the simplifying assumptions used	154
6.6	An assessment of the variation in how the data on treatment sequences was selected across modelling studies evaluating a similar decision problem in rheumatoid arthritis	161
	6.6.1 Modelling studies that evaluated sequential TNF-inhibitors	162
	6.6.2 The evidence used to inform sequencing effects of TNF-inhibitors	164
6.7	Discussion	168
	6.7.1 Summary of findings	168
	6.7.2 Previous reviews	170
	6.7.3 Assessing the validity of assumptions relating to sequencing effects and their impact of structural uncertainty	170
	6.7.4 Implications for practice	171
6.8	Next stage	173
Chapter 7 Methodological review of decision analytic modelling approaches used for evaluating treatment sequences		174
7.1	Chapter overview	174
7.2	Developing criteria for assessing modelling approaches used for evaluating treatment sequences	174
	7.2.1 Existing guidance on selecting an appropriate modelling technique	174
	7.2.2 Criteria used for categorising and assessing included modelling approaches	176
7.3	Modelling approaches used for evaluating treatment sequences	177
	7.3.1 Summary of modelling approaches used by included studies	177
	7.3.2 Decision tree	181
	7.3.3. State transition (Markov) cohort model	189
	7.3.4 Partitioned survival models	205
	7.3.5 Semi-Markov cohort process model	207
	7.3.6 Individual patient simulation	211
	7.3.7 Population based models or open cohort models	229
7.4	Discussion	231
	7.4.1 Summary of included studies	231
	7.4.2 Advantages and disadvantages of different modelling approaches	233
	7.4.3 Summary of software and implementation issues	244
	7.4.4 External validation of a sequential treatment model	245
	7.4.5 Structural complexity of the modelling approach	245

7.4.6	Using a conceptual framework to inform the structural complexity of the model	247
7.4.7	Developing a flexible model that can be re-used or further developed	248
7.4.8	Potential decision analytic modelling approaches not covered by included studies	249
7.4.9	Comprehensive decision analytic modelling	250
7.5	Chapter summary and next step	251
7.5.1	Chapter summary	251
7.5.2	Next step	252
Chapter 8: Framework for evaluating treatment sequences		253
8.1	Chapter summary	253
8.2	Definition and scope of the framework	253
8.3	The framework to inform the evaluation of treatment sequences	254
8.3.1	Framework component 1: Framework to guide practice in the conduct of health technology assessment	255
8.3.2	Framework component 2: List of references to related methodological development	270
8.3.3	Framework component 3: Coding scheme for simplifying assumptions used to represent treatment sequencing effects	273
8.4	Future development of the framework	273
8.5	Chapter summary	274
Chapter 9: Discussion and conclusions		275
9.1	Chapter overview	275
9.2	Summary of the main findings	275
9.3	Strengths and limitations of the research	279
9.4	Recommendations for practice and future research	280
9.4.1	Recommendations for policy and practice regarding health technology assessment of treatment sequences	280
9.4.2	Assessing the need to account for treatment sequencing effects and implications for future research	282
9.4.3	Recommendations for practice and research for decision modelling and the use of simplifying assumptions	282
9.4.4	Recommendations for practice and future research for meta-analytic methods	283
9.4.5	Recommendations for future practice and research regarding primary studies	285
9.4.6	Recommendations for future practice and research regarding patient registries	287
9.5	Conclusions	289
References		290
Appendix		313
	Volume I	314
	Volume II	384

LIST OF FIGURES

Figure 1.1: Example of a comparison of two treatment sequences

Figure 1.2: An illustration of two treatment sequences, (A,B) and (B,A), where the overall response rate was the same for both in scenario 1, and in scenario 2 favoured the B,A sequence due to the differential response rates in second-line treatments

Figure 1.3: Illustration of the type of evidence included in a meta-analysis of first-line treatments X, Y and Z

Figure 1.4: Illustration of the type of evidence included in a meta-analysis of third-line treatments X, Y and Z

Figure 1.5: Illustration of the type of evidence included in a meta-analysis comparing treatments X, Y, and Z where line of treatment is not considered

Figure 2.1: Flow diagram showing the number of references identified, publications retrieved for assessment, and studies included in the review

Figure 2.3: Network of treatment strategies for sciatica for comparative studies reporting global effect

Figure 2.3: Network of treatment strategies for sciatica for comparative studies reporting pain intensity

Figure 2.4: Plot of the odds ratios of global effects for different treatment strategies compared with inactive control from the network meta-analysis

Figure 2.5: Plot of the weighted mean difference for pain intensity for different treatment strategies compared with inactive control from the network meta-analysis

Figure 3.1: Flow diagram showing the number of references identified, publications retrieved for assessment, and studies included in the methodology review

Figure 6.1: Illustration of treatment sequences according to decision problem

Figure 6.2: Coding scheme used for coding simplifying assumptions relating to treatment sequencing effects used by included studies

Figure 7.1: A schematic diagram of the Dranitsaris model: a decision tree model for the treatment of metastatic colorectal cancer

Figure 7.2: A schematic diagram of the Cameron model: a Markov cohort model for the treatment of advanced breast cancer

Figure 7.3: A schematic diagram of the Heeg model: a response-based Markov cohort model for the treatment of multiple myeloma

Figure 7.4: A schematic diagram of the Lee model: a Markov cohort model for the treatment of recurrent ovarian cancer

Figure 7.5: A schematic diagram of the Maetzel model: a Markov cohort model for the treatment of rheumatoid arthritis

Figure 7.6: A schematic diagram of the Welsing model: a Markov cohort model for the treatment of rheumatoid arthritis

Figure 7.7: A schematic diagram of the Diamantpoulus model: a state transition individual patient simulation model for the treatment of rheumatoid arthritis

Figure 7.8: A schematic diagram of the Birmingham rheumatoid arthritis model (BRAM): a discrete event simulation model for the treatment of rheumatoid arthritis

Figure 8.1: A schematic of the categorisation scheme for practice recommendations

Figure 8.2: List of references relating to recent methodological development referred to in the framework

LIST OF TABLES

Table 2.1: Treatment categorisation

Table 2.2: Results (odds ratios, with 95% confidence intervals/credible intervals) of the network meta-analysis for global effect

Table 2.3: Results (weighted mean difference, with 95% confidence intervals/credible intervals) of the network meta-analysis for pain intensity

Table 4.1: Reasons why NICE guidance were identified by the NICE website search

Table 4.2: Technology appraisals and clinical guidelines identified during a search of the NICE website

Table 4.3: Technology appraisals included in the methodology review that were not identified during the NICE website search

Table 5.1: Primary aim of the included studies evaluating meta-analytic approaches for evaluating treatment sequences

Table 5.2: Overview of the meta-analytic approaches used by included studies

Table 5.3: Summary of included rheumatoid arthritis studies using subgroup-analysis

Table 5.4: Summary of included rheumatoid arthritis studies using regression methods

Table 6.1: Summary and coding of rheumatology studies

Table 6.2: Summary and coding of non-rheumatology studies

Table 6.3: Summary of the frequency of use of the simplifying assumptions

Table 6.4: The decision problem and treatment sequences evaluated by included modelling studies that considered sequential TNF-inhibitors

Table 7.1: Modelling approaches used by rheumatology studies

Table 7.2: Modelling approaches used by non-rheumatology

Table 7.3: Summary of the different modelling approaches used and their advantages and disadvantages for evaluating treatment sequences

Table 8.1: Framework for evaluating treatment sequences: recommendations for practice

Table 8.2: Advantages and disadvantages of current modelling approaches used for evaluating treatment sequences

Table 8.3: Coding scheme for simplifying assumptions relating to treatment sequencing effects

LIST OF BOXES

Box 1: Example illustrating the potential implicit assumptions made regarding sequencing effects when applying effect estimates taken from studies of single treatments

ABBREVIATIONS

ACR	American College of Rheumatology response
AHRQ	Agency for Healthcare Research and Quality
BRAM	Birmingham Rheumatoid Arthritis Model
BSRBR	British Society for Rheumatology Biologics Register
CADTH	Canadian Agency for Drugs and Technologies in Health
CBT	Cognitive-behavioural therapy
CER	Comparative effectiveness research
CG	Clinical guideline
DAS28	Disease Activity Score 28 joints
DES	Discrete event simulation
DMARD	Disease-modifying anti-rheumatic drug
DMARD-IR	Inadequate response to previous conventional DMARD
DSU	Decision Support Unit
EU	European Union
FU	Fluorouracil / leucovorin
HAQ	Health Assessment Questionnaire
HIV	Immunodeficiency virus
HRQL	Health related quality of life
HTA	Health technology assessment
IPD	Individual patient data
ISPOR-SMDM	International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making
mTORi	Mammalian target of rapamycin inhibitor
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NRTIs	Nucleoside reverse transcriptase inhibitors
NSAIDs	Non-steroidal anti-inflammatory drugs
NSCLC	Non-small cell lung cancer
ORs	Odds ratios
OS	Overall survival
PD	progressive disease
PFS	Progression-free survival
QALYs	Quality adjusted life years gained
Q-RCT	Quasi-RCT
RCT	Randomised controlled trial
SCQM	Swiss Clinical Quality Management in Rheumatic Diseases

SMART	Sequential multiple assignment randomised trial
SSATG	Southern Swedish Arthritis Treatment Group
SSRI	Selective serotonin reuptake inhibitor
TA	Technology assessment
TNF	Tumour necrosis factor alpha
TNF-IR	Inadequate response to TNF-inhibitors
TTNT	Time to next treatment
TTP	Time to progression
UKPDS	UK Prospective Diabetes Study
VAS	Visual Analogue Scale
VEGF TKI	Vascular endothelial growth factor receptor-tyrosine kinase inhibitor
WMD	Weighted mean difference

CHAPTER 1: AIMS AND BACKGROUND FOR THE THESIS

1.1 INTRODUCTION AND CHAPTER OVERVIEW

Good evidence of what treatment works, for which patient, under what circumstances is vital for policy and practice decision-making. An evaluation of the added value of new treatments is crucial for cost-constrained healthcare systems, such as the UK National Health Service (NHS). This thesis will contribute to the development of methods to achieve this. It describes the development of a framework for informing the use of quantitative evidence synthesis methods within the context of a health technology assessment, systematic review or decision analysis, in order to estimate the effectiveness of treatment sequences for decision-making.

Treatment sequences relate to the order in which interventions are administered within the treatment or care pathway for a specific condition. Systematic review and evidence synthesis is a specific methodology that encompasses searching for, appraising and synthesising findings of primary studies using transparent and objective processes in order to minimise bias and reduce chance findings (random effects).¹ This method has rapidly become a cornerstone of the evidence-based practice and policy movement.² Decision analysis is a quantitative approach used for assessing the relative value, in terms of both benefit and risk, with or without costs, of different decision options.³ Decision analyses can be used to decide how to manage an individual patient, formulate policy recommendations about groups of similar patients or, in the form of decision aids, help individuals make decisions about therapies.⁴

This chapter describes the nature of 'sequential treatments'; and why they are important when evaluating the effectiveness of interventions for clinical and policy decision-making. It describes the limitations of the evidence base for evaluating the effectiveness of treatment sequences and how this impacts healthcare decision-making. It introduces meta-analysis and decision analytic modelling as important quantitative evidence synthesis methods to support evidence-based clinical and policy decision-making. It also outlines some of the challenges faced by healthcare decision-making organisations, such as the National Institute for Health and Care Excellence (NICE) in the UK, regarding the introduction of new treatments within a current treatment pathway, or identifying the optimum sequence of treatments to use in practice. It provides a summary of the research problem and outlines the overall aims and objectives of the thesis. It sums up the structure of the thesis and ends with an outline of my relevant research experience and the proposed novel contribution of the thesis.

1.2 TREATMENT SEQUENCING

1.2.1 Treatment sequencing and implications for healthcare decision-making

The availability of multiple interventions for the same condition or indication is increasingly common;⁵ when the current treatment is no longer effective or the condition progresses, further treatments are available. The same is true in the case of adverse effects, intolerance to specific treatments, or where interventions are contraindicated. Treatments that are only partially effective can be progressively intensified or enhanced by increasing the dose, or adding new treatments to produce combination therapies. Hence, for many conditions, in order to find and maintain an effective treatment regimen, a sequence of treatments is likely to be used. This is especially true in the treatment of chronic diseases, such as heart disease, cancer, depression, and diabetes, which are 'slow in progression, long in duration, and devoid of spontaneous resolution'.⁶ However, despite the availability of multiple treatments, effective and safe control of disease activity is not always easy to achieve and there is usually no standard therapeutic approach.^{7 8} Furthermore, patients suffering from chronic diseases often have multiple conditions or co-morbidities which are likely to affect treatment choice.

As the number of therapeutic options increases, treatment decisions become increasingly complex. Clinical decisions on when to use an intervention in the course of a patient's illness or care pathway are dependent on a number of interrelated treatment, patient, and disease characteristics including prior treatment history and response to previous treatments. The disease process for many conditions tends to wax and wane, with intermittent acute or exacerbated stages accompanied by an overall gradual progression. Treatments that are effective early on in the disease process might not be effective later. There may be intervals where a more intensive treatment with potential adverse effects is required to combat more serious symptoms or reduce the risk of relapse or recurrence. At other times a less intensive, and less burdensome, treatment is sufficient. The sequencing of treatments is also important for some infectious diseases where treatment resistance can become an issue, for example human immunodeficiency virus (HIV). This results in the need for repeated clinical decisions on when to change treatment, or treatment intensity, and to what type. For example, if a new treatment results in symptom-reduction, but not elimination:⁹

- i. Should a new treatment be used at the risk of losing the benefit already achieved?
- ii. Should the current treatment dose be increased, or another treatment added at the risk of increasing side effects?
- iii. Should the condition just be monitored, with the hope of a gradual improvement?

In clinical practice, treatment sequences tend to be established in a 'trial and error' fashion until a response is observed⁹ or using the 'play-the-winner-drop-the-loser' algorithm.¹⁰ This allows treatment sequences to be tailored to individual patients. However, decisions on which treatment to use first, or the optimum sequence, may not be straightforward. This is explored further in Section 1.2.2.

Treatment-sequencing choices may be dependent on the overall aim of the treatment, for example, whether it is intended to provide benefit beyond the treatment period (e.g. in rheumatoid arthritis) or only whilst the treatment is being used (e.g. neuropathic pain). They may also be influenced by the

type of treatment being considered, for example some treatments, such as antiviral therapy, may be associated with drug resistance or poor adherence.

In some circumstances, upfront treatment-sequencing decisions on the overall management approach may be required. Examples include the following:

- i. Whether to use a step-up approach, starting with less effective treatments that may be associated with fewer adverse effects and possibly cheaper, and then incrementally adding more potent agents or increasing the dose as required
- ii. Whether to use a step-down approach, starting with the most effective treatment, which may be more expensive and carry a greater risk of serious adverse effects, and then gradually 'climbing down the therapeutic ladder'

For progressive conditions such as rheumatoid arthritis, the second option might modify the course of the disease sufficiently to mean that, over a lifetime, the resulting quality of life outcomes are better than those obtained with the first option. This is explored in more detail in the Appendix Volume I (Section C3).

Treatment-sequencing decisions should be based on the best available evidence of effectiveness. However, developing the evidence base to inform such decisions is not straightforward. Research conducted to inform clinical effectiveness tends to concentrate on the evaluation of discrete treatments used at a single point in the treatment pathway, rather than a sequence of treatments. This does not account for the complexities and dynamic nature of factors that are likely to impact the choice and effectiveness of each treatment used within a sequence, which can also affect the optimum ordering of these treatments, when considering the effectiveness of the whole sequence. Furthermore, research evidence is used to inform clinical practice in general, and therefore targets the majority of the patient population rather than informing decision-making for the individual. The importance of individual factors is likely to be heightened when it comes to decision-making around treatment sequences. The tension between individually based decision-making and policy decisions in the context of treatment sequences is discussed in Section 1.7 (Personalised medicine). The thesis, however, focuses on developing the evidence base to inform decision-making on treatment sequencing for the majority.

1.2.2 Factors that influence the effectiveness of individual treatments when used as part of a sequence

The effectiveness of specific interventions is affected by the order in which they are used within a treatment sequence, with the performance of individual treatments being conditional, or dependant, on the previous treatments used.¹¹⁻¹⁴ The effectiveness of a specific treatment, when considered in isolation, may not provide a good indication for its use as the first treatment within a planned sequence, where some treatments can have an impact or delayed effect on future treatments.^{15 16}

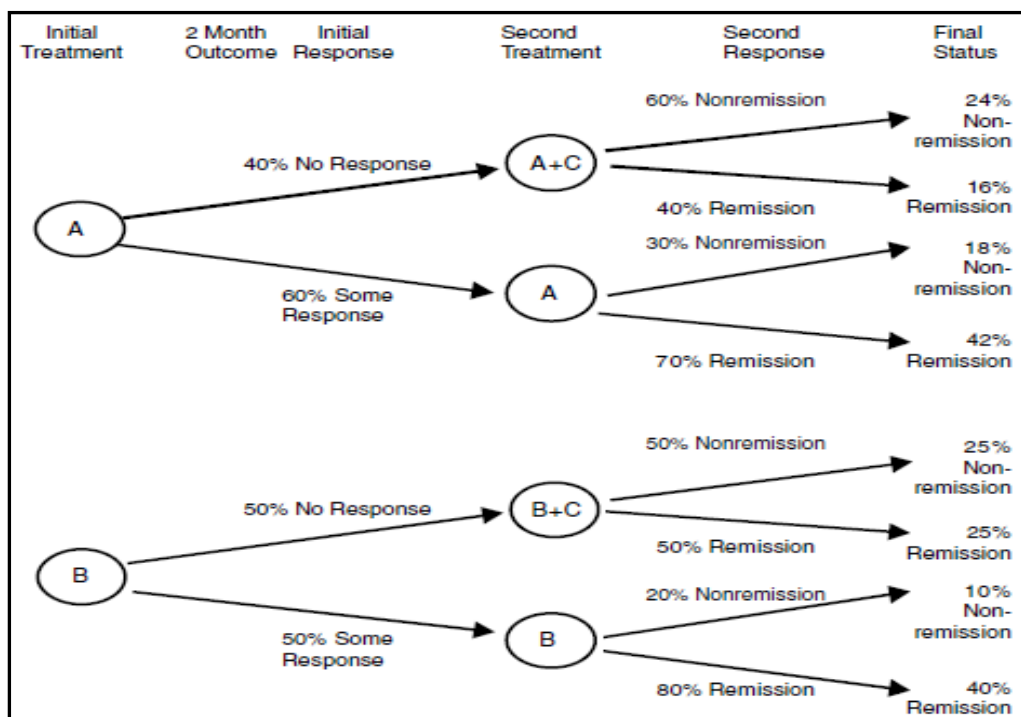
A number of factors are likely to impact the effectiveness of individual treatments used as subsequent treatments. The treatment effect may, for example, be dependent on changes in pathophysiology over time, the disease condition left by the previous treatments, the residual effect left by previous treatment, immunogenicity, and drug kinetics.¹³ The reason for treatment discontinuation is also likely to have differential effects on the effectiveness of subsequent treatments. For example, when a treatment is discontinued due to inadequate response or loss of responsiveness over time, the effectiveness of the next treatment is likely to be reduced. However, changing treatment due to intolerance or adverse effects may not have an impact the subsequent treatment's effect.^{17 18} The adverse event profiles of initial treatments can also have a differential impact on patients' willingness or ability to adhere subsequent treatments.¹⁹

Treatments for chronic conditions tend to become less effective over time, as a patient progresses through their disease journey. For example in rheumatoid arthritis, a progressive disease, a second tumour necrosis factor alpha (TNF)-inhibitor, which is an important class of biological agents, has been shown to more effective in patients who have not previously received a TNF-inhibitor than those who have.^{12 20 21} Similarly, the use of conventional disease modifying anti-rheumatic drugs (DMARDs), which are generally tried before biological therapy, have also been shown to be more effective when only non-steroidal anti-inflammatory drugs (NSAIDs) have previously been used and less effective following another DMARD.¹¹ However, it is not axiomatic that treatment effect declines with successive treatments. Dimopoulous *et al.* found second-line antibiotics for treating acute exacerbations of chronic bronchitis to be more effective, but not less safe, than first-line.²² It was concluded that, given the increasing resistance to older antimicrobial agents within this patient population, treatment with selected second or third generation antibiotics may be preferable. However, the available data did not allow for stratified analysis according to the presence of risk factors for poor outcomes. Early research findings show that the presence of antidrug antibodies to the first TNF-inhibitor for rheumatoid arthritis is associated with a better clinical response to a second.²³ Exposure to some biologically targeted therapies in oncology can also sensitise the tumour to subsequent treatments.²⁴ Overall, it is therefore not trivial or intuitive to optimise a treatment sequence from the start.¹⁵

The theoretical case for starting with a suboptimal treatment, rather than the most effective treatment, because it may lay down a better foundation or enhance the effect of subsequent treatments is illustrated in an example presented by Murphy *et al.* shown here in Figure 1.1.⁹ It considers two treatments, A and B, that differ in terms of immediate response, favouring A. When these are used within a sequence, where the initial treatment is augmented with treatment C when a patient does not respond, then the long-term response during the entire period for the sequence starting with treatment B may exceed the effect of A followed by A augmented by C. The sequence beginning with A has an overall remission rate of four months of 58% (16 + 42%) whereas the sequence beginning with B has an overall remission rate of 65% (25 + 40%). This means that treatment A is best if as a standalone treatment, but treatment B is the best initial treatment as part of a sequence. This is due to the better

synergistic effect of B and C for non-responders to treatment B and the greater likelihood of those who had an initial partial response progressing to remission whilst remaining on treatment B. This issue is also referred to as delayed effect.¹⁶ For example, consider addiction management for alcohol dependency, which can be viewed as a chronic relapsing condition.²⁵ A clinician may prescribe either an opiate antagonist (naltrexone) or cognitive-behavioural therapy (CBT) as the initial treatment. Patients who could then go on to receive telephone monitoring or telephone monitoring plus counselling to prevent relapse. The subsequent telephone counselling may be more effective for CBT responders, who have already experienced a 'talking therapy' and are better prepared to take advantage of counselling, than for responders to naltrexone who have no such experience.¹⁶ Thus, even if CBT and naltrexone result in the same proportion of responders, or if CBT appears less effective than naltrexone, CBT may be the best initial treatment when considered as part of a sequence.¹⁶ This example is also used in Appendix Volume I Section B2.

Figure 1.1: Example of a comparison of two treatment sequences



Taken from: Murphy, S.A., Oslin, D.W., Rush, A.J., Zhu, J. (2007) Methodological challenges in constructing effective treatment sequences for chronic psychiatric disorders. *Neuropsychopharmacology*, 32, 257–262.

Three reasons are proposed why studies focusing on a single point in the treatment pathway provide poor evidence on treatment sequences:^{15 16 25 26}

1. *Delayed therapeutic effects or sequential treatment interaction*

i) Positive synergies:

Treatment A may not appear best initially but may have enhanced long-term effectiveness when followed by a particular maintenance treatment.

ii) Negative synergies:

Treatment A may produce a higher proportion of responders but also result in burdensome side effects that may reduce the willingness of non-responders adhere to subsequent treatments.

2. *Prescriptive (diagnostic) effect*

Treatment A may produce fewer responders than treatment B initially, but treatment A may elicit symptoms that allow better matching of the subsequent treatment to the patient and thus achieve improved response to the sequence of treatments.

3. *Cohort (selection) effect*

Participants who enrol and are adherent in a clinical trial of a single stage treatment may be quite different from subjects in a trial of treatment sequences where alternative, subsequent treatment options are available.^{15 16 25 26} Conversely, in a sequential study, the decision to end first-line treatments may be influenced by the knowledge there is a second-line treatment readily available.¹⁹

1.3 THE POTENTIAL LIMITATIONS OF THE EVIDENCE BASE FOR EVALUATING TREATMENT SEQUENCES

1.3.1 Randomised controlled trials for informing treatment sequences

The randomised controlled trial (RCT), when properly designed and implemented, is widely considered the “gold standard” for assessing the efficacy or effectiveness of healthcare interventions. The underpinning reasons for this contention is discussed in more detail in the Appendix Volume I (Section A).

Conventional RCTs tend to evaluate single treatments. They rarely provide long-term data on disease trajectories, or address questions about whether, or when specific treatments should be used sequentially.⁹ The challenge is that, even when RCTs of treatment sequences are available, they are unlikely to be sufficient to inform decision-making. The number of available treatments continues to increase in most cases, and consequently the number of unique sequences will increase geometrically.^{27 28} It is therefore both impractical and prohibitively costly to evaluate all conceivable treatment sequences in RCTs. The number of possible sequencing scenarios used in clinical practice is likely to outnumber the volume of trials conducted. Individual genetic makeup or biologic characteristics will increasingly determine the suitability of new treatments.²⁹ This means that personalised medicine will evolve; I re-visit this in Section 1.7.

Rather than RCTs of whole treatment pathways (sequences) it is more usual to have RCTs comparing individual interventions used at specific points in the treatment pathway, for example second- or third-line treatments, for informing treatment sequencing decision-making. However, although these RCTs potentially include participants who are at the same point in the treatment pathway, the specific prior treatments received by participants are likely to differ. Furthermore, as discussed in Section 1.2, the effect of the overall sequence cannot be accurately estimated by

evaluating a single treatment episode, and choosing the best initial treatment on the basis of an RCT of first-line treatments is suboptimal.⁹ I re-visit the implications of relying on RCTs for treatments used at a specific point in the pathway within the evidence synthesis in Section 1.5. Of note here is that when interventions are no longer effective or cause adverse effects, patients generally cross over to the alternative treatment, revert to current clinical practice, or drop out. A number of methodological studies provide advice on how best to adjust for the cross-over effect when analysing the data from RCTs of single treatments (also discussed in Appendix Volume I, Section A4),³⁰ but these methods do not consider the clinical effectiveness of treatment sequences, or deal with secondary research. This also means that where RCTs with long-term follow up do exist, they may be subject to the same confounding and biases as observational studies due to non-adherence and loss to follow-up. This is discussed further in Section 1.3.3 and the Appendix Volume I (Section A4).

1.3.2 Randomised controlled trials to inform the introduction of new technologies

A common question faced by policy makers and reimbursement agencies is the optimal positioning of a new technology within the current treatment pathway. The nature and availability of the evidence base to inform such decisions is often very limited and tends to be an artefact of the clinical trial design and licencing process.³¹ In most cases, a new treatment is only tested and licensed for a single point in a treatment pathway, so there is little evidence for how well they might perform at a different point in the sequence even though they might do well, or even better than a current treatment.

The clinical trials of new treatments conducted for licencing purposes frequently limit inclusion to patients who are intolerant to, or who have failed to respond to the 'first-in-class' drug, or all available active treatments, and are therefore receiving best supportive care. This type of trial design is accepted by regulators, who grant a licence on this basis. The role of the regulatory authorities, such as The European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) is to assess the efficacy, safety, and quality of the drug.³² They allow companies to sell their products, and ideally want to grant approval to as many new and safe drugs as possible in order to increase the availability of treatments on the market. The new drug may actually be as good as the first-in-class drug, but manufacturers generally consider it too much of a risk to make this comparison. This discrepancy between what is accepted by the regulatory authorities for licencing purposes, and what is needed by policy decision makers (e.g. NICE) and health technology assessment agencies leads to important gaps in the evidence base, especially for informing the optimum sequencing or positioning of new treatments. It can also lead to the available drug sequences being dictated by when and how individual drugs were introduced into the market rather than their optimum positioning based on effectiveness. For example, the introduction of a new drug for advanced cancer, compared in a phase III trial to best supportive care in participants who had failed to respond to initial chemotherapy treatment. Other second-line treatment may be available for treating the same cancer. The policy decision question about placing the new drug in established therapy then becomes one about whether it should be second-line or reserved for third-line. The evidence base for answering this

question would have been limited, but the question of first-line use, and optimal overall sequence would not have even been considered. Some treatments however, do get studied and licenced at more than one point. For example, both targeted therapies, crizotinib and ceritinib, for anaplastic lymphoma kinase (ALK)-positive Non-small cell lung cancer (NSCLC) were first studied, licensed, and NICE approved for second-line treatment and then subsequently studied, licensed, and NICE approved as first-line-treatment.

The access to many new targeted therapies for conditions such as cancer are determined by gene expression and receptor status. These drugs are usually very expensive, especially when including the cost of testing. These agents act at the molecular level, which means that the limitations of the clinical trials to inform decision-making is further complicated by the fact that some genetic mutations, which predict treatment response, are not identified until the analyses of the pivotal trial or in some cases after their introduction into practice.³⁷ This means that the evidence to inform clinical practice is further limited to subgroup analyses, which is no longer protected by randomisation, and subject to the limitation of interpreting subgroup data. A summary of the recent developments in trial design for evaluating these new targeted therapies is provided in the Appendix Volume I (Section B).

1.3.3 Non-randomised data as evidence for treatment sequences

Non-randomised studies are an alternative source of evidence to inform treatment sequencing decisions. Examples include observational studies based on patient registries or non-controlled long-term phase IV follow-up studies. Non-randomised studies are likely to be more readily available and provide timelier data than prospective RCTs.^{38 39} Real-world observational studies are also more likely to reflect treatment sequences used in clinical practice and include patients generally excluded from RCTs. However, treatment-effect estimates derived from non-randomised studies are at greater risk of bias and confounding than those taken from RCTs.⁴⁰ The generic advantages and disadvantages of non-randomised studies is discussed in more detail in the Appendix Volume I (Section A).

Real-world data from sources such as disease-specific registries could provide crucial information on treatment sequencing. The sequential use of TNF-inhibitors for inflammatory arthritis provides a useful example. In the absence of RCTs evaluating the use of a second-line TNF-inhibitor for inflammatory arthritis, a number of observational studies were based on patient registry data. at a time when only first generation TNF-inhibitors were available, Hyrich *et al.* (2007)⁴¹ reported that designing a randomised experiment for patients to receive a second TNF-inhibitor on the basis of inefficacy or toxicity would present considerable practical and ethical difficulties. More recently, the first RCT (EXXELERATE study) to directly compare two different TNF-inhibitors for the treatment of rheumatoid arthritis, published in 2016, allowed non-responders to switch to the alternative TNF-inhibitor.⁴² However, the RCT compared the use of a second generation TNF-inhibitor (certolizumab pegol) with a first generation TNF-inhibitor (adalimumab). Whilst not without their limitations, non-randomised studies may therefore, be the best data obtainable.¹⁷

With the increasing popularity of analysing real-world observational data (also referred to as big data) and the use of linked databases, the availability of this type of data is likely to improve.^{43 44} The validity and quality of routinely-collected data is key here. Furthermore, selection bias or confounding caused by an imbalance in prognostic factors between intervention and control groups, is a concern even in rigorously conducted observational studies.^{43 45} I come back to this in Chapter 9 (Section 9.4.6), and to specific limitations of using data from non-randomised studies to inform treatment sequencing effects in Chapter 5.

1.4 SYSTEMATIC REVIEW AND QUANTITATIVE EVIDENCE SYNTHESIS

Evidence synthesis is the process of bringing together the results of individual research studies in order to better map the knowledge base.⁴⁶ It is based on the principle that science is cumulative.⁴⁷ The term is used for any method used to combine the results of studies including quantitative,⁴⁸ qualitative,^{2 49} and narrative methods.⁵⁰ This thesis focuses on the use of quantitative evidence synthesis, and includes both meta-analysis methods and decision analytic modelling.

1.4.1 Meta-analysis

1.4.1.1 Meta-analysis for informing decision making

Meta-analysis, when conducted as part of a systematic literature review, is an essential part of evidence-based medicine, and refers to a collection of statistical methods and techniques used to synthesise the results from several independent studies, generally with the aim of producing a single estimate of treatment effect, or explore variation in their findings.⁴⁸ It enhances the precision, or statistical power, of the summary estimates of treatment effects, and the assessment of inconsistency of effects between studies enables a better understanding of moderator variables, boundary conditions, and generalisability.⁵ Meta-analysis aims to resolve uncertainty in the decision process, which relates to what conclusions to draw from a body of research studies on the same topic.⁴

Conventional pairwise meta-analysis estimates the effects between two alternative treatments using a single outcome measure. However, effective policy making needs to be underpinned by a broader and more complex evidence base, which is likely to require a more sophisticated form of evidence synthesis. Clinical and policy decision-making needs to consider all available interventions for a particular condition, including multiple compounds from the same pharmacological class, and compounds from different classes used for the same indication.⁵¹ They also need to incorporate a diverse evidence base due to, among other things, the lack of primary data from head-to-head trials for a number of comparators, information for relevant subpopulations, and representation of real-world factors such as variable adherence.^{52 53} The assessment of the evidence base is therefore likely to require the following:

- i. Simultaneous comparison of multiple treatment options
- ii. Include heterogeneous studies with differing baseline risks
- iii. Include evidence from multiple sources of variable quality

- iv. Include long-term effects, where only RCTs of efficiency with short-term follow-up are available
- v. Incorporate outcome data reported in various ways by different studies

Recent developments in meta-analysis include the use of multi-parameter methods which capture further complexities in the evidence base by facilitating the utilisation and integration of diverse sources of evidence.^{54 55} Multi-parameter evidence synthesis is a generalization of meta-analysis in which several parameters are estimated jointly.^{54 55} Network meta-analysis, which is increasingly used within health technology assessment, is an important class of multi-parameter synthesis models.^{56 57} Further methods to enhance and complement conventional meta-analysis are continually being developed,^{58 59} for example methods that allow for the inclusion of multiple endpoints or observations over time,⁶⁰⁻⁶³ bias modelling in order to synthesise data from different study designs,^{38 64-66} statistical methods for combining individual patient data and aggregate data in the same meta-analysis,⁶⁷⁻⁷¹ and methods that can allow the relaxation of the main assumption of similarity (transitivity) within a network meta-analysis.^{72 73} A number of these methods are grounded in Bayesian statistical theory, and implemented using Markov chain Monte Carlo (MCMC) simulation methods. An example of a network meta-analysis conducted within the Bayesian Framework is presented in Chapter 2. This includes a more detailed description of these methods.

The decision framework for treatment sequencing is more complex than the decision framework for a single treatment. The evidence base for sequencing is likely to be seriously limited in many instances. The recent developments in meta-analytical methods that allow the inclusion of broad and diverse evidence required to inform decision-making will be especially important. The findings of a review of meta-analytic methods for evaluating treatment sequences is presented in Chapter 5.

1.4.1.2 Meta-analysis of individual patient data

Meta-analytic methods that utilise individual patient data may be particularly useful for developing treatment sequencing effects. Meta-analysis is usually based on aggregate data obtained from study publications. In a meta-analysis of individual participant data, the synthesis is based on the original participant data from the relevant studies. This relies on extensive collaboration between researchers, as the trial groups must share their data with the reviewers. There has been some progress on wider access to individual patient-level data, however there is a long way to go before this is routinely available for most studies.⁷⁴⁻⁷⁷ There is growing recognition of the imperative for data-sharing.⁷⁸ Many funders now require that data arising from their grants is shared.⁷⁹ However, important impediments to data-sharing include concerns that patient confidentiality and consent may be breached,⁸⁰ a lack of trust between researchers and the community they share,⁷⁸ and the fact that providing access to the data requires considerable time and effort.⁷⁸ Where individual patient-level data is not available for all trials, there is a need to incorporate both aggregate and individual patient data within the same meta-analysis.⁶⁷ I come back to the use of individual patient data in meta-analysis in Chapter 5, Section 5.9.8.

Systematic reviews and meta-analyses aim to provide the best estimate of the true 'average' population effect, and do not apply to the individual, even when the meta-analysis is based on individual patient data. This clearly limits their application to individual decision-making.

1.4.2 Decision analytic modelling

Decision analytic modelling provides an alternative framework for the synthesis of the broad and complex evidence base required to support decision-making. It is especially useful where the evidence base is limited. However, unlike meta-analysis, which is used to obtain the best estimate of the true treatment effect in a particular population, decision analytic modelling is used to obtain a better informed and rational decision in the face of uncertain or incomplete information.⁸¹ A more detailed description of decision analytic modelling is provided in Chapters 6 and 7.

1.4.2.1 Decision modelling for health economic evaluation

Decision analytic models are often used as part of an economic evaluation to simultaneously compare the expected consequences of pursuing different strategies, or sequences of clinical decisions, and quantify the extent of the uncertainty involved.⁸² Statistical evidence synthesis techniques and decision analytic models provide an ideal mechanism to structure the decision problem, combine all available data, and characterise the various sources of uncertainty associated with the decision problem.⁸³ The usefulness and validity of the results obtained from the economic model is dependent on the suitability of the model structure, the quality of the data inputs, and the methods used to derive these.^{83 84} The evidence base informing the model parameters is generally derived from multiple sources, including clinical trials, observational studies, administrative databases, expert opinion, and secondary analysis (such as meta-analysis).^{83 84} The data used to inform the clinical effectiveness parameters should be based on a systematic review and meta-analysis where feasible.⁸³⁻⁸⁵ However, the model parameters can also be derived almost exclusively from a single RCT.⁸³

The requirements of decision analytic modelling have placed some important demands on meta-analytic methods, which include:⁸⁶

- i. The need to estimate the effectiveness of interventions despite the absence of head-to-head RCTs. This includes the use of network meta-analysis
- ii. The need to obtain probabilities of clinical events for models over a standardised follow-up period, despite the available research studies reporting data for various follow-up periods
- iii. The need for estimates of treatment effectiveness for a common endpoint, despite available research studies reporting data for various outcome measures
- iv. The need to assess heterogeneity in measures between different types of patients

I re-visit these issues in Chapter 2.

Further developments for evaluating the evidence to inform decision-making include the use of comprehensive decision modelling, which incorporates both evidence synthesis and economic decision modelling in the same coherent framework.⁸⁷⁻⁸⁹ This enables all the evidence, uncertainties

and correlations to be captured in the economic analysis.⁸² It has the advantage of facilitating a fuller expression of the uncertainty in the evidence base,⁹⁰ and can make use of the posterior distributions of the relevant parameters directly, making it ideal for implementing probabilistic decision analytic modelling. This approach is currently based on the use of Markov cohort modelling technique, and implemented within the Bayesian framework. I return to this in Chapter 7, Section 7.4.9.

1.4.2.2 Whole disease modelling

Economic models tend to be static, focusing on a single decision-point within the treatment pathway, for example the comparison of second-line treatments, and then mapping the long-term outcomes. Models can also be used to incorporate the projected impact of future treatments.⁹¹ Restricting the scope of the economic analysis to a single decision-point means that other adoption decisions elsewhere in the disease pathway, and their knock-on impacts, are often treated as independent of the decision problem under consideration.⁹² The issue regarding the use of static models may be a particular concern in treatments of communicable diseases where drug resistance is problematic. Here treatment sequencing sometimes means reserving the most efficacious treatment as a last-line of therapy, but the use of inappropriate sequences could lead to selection pressure and increase drug resistance.

Recent developments in healthcare modelling include whole disease pathways or system models.⁹²⁻⁹⁴ This includes the 'whole disease model' based on an individual patient-level simulation modelling technique, the application of which was first demonstrated in colorectal cancer.^{92 93} The model incorporates the whole disease pathway from pre-diagnosis to end stage disease, and includes all the diagnostics, treatment and follow-up interventions.^{92 93} Another example is the Archimedes model, which incorporates the whole system including not only the individual patients but also other important aspects of the healthcare system.^{95 96} The Archimedes model has been developed to evaluate diabetes care, cardiovascular disease, and cancer, and includes its use for the evaluation of prevention and screening, not just treatment.^{95 96} I come back to this in Chapter 7 (Section 7.4.8). A project funded by the Medical Research Council, on behalf of the NICE, investigated the feasibility of using whole disease modelling for informing clinical guidelines, using atrial fibrillation and prostate cancer as case studies.⁹⁴ The project was able to demonstrate feasibility, with successful development of models representing the complicated guideline pathway and disease process being completed. However, barriers to routine adoption included the extensive time and resource required to implement, and the need for specialist expertise in discrete event simulation. The models included treatment sequences, but the available data to inform these were limited, necessitating the need to make simplifying assumptions regarding treatment-sequencing effects. The funding of this project showed that there was recognition and desire for modelling exercises to more robustly account for both the 'upstream' and 'downstream' impact of treatments. The work done as part of this thesis would fit in with this goal. However, this type of whole-pathway modelling approach has not subsequently been adopted within the health technology funding decision process or health economic evaluations in general.^{97 98} There is also a lack of general agreement between key health technologies

assessment stakeholders, e.g. funding bodies and industry, on the need for disease-specific models and a framework to develop and use such models.⁹⁸ Finally, a substantial investment in time and resources is required at the onset.⁹⁸ A compromise might be to consider whole treatment pathways, rather than the whole disease. This could address the limitations of focusing on a single decision-point.

1.5 THE CHALLENGE OF APPLYING META-ANALYSIS OF THE BEST AVAILABLE EVIDENCE TO INFORM TREATMENT SEQUENCING

1.5.1 Meta-analysis of treatment sequencing as a complex intervention

A recent project for developing practical tools and guidance for systematic reviews of complex interventions developed the following definition:⁹⁹

All complex interventions have two common characteristics; they have multiple components (*intervention complexity*) and complicated/multiple causal pathways, feedback loops, synergies, and/or mediators and moderators of effect (*pathway complexity*). In addition, they may also have one or more of the following three additional characteristics; target multiple participants, groups, or organizational levels (*population complexity*); require multifaceted adoption, uptake, or integration strategies (*implementation complexity*); or work in a dynamic multidimensional environment (*contextual complexity*).

This definition was developed by consolidating prior definitions including, among others, the one produced by the Medical Research Council.¹⁰⁰ It shows that, in most instances, treatment sequencing represents a multicomponent complex intervention, a quantitative evidence synthesis of which is unlikely to be straightforward.⁹⁹

Treatment sequencing represents a complex intervention because, as discussed in Section 1.2, treatment history and patient characteristics can have an effect on both the choice and the effectiveness of subsequent treatments. The subsequent evidence synthesis, developed to inform treatment sequencing decisions, also becomes complicated as it needs to account for both treatment history and the potential effect of subsequent treatment. The complexity of treatment sequencing is compounded by the fact that treatment history can represent many factors including the following:

- i. Carry-over effect of prior treatments
- ii. Type and level of response to previous treatment
- iii. Duration of treatment response
- iv. Time on treatment
- v. Intolerance or toxicity
- vi. The development of drug resistance
- vii. The burden of preceding treatments that can impact subsequent adherence

The choice of which treatment to use next can involve drug escalation, using an add-on therapy, trying a completely new treatment, or re-using of a previously effective treatment. Adding further to the complexity of the decision, some new treatments are only available for a subset of patients who

are intolerant or have failed to respond to a specific treatment, generally representing current practice. Furthermore, some biological or targeted therapies are only licenced for a specific biomarker-defined sub-group of patients, which I come back to in Section 1.7. Time and disease-trajectory are also important factors that can influence the effectiveness of treatment, the impact of which can be both dependent and independent of previous treatments. I come back to this point in Chapter 5, Section 5.5.

1.5.2 Applying meta-analysis of discrete treatments to inform treatment sequences

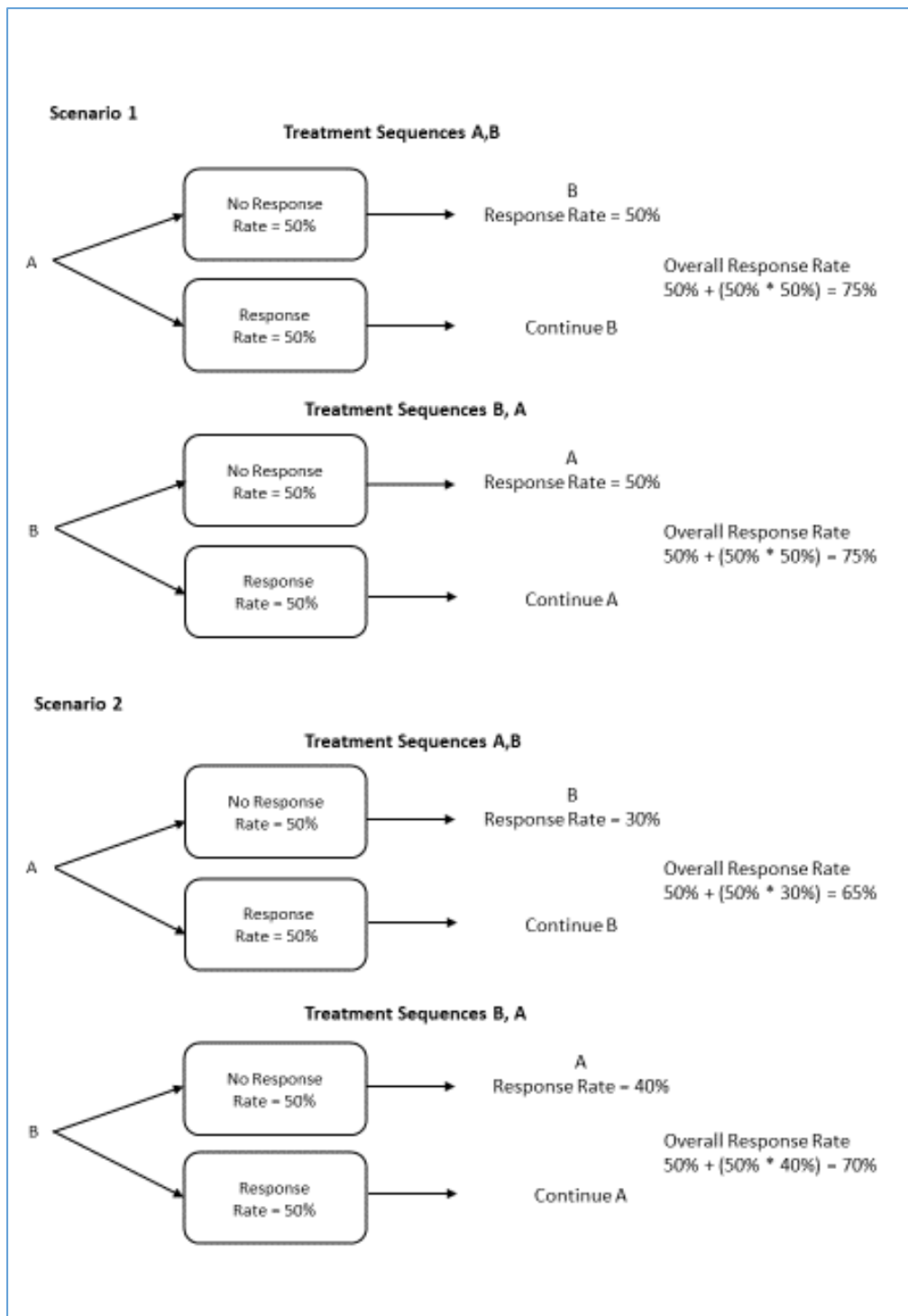
Where RCTs or comparative studies of whole sequences are available, summary effect estimates of each sequence can be developed using established meta-analytic methods, however, this may not be as straightforward as a meta-analysis of single, or discrete treatments, due to the complexity and dynamic nature of the intervention. In reality though, the available RCT evidence for informing treatment sequences is likely to be limited to studies of discrete treatments, used at single points in the treatment pathway. However, in conceiving treatment sequences as a series of discrete treatments and applying the RCT evidence to each treatment, implicit assumptions are likely to be made about the impact of treatment sequencing effects, as illustrated in the example provided in Box 1 and Figure 1.2. Meta-analyses of single treatments are also likely to provide poor evidence on sequencing effects, and generally fail to consider the long-term impact of multiple treatment sequences. This is illustrated in Figures 1.3 to 1.5. These examples (presented in Box 1 and Figures 1.2-1.5), however, represent a considerable simplification of the treatment sequencing issues, and are only aimed at providing an introduction and illustration of the potential limitations. This is explored in more depth in the review of methods presented in Chapters 3-7, which provides real examples and evaluates what is actually done in practice.

BOX 1: Example illustrating the potential implicit assumptions made regarding sequencing effects when applying effect estimates taken from studies of single treatments

Consider the hypothetical sequences of two drug treatments A and B, both of which are associated with a response rate of 50%, but A is considered to be less toxic, or more expensive, than B. The estimates for clinical effectiveness may be derived from a published RCT or meta-analysis but more importantly, for the purpose of this example, they are based on the evaluation of discrete treatments. In this circumstance, the decision may be to reserve drug B as second-line because both sequences (A,B and B,A) are considered to have the same overall response rate of 75%, but patients would be exposed to more harm when using a sequence starting with B (Scenario 1 in Figure 1.2). In absolute terms drug A is half as effective as B, but in relative terms their effectiveness is the same but used in different contexts. The underlying assumption here is that the treatment response (50%) for both drugs is the same irrespective of positioning in the sequence (Scenario 1 in Figure 1.2). The generic estimate of 50% response rate may not be representative of treatment B when used in patients who have failed to respond to drug A, and may therefore have a more treatment resistant and aggressive disease. Another approach would be to use an estimate of

the treatment effect for the first treatment derived from clinical trials of patients who were treatment naïve, and the second drug from trials that include patients with refractory disease, or receiving second-line-treatment. Drug A might be more effective than B in the second-line setting (Scenario 2 in Figure 1.2). This may render the overall response rate for the sequence A,B to be 65% and B,A to be 70%. However, this does not take into account that the effect of a treatment may differ according to the specific previous treatment used, for example the presence of antidrug antibodies to treatment A may be associated with a better response to treatment B when used after A (as discussed in Section 1.2.2).

Figure 1.2: An illustration of two treatment sequences, (A,B) and (B,A), where the overall response rate was the same for both in scenario 1, and in scenario 2 favoured the B,A sequence due to the differential response rates in second-line treatments

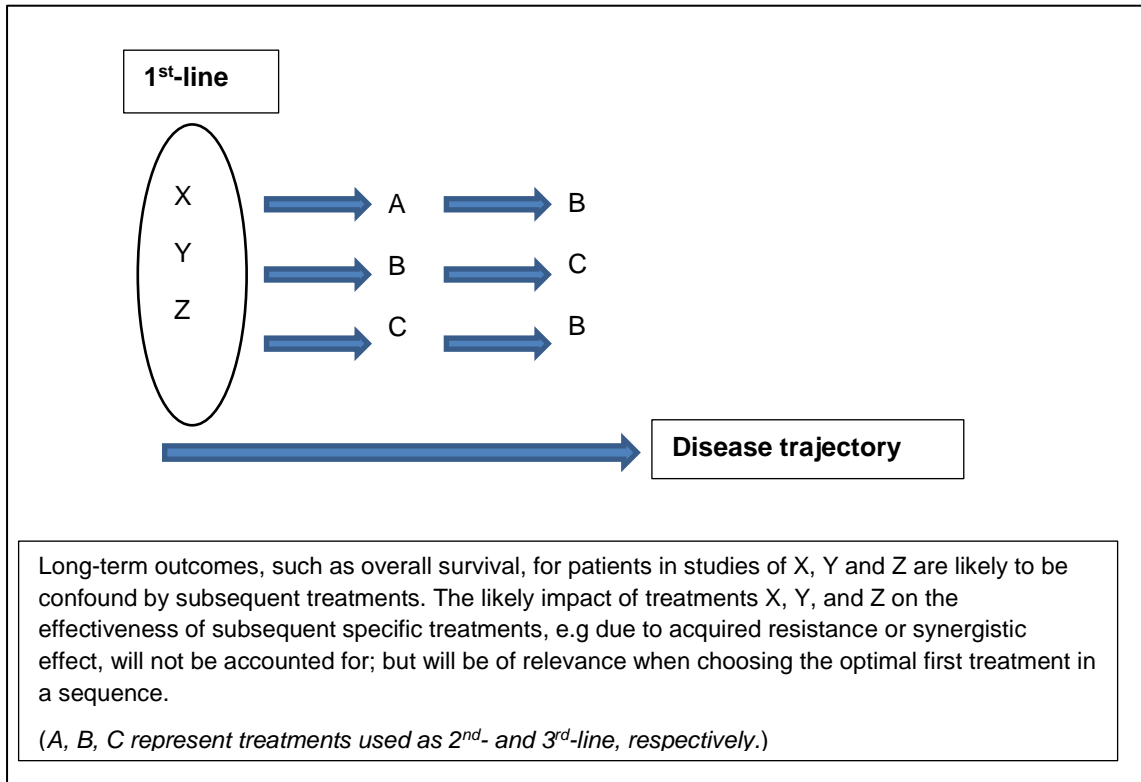


Adapted from a figure presented in: Kidwell, K. M. (2014). SMART designs in cancer research: past, present and future. Clinical Trials, 11(4), 445–456.

A meta-analysis of first-line treatments (Figure 1.3) may fail to provide good evidence for the optimal initial treatment as it cannot account for the prospective impact of subsequent treatments (as discussed in Section 1.2.2). The long-term outcomes of first-line treatments can also be confounded

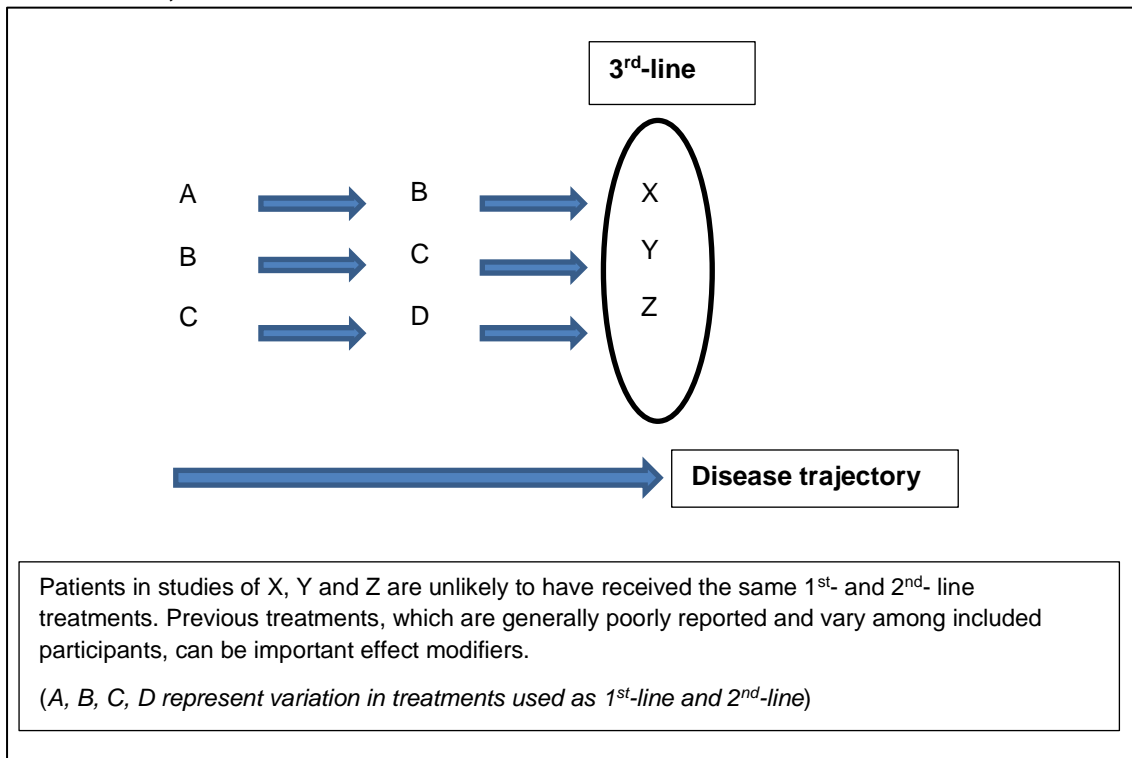
by the differential use of subsequent treatments. The meta-analysis may also fail to provide a good summary estimates of the same treatments used at a later point in a sequence as it will not account for the potential impact of treatment history and its influence on the choice of treatment and adherence, nor does it allow for the consequence of time or disease trajectory.¹⁷

Figure 1.3: Illustration of the type of evidence included in a meta-analysis of first-line treatments X, Y and Z



A meta-analysis of treatments used at a specific line (Figure 1.4), for example third-line, or in patients who are refractory to treatment, may potentially be problematic as it is unlikely to include RCTs with patients who received the same prior treatments. It is also unlikely to take into account the reason for discontinuing a previous treatment, which can lead to a differential effect of current treatment (as discussed in Section 1.2.2).

Figure 1.4: Illustration of the type of evidence included in a meta-analysis of third-line treatments X, Y and Z



A meta-analysis that includes RCTs of any treatment line (Figure 1.5), in other words, ignoring point of use within the treatment pathway, is likely to have considerable heterogeneity of patient characteristics. An example of this is the conventional pair-wise meta-analysis and network meta-analysis of sciatica treatments presented in Chapter 2. In addition to prior treatment, important patient characteristics that might act as effect modifiers include duration of symptoms, disease progression, severity, and type. In particular, consolidating evidence from various reviews investigating different points along the treatment pathway is not the ideal way of informing decision problems that involve an on-going risk that changes over time.

Consortium, and the All Wales Medicines Strategy Group (AWMSG). Health technology assessment is also an integral part of national decision-making processes in many other countries,¹⁰² and used by organisations such as the Institute for Quality and Efficacy in Healthcare in Germany, the Canadian Agency for Drugs and technologies in Health (CADTH), the Swedish Council for Health Technology Assessment (SBU) and most large payers in the United States (e.g. WellPoint and UnitedHealth Group).¹⁰² Health Technology Assessment plays an increasingly important role in the process of allocating future healthcare resources, especially in the context of innovative, high cost, interventions such as biological therapies.¹⁰⁴

Clinical guidelines advise on how healthcare professionals should care for people with specific conditions. A number of organisations publish evidence-based clinical practice guidelines in the UK, including NICE, Scottish Intercollegiate Guidelines Network (SIGN) and various professional bodies.

Comparative effectiveness research, conceived in the United States, was designed to inform healthcare decisions by providing evidence on the effectiveness, benefits, and harms of different treatment options. Evidence is generated in two ways, reviewing existing research, or conducting a new pragmatic clinical trial. The comparative effectiveness research review is a unique type of 'effectiveness review' or systematic review that depicts how the relative benefits and harms of a range of clinical options compared in the context of real-world healthcare decisions.^{105 106} They adhere to the same rigorous criteria which all systematic reviews must follow, but aim to answer more than the narrow question of whether a single therapy is safe and effective. In the absence of head-to-head trials, comparative effectiveness research reviews may use indirect comparisons following explicit methodological criteria. The comparative effectiveness research review compares the 'trade-offs' of multiple alternatives, each of which may vary with the underlying population and setting.

Quantitative evidence synthesis is an essential part of health technology assessments, comparative effectiveness reviews, and for developing clinical guidelines. A framework outlining the key issues and recommendations for performing quantitative evidence synthesis of the effectiveness of sequential treatment options is likely to be very useful for all stakeholders involved in any of these processes to inform policy and clinical decision-making both nationally and internationally. The process implemented by NICE is used as an exemplar to ensure that the thesis can achieve this. This is discussed in more detail in Chapter 4.

1.7 PERSONALISED MEDICINE

'Personalised' or 'individualised' medicine' is about targeting healthcare at patients who will benefit the most, using risk scores, patient characteristics, or other methods of stratification.⁹⁷ It is a growing area of research, especially for the use of individualised genomic, or biologic characteristics, also known as biomarkers.¹⁰⁷ However, the terms 'stratified' or 'precision medicine' are considered to be more accurate by some, as the objective of this approach is to identify demographic- or biomarker-

defined subgroups and therefore still remains a population and not an individualised approach.^{29 107} Many clinicians would argue that they have always practiced individualised or personalised medicine, whilst this new field of biomarker-based personalised medicine is driven by new diagnostics and therapeutics.²⁹

Personalised medicine is about the right patient receiving the right treatment at the right time. It is particularly prominent in oncology, where a growing proportion of new drugs target specific molecular mechanisms and only provide therapeutic benefit to a subset of patients defined by the presence of the altered target biomarker in the tumour.¹⁰⁸ The exclusion of patients without these mutations makes it possible to minimise exposure to costly and potentially toxic therapies that are unlikely to help them.²⁹ Research is also trying to identify biomarkers for predicting treatment responsiveness to biological agents for rheumatoid arthritis in order to identify patients most likely to benefit from each class or type of biological agent.¹⁰⁹ In response to personalised medicine, new clinical trial designs have been developed and implemented, particularly in oncology.¹⁰⁸ These adaptive trial designs are discussed in more detail in the Appendix Volume I (Section B).

The use of targeted drugs may introduce the need for individual level decision-making, as they allow treatments to be tailored for a specific patient. However, current research methods used for developing evidence to inform policy decision-making are based on the 'average' patient. The focus of NICE, which provides evidence-based recommendations to the UK National Health Service, is also in making policy decisions for the majority rather than the individual patient. This thesis therefore concentrates on investigating the best way to develop 'average' effect estimates to inform decision-making as a starting point. The expanding range of targeted and biological therapies available means that further work is likely to be needed for establishing the best way to inform individual patient-level clinical decision-making, for which this thesis will provide good preliminary work.

Single person trials, also known as N-of-1 trials, are increasingly recognised as potentially useful for informing personalised treatment decisions for patients with chronic conditions.¹¹⁰ N-of-1 trial is a special case of a cross-over trial in which the same patient is repeatedly randomised to receive either the experimental treatment or its control.¹¹¹ Because patients in cross-over trials and N-of-1 trials have received both control and intervention, these designs have the added advantage over conventional parallel trials of being able to provide information on patient preferences. An N-of-1 trial, which mimics usual clinical practice in its ability to allow flexible dosing and follow-up,¹¹² also provides data on the effectiveness of an intervention for 'this patient' rather than an average population, although the findings of multiple N-of-1 trials can be pooled to inform the latter.¹¹³ The main limitation of using these designs to inform sequential treatment is that potential carry-over effects, where the effect of one treatment persists into subsequent treatment periods, are minimised as much as possible. In clinical practice, once a treatment is discontinued, subsequent treatments are generally initiated without delay and, paradoxically, carry-over effect is the main disadvantage of these designs for evaluating effectiveness of individual treatments. When estimating the treatment effects from cross

over trials, attempts are also made to limit the period effect. However, knowing the effect of the time period (or duration of the condition) in which the treatment was administered would also be useful for assessing sequential treatment options. I re-visit to the use of N-of-1 trials in Chapter 5, Section 5.9.7.

1.8 SUMMARY OF THE PROBLEM RELATING TO SEQUENTIAL TREATMENTS

Evidence-based decisions on the optimum sequence of treatments, or optimal positioning of individual treatments within a treatment pathway, are needed to inform both clinical practice and policy. A number of interventions are available to treat many chronic diseases, which are likely to be used within a sequence of treatments. However, estimating the effectiveness of a sequence of treatment options is not straightforward. The evidence required to inform clinical and policy decision-making is likely to include studies with different designs, estimating different quantities related to the decision problem in question. Recent developments in meta-analytic techniques include multi-parameter evidence synthesis methods, such as network meta-analyses. These have been developed to encompass the broad evidence base needed to inform decision-making and to overcome some of the limitations of conventional pairwise meta-analyses. These methods may be useful for evaluating treatment sequences, however, their implementation may be hampered by gaps in the evidence base. Varying evidential requirements of regulatory authorities for the introduction of new medicines, and those of policy makers and clinicians for informing practice, have also introduced important gaps in the evidence base. An alternative approach is to use decision analytic modelling, which also utilises analytic judgements and assumptions. However, the extent and impact of the assumptions required to overcome the limitations of the evidence base on the uncertainty in the decision-making would likely be an important factor.

1.9 AIMS AND OBJECTIVES OF THE THESIS

The aim of the thesis is to develop a framework for conducting quantitative evidence synthesis of the effectiveness of sequential treatment options, within the context of a health technology assessment, for informing clinical and policy decision-making. It will be underpinned by a comprehensive review exploring and comparing current methods for estimating the treatment effects of an intervention conditional on the previous treatments administered. The added value of using complex evidence synthesis methods over more simplistic approaches will be investigated. The framework will consider the decision problem from the perspective of the policy maker. It will take into account that variation in disease characteristics could influence the solutions or methods needed in terms of the evidence synthesis.

Specific objectives are to:

- i. Identify the challenges of evaluating the clinical effectiveness of treatment sequences within the context of a health technology assessment using a case study that includes network meta-analyses and pair-wise meta-analyses of sciatica treatments

- ii. Identify clinical scenarios where treatment sequencing was an important consideration for policy or clinical decision-making, using NICE as an exemplar policy maker
- iii. Identify and review different meta-analytic methods developed to assess the clinical effectiveness of treatment sequences
- iv. Identify the type of simplifying assumptions relating to treatment sequencing effects made by decision analytic modelling studies in the absence of an adequate evidence base
- v. Identify and review decision analytic modelling approaches used to evaluate the effectiveness of treatment sequences
- vi. Develop an initial framework outlining the key issues and recommendations for undertaking quantitative evidence synthesis of the effectiveness of sequential treatment options, which can be further refined and tested and as part of a future research project

1.10 STRUCTURE OF THESIS AND SUMMARY OF CONTRIBUTION

1.10.1 Structure of the thesis

Chapter 2 represents the start of my exploration of the evaluation of treatment sequences. It is based on a health technology assessment of sciatica treatments, which identified some key challenges of conducting quantitative evidence synthesis to inform treatment sequences. This includes both network and pair-wise meta-analyses, which I conducted.

Chapter 3 introduces and presents the methods for the methodology review I conducted to identify previous quantitative evidence synthesis methods developed to inform treatment sequences. This represents the first comprehensive review of methods to investigate the evaluation of treatment sequencing for any disease condition, and incorporates both meta-analytic techniques and decision analytic modelling.

Chapter 4 explores the challenges faced by policy makers and health technology assessment agencies for evaluating treatment sequences. It uses NICE as the exemplar policy maker, and identifies the range of clinical conditions where treatments sequencing is likely to be an important consideration within NICE's technology appraisal and clinical guideline process. Preliminary work identified rheumatoid arthritis as a clinical condition for which treatment sequencing was particularly pertinent for NICE and is used as an exemplar in the thesis and the review of methods.

The findings of the review of methods are presented in **Chapters 5, 6, and 7**: divided into two parts covering meta-analytic techniques and decision modelling approaches separately, with the former presented in Chapter 5, and the latter in Chapters 6-7. In the absence of data on treatment sequencing effects, the decision analytic model conducted as part of the health technology assessment of sciatica treatments was based on the application of simplifying assumptions to effect estimates of single treatments (derived the network meta-analyses) in order to represent treatment sequencing effects. The type and range of assumptions used in other decision analytic models in

practice, to represent treatment sequencing effects, are presented in **Chapter 6**. This culminated in the development of a proposed new coding scheme of all assumptions possible for moving the science forward in this area. The review also showed that a wide range of different decision analytic modelling approaches have been used to model treatment sequences, which are summarised in **Chapter 7**, including an assessment of their advantages and disadvantages for achieving this.

The lessons learnt and the challenges identified from the various stages of the thesis are used to develop a novel framework that provides guidance for commissioners, producers, and users of health technology assessment, or similar process, on the evaluation of treatment sequences to inform policy and clinical decision-making. This is presented in **Chapter 8**.

The discussion and conclusions are presented in **Chapter 9**.

The Appendix for the thesis is presented as two separate volumes, which serve different purposes. **Volume I** provides supplementary information to the main text, for example the description of three clinical conditions (rheumatoid arthritis, advanced cancer, and epilepsy) for which treatment sequencing was important as identified in Chapter 4. **Volume II** provides supplementary results and the data extraction tables that relate to each included review.

1.10.2 Summary of contribution

The network meta-analyses presented in Chapter 2 were originally performed as part a health technology assessment commissioned by the National Institute for Health Research (NIHR) Health Technology Assessment programme (project number: 06/79/01).⁷ The health technology assessment, which included both clinical and economic evaluation, was based on a systematic review, meta-analyses, and a decision analytic model. The network meta-analyses and the economic model were subsequently published as separate articles in *The Spine Journal*,¹¹⁴ and *Pain*,¹¹⁵ respectively. The health technology assessment was conducted by a team of researchers and health economists. My role involved conducting all the pairwise and network meta-analyses under the supervision of one of my PhD supervisors, Professor Alex Sutton. I was involved in developing the search strategies, as well as selecting, summarising and critically appraising the included studies. I also prepared the methodology for the proposal of the initial project, developed the quality appraisal checklist, and wrote a statistical analysis plan. I was the lead author of both the health technology assessment monograph⁷ and the subsequent published article on the network meta-analyses.¹¹⁴ The findings presented in the journal publication represent the re-analysis of the data using a revised treatment categorisation, based on peer review comments. I undertook all the analyses for this publication, which included both network meta-analyses and accompanying conventional pair-wise meta-analysis, which were used to evaluate the robustness of the former.

1.10.3 My research experience and novel contribution of the thesis

My main research interest is evidence synthesis to inform clinical and policy decision-making. I have extensive systematic review experience covering a wide range of therapeutic indications and interventions, funded by the National Institute for Health Research Health Technology Assessment (NIHR HTA) programme, Cancer Research UK, Department of Health (England) and the Welsh Government. I have also conducted evidence reviews to underpin NICE Cancer guidelines. I have conducted evidence synthesis of both qualitative and quantitative data, with most meta-analyses being performed using frequentist statistics. The sciatica project represents my first network meta-analysis using Bayesian statistics. I believe that network meta-analysis, as well as other multi-parameter meta-analyses, epitomize an important step-forward in quantitative evidence synthesis methods for informing policy decision-making, as it allows multiple treatments to be compared simultaneously, and summary effect estimates to be developed for the comparison of treatments which have not been directly compared. However, working on my thesis has made me realise that it is too simplistic to consider the multitude of treatments now available for most chronic conditions as discrete treatments, or individual treatment lines. We now need to consider how to account for positioning in the treatment sequence, including reasons for discontinuing previous treatments, and the disease trajectory. This is not straightforward, and requires a more informative evidence base than is currently available. Three important trends that are likely to impact the available evidence base include: a) the drug marketing process, which can influence the type of RCTs available; b) access to individual patient data for researchers; and c) the availability of linked or big-data sets.

The original aim of my PhD was to develop quantitative evidence synthesis methods to evaluate the clinical effectiveness of sequential treatments from a meta-analytical and reviewer perspective only. However, as the project evolved it became clear that the evidence base is limited in most cases, necessitating the use of treatment sequencing assumptions and theoretical modelling. As a result, I became interested in developing my methodological knowledge of economic decision analytic modelling. This enabled me to explore the evaluation of treatment sequences from the viewpoint of the complementary disciplines of health economics and clinical effectiveness. The thesis provides a unique resource to inform the overall policy decision-making process, and for developing future guidance to inform the practice and methods for synthesising quantitative evidence on treatment sequencing options within health technology assessment or similar procedures. This is presented as a framework in Chapter 8, which lists all the key issues identified as part of the research project and their associated recommendations for practice. I am not a statistician or a modeller, and the thesis, therefore, does not include developing a new method for producing a summary effect estimate that is conditional on positioning in a treatment sequence. It does however, provide an important background and summary of the overall research position and challenges, which would be essential to inform such novel developments.

Prior to working on systematic reviews, I worked as both a State Registered Chiropractor and a Clinical Audit and Effectiveness Facilitator within the UK National Health Service. My progression from

experienced reviewer to studying for a PhD in health economics, complemented by my early career working in the NHS and my MSc in Information Science, provides me with a full spectrum of skills and expertise needed for conducting health technology assessment. This thesis considers the problem of how best to evaluate treatment sequences to inform policy decision-making from this broad perspective.

1.10.4 Dissemination

I presented earlier findings of the methodology review at the Cochrane Colloquium, the network meta-analyses of sciatica treatments in a published paper, and more recently the review of decision modelling studies at the ISPOR 20th Annual European Congress.

Lewis R, Wilkinson C, Sutton AJ, Woolacott N, Hughes DA, Ruiz F, Williams NH, Philips CJ. Methods for the assessment of the effectiveness of treatment sequences for clinical and economic decision-making. *21st Cochrane Colloquium*. Quebec, 19-23 September 2013 (*Oral presentation*)

Lewis R, Williams NH, Sutton AJ, Burton K, Ud Din N, Matar HE, Hendry M, Phillips CJ, Nafees S, Fitzsimmons D, Rickard I, Wilkinson C. Comparative clinical effectiveness of management strategies for sciatica: systematic review and network meta-analyses. *The Spine Journal* 2015 Jun 1;15(6):1461-77

Lewis RA, Hughes DA, Wilkinson C. PRM98: Decision analytic modelling methods for the assessing the effectiveness of treatment sequences for clinical and economic decision-making: a methodological review. *ISPOR 20th Annual European Congress*. Glasgow, 4-8 November 2017 (*Poster presentation*)

CHAPTER 2: A SYSTEMATIC REVIEW AND NETWORK META-ANALYSES OF THE CLINICAL EFFECTIVENESS OF DIFFERENT MANAGEMENT STRATEGIES FOR SCIATICA – AN EXAMPLE OF A HEALTH TECHNOLOGY ASSESSMENT

2.1 CHAPTER OVERVIEW

My interest in treatment sequencing started whilst developing an understanding to the limitations of a health technology assessment (HTA) of the clinical and cost-effectiveness of sciatica treatments. Sciatica is a condition where a number of different treatments are available, the majority of which have not been directly compared within primary studies. The commissioning brief for this project therefore commended the use of indirect treatment comparisons. This chapter presents the findings of network meta-analyses of the clinical effectiveness of treatments for sciatica, which I originally conducted for this health technology assessment. It includes a description and illustration of this relatively new statistical technique, and demonstrates the advantages of using network meta-analysis over a series of pairwise meta-analyses when multiple interventions are available for the same condition. As part of the clinical evaluation we were interested in assessing the effect of previous treatment as a potential effect modifier, but were unable to evaluate this in any depth due to the paucity of data. The clinical management of sciatica is generally based on a stepped-care approach, starting with non-invasive treatments such as advice and analgesia, followed by conservative or more invasive interventions if the symptoms do not resolve. The economic evaluation therefore investigated the cost-effectiveness of whole sequences. This raised an important question regarding whether the clinical evaluation should have also aimed to evaluate the clinical effectiveness of whole treatment sequences, or to develop treatment effect estimates of individual treatments conditional on positioning in the pathway. This in turn interested me in how best to evaluate the clinical effectiveness of treatment sequences, and how other health technology assessments had tackled this issue. It made me question the appropriateness of focusing on single treatments and essentially assuming that treatment effectiveness is independent of positioning in the pathway. This chapter includes a brief reflection of the potential impact of using this approach on clinical heterogeneity within the meta-analyses. The chapter ends with a rumination on the advantages and disadvantages of using a Bayesian random-effects model and its potential use for evaluating treatment sequences. It also introduces some recent methodological developments in meta-analytic techniques that may be used in existing health technology assessments to tackle some of the issues regarding treatment sequencing.

2.2 BACKGROUND TO THE HEALTH TECHNOLOGY ASSESSMENT

The network meta-analyses presented here represent an example of a health technology assessment where the latest methodological development in meta-analytic techniques was used. Despite the innovative method, it had not been feasible to develop treatment sequencing effects. My original intention was to explore the impact of previous treatments, or disease duration, which can be considered a proxy for the number of previous treatments used, as potential effect modifiers using

meta-regression. The poor reporting of both the previous treatments used, and chronicity of the disease within most included studies, precluded this.

These network meta-analyses were first conceived as part of a health technology assessment funded by the National Institute for Health Research (NIHR) HTA programme (project number: 06/79/01) and published as one of their monographs.⁷ The monograph included an evaluation of both the clinical and cost-effectiveness of all the treatment strategies for sciatica. The funding body had no role in design and conduct of the study; data collection; management, analyses, and interpretation of the data; preparation of the manuscript; or the decision to submit the article for publication. The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the HTA program, NIHR, NHS, or the Department of Health.

The clinical evaluation included a systematic review with an initial assessment of the findings for each individual treatment. This was based on a combination of a narrative synthesis and a series of 'head to head' pairwise meta-analysis for three relevant patient-based outcome measures. This also used three different follow-up periods, to represent short, medium, and long-term outcome. The analyses conducted for this initial evaluation are discussed in a little more detail in the Appendix Volume II: Appendix A1-2). The review also included the simultaneous comparison of all treatments, based on both direct and indirect comparisons, using network meta-analyses for all three outcome measures at a single time-point. I was a co-lead for the clinical assessment and undertook all the meta-analyses for this project.

The economic evaluation included a review of current cost-effectiveness studies and the development of a *de novo* decision analytic model to estimate the cost per quality adjusted life-year gained for each treatment strategy. Although the economic model considered treatment sequences, the clinical evaluation was not able to achieve this, due primarily to the limitations of the evidence base.

The network meta-analyses conducted for the clinical evaluation were subsequently published in *The Spine Journal*,¹¹⁴ and the decision analytic model undertaken as part of the economic evaluation in *Pain*.¹¹⁵

This chapter summarises the findings published in *The Spine Journal*, the target audience for which is primarily surgeons. The article reported updated network meta-analyses, which I conducted in response to peer review comments. This included a slightly different treatment categorisation to that reported in the original health technology assessment. It also focused on two outcome measures (global effect or pain intensity) at a single composite time-point; the third outcome measure included in the original health technology assessment was a composite of condition-specific outcome measures. The published article is presented here as it appears in the journal, which included an explanation of network meta-analysis and Bayesian analyses, as requested by the peer reviewers.¹¹⁴ It also represents work undertaken by a team of researchers.

2.3 INTRODUCTION TO THE CLINICAL EVALUATION OF SCIATICA TREATMENTS

Sciatica is the term used for the syndrome characterised by radicular leg pain, with or without sensory deficits, radiating along the distribution of the sciatic nerve.¹¹⁶⁻¹¹⁸ In about 90% of cases, it is caused by an intervertebral disc herniation resulting in nerve root irritation.¹¹⁹⁻¹²¹ It is a common reason for seeking medical advice,^{122 123} and has considerable economic consequence in terms of healthcare resources and lost productivity.¹²² The diagnosis and management of sciatica varies considerably within and between countries,¹¹⁹ which may reflect treatment availability, clinician preference and socio-economic variables rather than evidence-based practice.

Previous systematic reviews, including meta-analyses, have evaluated the effectiveness of various individual treatment approaches for sciatica, including conservative treatments,¹²⁴⁻¹²⁷ epidural steroid injections,^{124 126 128 129} and surgical procedures.¹³⁰ However, numerous treatments have not been directly compared. Furthermore, in order to choose the optimal treatment, it would be more helpful if all candidate treatments could be compared in the same analysis, as opposed to using a series of simple but inefficient standard pairwise meta-analyses comparing only two treatments at a time. It has been acknowledged that there is difficulty in interpreting the findings of multiple comparisons with low power, due to the small number of participants or events, which are inclined to result in statistically insignificant findings.^{7 131}

A network meta-analysis,¹³² by contrast, enables the simultaneous comparison of more than two treatment approaches, whilst combining data derived from both direct within-study comparisons between two treatment strategies (e.g. A vs B) and comparisons constructed from two studies that have one treatment in common (e.g. A vs B, B vs C).¹³¹ This type of analysis can only be applied to connected networks of randomised controlled trials (RCTs),¹³³ but preserves the within trial randomised comparison of each study¹³³ and allows information on treatment strategies to be “borrowed” from other studies within the network, thereby increasing the total sample size.^{134 135} Network meta-analysis conducted using Bayesian methods¹³⁶⁻¹³⁸ also allows the treatment strategies to be ranked in terms of clinical effectiveness with an estimate of the probability that each strategy is ‘best’.¹³⁹

The primary aims of the health technology assessment were to simultaneously compare the clinical effectiveness of different treatment strategies for sciatica using network meta-analyses, in order to identify the best treatment and to provide estimates for all possible pairwise comparisons, based on both direct and indirect evidence. The secondary aims were to demonstrate the feasibility of using network meta-analyses as a rational basis for clinical decision-making when a number of treatment options are available and where a series of conventional systematic reviews have failed to help with real-world treatment decisions.

2.4 METHODS

2.4.1 Review methods

A full account of the study methods and literature search are presented in the health technology assessment monograph (which also includes the protocol).⁷ The review was conducted in line with the principles of good practice¹⁴⁰ and presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.¹⁴¹

2.4.2 Search strategy

Included studies were identified via an extensive literature search described in full, including the search strategy, in the health technology assessment monograph.⁷ The search incorporated 28 electronic databases and trial registries including MEDLINE, EMBASE, and AMED. Databases were searched from inception until December 2009 without language restriction. The reference lists of previous systematic reviews and included studies were also scanned for further references.

2.4.3 Study selection and data extraction

This review included any comparative study (experimental or observational) with adults who had sciatica diagnosed clinically, or where clinical imaging confirmed lumbar disc prolapse consistent with the clinical findings. The essential clinical criterion was radicular leg pain worse than back pain.⁷ Studies of sciatica caused by conditions other than a prolapsed intervertebral disc were included if it was documented that radicular leg pain was worse than back pain. If imaging was used, it had to demonstrate evidence of nerve root compromise. Studies that included participants with non-specific low back pain were only included if the findings for patients with sciatica were reported separately. Any type of intervention to treat sciatica was considered. These were categorised, for the purpose of the present analyses, into one of 21 categories (See Table 2.1). Interventions that included a combination of more than one treatment strategy (or mixed treatments) were excluded from the network meta-analyses due to uncertainty regarding the extent of interaction between the combined interventions. The same applied to post-surgical interventions due to surgery being included as a separate treatment category. Studies comparing interventions that were grouped under the same treatment strategy were also excluded. Three further studies evaluating experimental interventions for sciatica (common peroneal nerve block, proteolytic enzyme, and colchicine) were excluded from the analyses as these interventions did not fit the treatment categorisation. The analyses presented were also limited to studies that reported data on overall response or pain intensity.

Table 2.1: Treatment categorisation

Treatment strategy ^a	Treatment strategy Code ^a	Type of treatment
Inactive control	A	Placebo or sham treatment in any type or format: tablet, injection, epidural etc. No treatment
Conventional care	B	Conservative therapy Conventional care Non-surgical treatments General practitioner care
Disc surgery	C	Discectomy

Treatment strategy ^a	Treatment strategy Code ^a	Type of treatment
		Micro-discectomy Laminectomy Hemilaminectomy Surgical decompression
Epidural injections (includes spinal nerve block)	D	Lumbar epidural injection Transforminal epidural injection Intraforaminal injection Interlaminar epidural injection Caudal epidural injection Periradicular injection Spinal nerve root block
Chemoneucleolysis	E	Chymopapain Collagenase Ozone
Non-opioids	F	Conventional pain or anti-inflammatory medication used as oral, intravenous or intramuscular: Steroids Non-steroidal anti-inflammatory drugs (NSAIDs) Paracetamol
Intra-operative interventions	G	Barrier membranes: - Antiadhesion barrier - Fat graft Intra-operative steroid +/- local anaesthetic
Traction	H	Mechanical traction Antigravitational traction Auto-traction Manual traction
Manipulation	I	Chiropractic Osteopathic
Acupuncture	J	
Exercise therapy	K	Exercise therapy Isometric exercises Mobilising and strengthening exercises
Passive physical therapy	L	Ultrasound Transcutaneous electrical nerve stimulation (TENS) Infra-red heat Physical therapy programme (hot pack, continuous ultrasound, and diadynamic currents) Conservative physiotherapy
Biological agents	M	Cytokine modulating treatments targeting tumour necrosis factor alpha: - Entanercept - Infliximab - Autologous Conditioned Serum
Bed rest	N	
Opioids	O	Oral, intravenous or intramuscular opioids
Education/Advice	P	Advice to keep active Advised to continue activities of daily living
Percutaneous discectomy	Q	Automated percutaneous discectomy Percutaneous automated nucleotomy Nucleoplasty Laser discectomy
Neuropathic painmodulators	R	Pharmaceutical treatment used for neuropathic pain: Anti-epileptic medication Tricyclic antidepressants
Intra-discal injections	S	
Spinal cord stimulation	T	
Radiofrequency treatment	U	

^a Interventions are summarised using these codes for displaying the results of the network-meta-analyses

Two reviewers screened studies for inclusion independently. Data were extracted by one reviewer and checked by a second using the original paper, whilst quality assessment was done by two reviewers independently. Any disagreements were resolved by discussion. The quality of both trials and observational studies was assessed using the same checklist, which was based on one used by the Back Review Group of the Cochrane Collaboration for RCTs¹⁴² and another recommended by the Guidelines for Systematic Reviews in Health Promotion and Public Health Taskforce,¹⁴³ which was developed by the Effective Public Health Practice Project, Canada.¹⁴⁴ The criteria covered external validity, selection bias and confounding, detection bias, performance bias, and attrition bias. Studies were coded as strong, moderate or weak for each domain, estimating the risk of bias.

2.4.4 Outcome measures

Overall response or global effect was analysed as a binary outcome (treatment success vs failure) and synthesised using odds ratios (ORs). Where studies reported overall response in terms of both overall improvement and improvement in leg pain, the data on overall improvement were used. For studies that reported both physician and patient perceived global effect, the data for patients' perceived effects were used.

Pain intensity (on a scale of 0-100) was analysed as a continuous outcome measure using weighted mean difference (WMD). We included only pain assessment from one location from each study using the preference hierarchy of leg pain then overall pain. Where feasible, missing data were estimated from the published data, using standard methods, such as standard deviations derived from standard errors.¹⁴⁵ Where mean values were unavailable but the medians reported, these were used instead. If standard deviations for baseline values were available these were substituted for missing standard deviations. For studies that did not report sufficient data to derive the standard deviations, they were imputed using the weighted mean,¹⁴⁶ which was calculated separately for each intervention category.

2.4.5 Statistical analysis

The network meta-analyses were based on a single time-point, using the findings from individual studies closest to six months follow up. Sensitivity analyses were conducted to assess the impact of excluding non-randomised studies (observational studies and non-RCTs).

The network meta-analyses were conducted using a hierarchical random-effects model¹³² within the Bayesian framework. Bayesian methods are based on the idea that unknown quantities, such as population means or proportions, have probability distributions.¹³⁷ One starts with a distribution that is based on prior knowledge or subjective belief about the population and then update this using data from your included studies. However, using non-informative priors (such as, a normal distribution with a large variance) means that the results are based predominantly on the data from the included studies, and as such will mirror those obtained using frequentist or classical meta-analysis methods. Bayesian methods are implemented using model-based simulations, which means that they can be

used to perform complex analyses that incorporate multiple data sources and allow for various parameter uncertainties within a single coherent model, which is why we chose to use these methods.

The network meta-analyses were conducted using WinBUGS 1.4.3 software (MRC Biostatistics Unit, Cambridge, UK),¹⁴⁷ which uses Markov chain Monte Carlo simulation methods to run thousands of simulated iterations based on the data and description of the proposed distributions for relevant parameters. The iterative simulations are generally started at multiple points, in order to ensure the samples are drawn from the whole sampling frame. The first 50,000 iterations (or burn-in) were discarded, and the results based on a further sample of at least 100,000 simulations, ensuring that the multiple simulation strings have converged and distributions were informed by later simulations. Numerical methods such as the Brooks-Gelman-Rubin statistic¹⁴⁷ and the inspection of the auto-correlation and history plots, which are routine assessments made when using Markov chain Monte Carlo methods, were used to check that convergence had occurred. The model fit was checked by the global goodness of fit statistic, residual deviance. If the model is an adequate fit, it is expected that the residual deviance should be roughly equal to the number of data points.¹³³ Non-informative priors were used for normal distributions for means, and uniform distributions for standard deviations. The treatment strategy 'inactive control' was used as the reference treatment. This included interventions that represent the non (active) treatment of sciatica, such as no treatment, sham treatment, or placebo (two studies used active placebo). The WinBUGS codes (or models) that we used are presented in the Appendix Volume II (Appendix A3). The robustness of the network meta-analyses were also evaluated by comparing the findings (where head to head studies were available) with those of standard 'direct' pairwise meta-analyses⁷ conducted using a random-effects model¹⁴⁸ based on frequentist methods¹³⁶⁻¹³⁸ in Stata 10 (Stata Corp LP, College Station, TX, USA).

The assumptions of a random-effects network meta-analysis are that (1) the treatment effects are additive (i.e. the relative effect of treatment A vs C can be estimated from the effect of A vs B and B vs C),^{133 149 150} (2) study-specific treatment effects are drawn from a common distribution (exchangeable);^{56 133} and (3) this common distribution or heterogeneity is constant between the different comparisons.^{56 133} The heterogeneity between studies, defined as the variability of the results across studies within each treatment comparison over and above chance,¹⁵¹ was evaluated by examining the findings of standard pairwise meta-analyses using visual inspection of the forest plots, using the Chi² statistic to test for statistical heterogeneity, and the I² statistic to quantify this.^{152 153}

2.5 RESULTS

2.5.1 Included studies

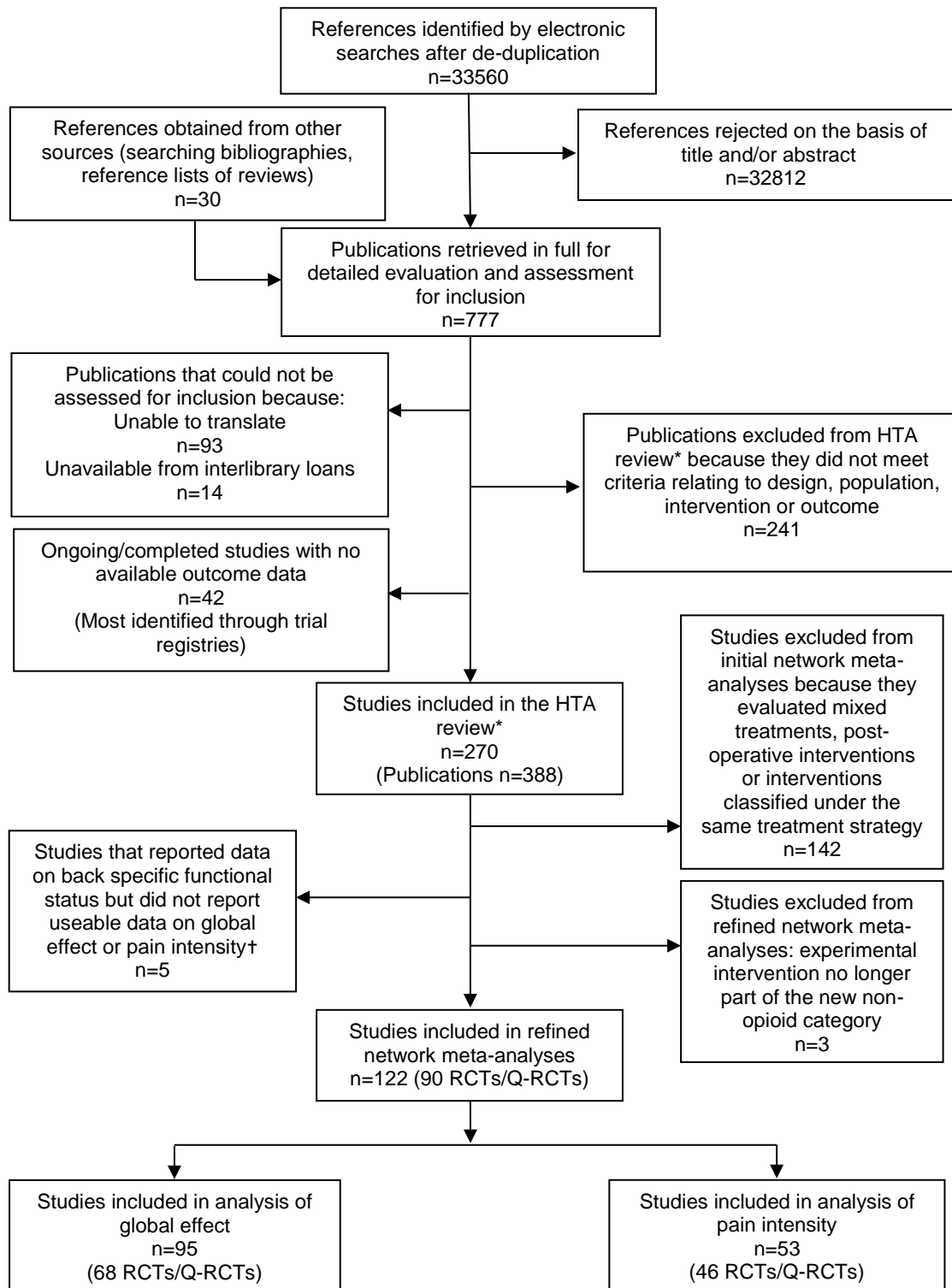
As seen in Figure 2.1, 122 studies were included in the revised network meta-analyses (one publication included two studies), 86 were RCTs and four quasi-RCTs (Q-RCTs). The network meta-analysis of global effect included 95 studies (68 RCTs/Q-RCTs) and pain intensity 53 studies (46 RCTs/Q-RCTs). A list of included studies is presented in the Appendix Volume II (Appendix A4). This

also includes a summary of the quality assessment of included studies. A description of the interventions, populations, study design, and outcome data for the included studies are available in the extensive supplementary data accompanying the published paper¹¹⁴ and on request.

Eleven (9%) studies had a strong overall quality rating and eight (7%) had a strong overall external validity rating; five (4%) of which had a strong rating for both. Only 26 (21%) studies used both adequate randomisation and adequate or partially adequate (using sealed envelopes, n=16) allocation concealment.

The proportion of studies that limited inclusion to patients with acute sciatica (duration of symptoms <3 months) was much higher in conservative treatments, such as traction (71%), bed rest (80%), and non-opioid medication (53%), than more invasive treatments (such as disc surgery 8%, chemonucleolysis 3%, and epidural 5%). However, most studies did not report the duration of sciatica, or included patients with acute and chronic sciatica. The presence of disc herniation was also confirmed by imaging in a high proportion of studies evaluating invasive treatments such as percutaneous discectomy (100%), disc surgery (86%) and chemonucleolysis (84%). Previous treatment was poorly reported in many studies, but the proportion of studies that reported patients who had received previous treatment was higher for invasive treatments such as disc surgery (70%), percutaneous discectomy (100%), and chemonucleolysis (88%), than for conservative treatments such as non-opioids (20%), traction (29%), and acupuncture (33%). The mean pain score (where reported), at baseline for each treatment strategy were fairly similar (ranging from 59 to 69) with the exception of biological agents (78).

Figure 2.1: Flow diagram showing the number of references identified, publications retrieved for assessment, and studies included in the review

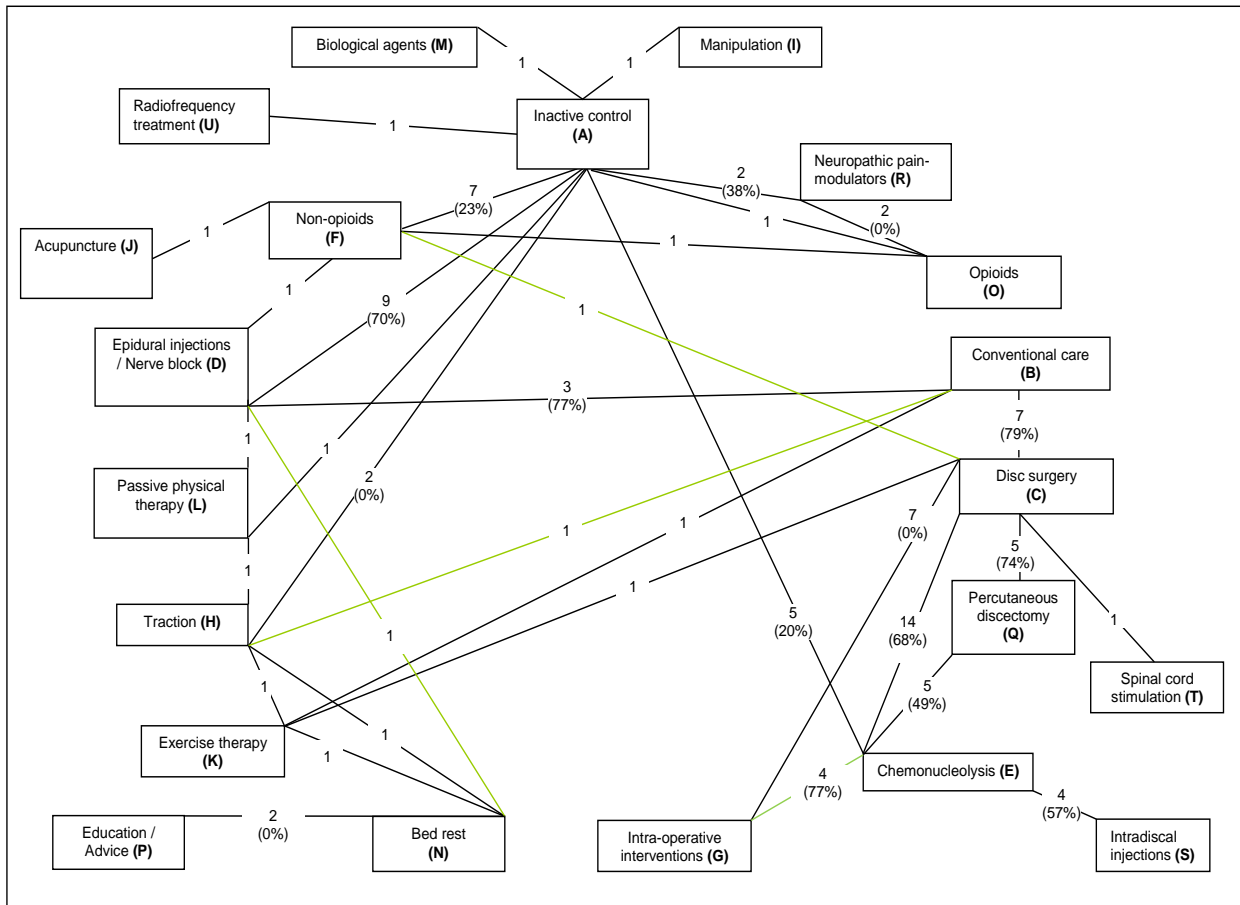


*The current study represents further refinements of initial network meta-analyses conducted as part of a broader Health Technology Assessment (HTA) review evaluating the clinical and cost effectiveness of sciatica treatments;⁷ the original review included a narrative synthesis, conventional pairwise analyses, network meta-analyses, a review of economic evaluations and an economic evaluation.

†Five studies reported data on back specific functional status (an outcome considered as part of the initial network meta-analyses), but did not report useable data for global effect or pain intensity.

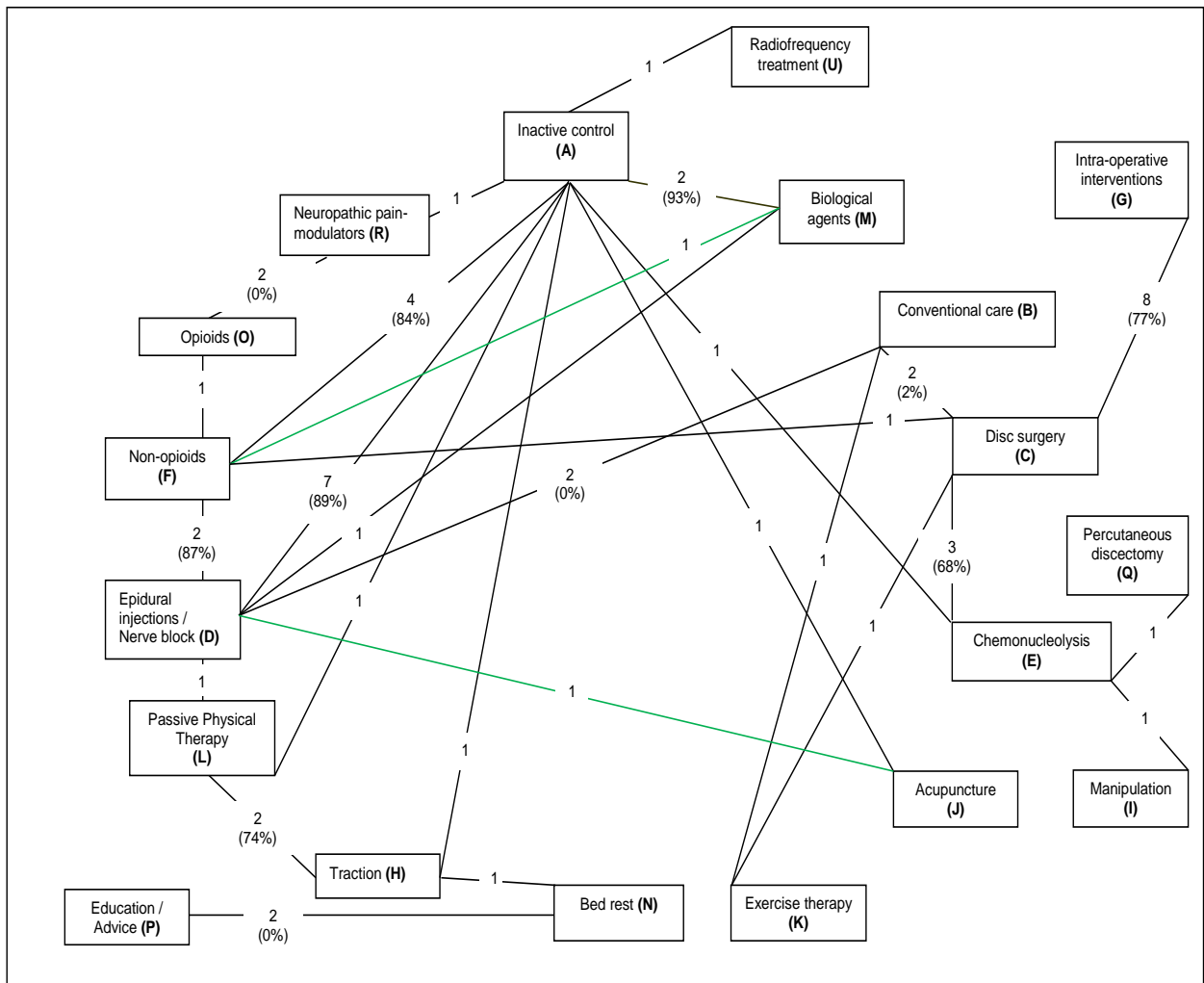
Figure 2.2 shows the network of treatment comparisons for the network meta-analysis of global effect and Figure 2.3 shows the same for the analysis of pain intensity.

Figure 2.3: Network of treatment strategies for sciatica for comparative studies reporting global effect



Note: Each treatment strategy is a node in the network. The links between the nodes are arms in head-to-head (direct) comparisons in eligible studies. The numbers along the link lines indicate the number of studies or pairs of study arms for that link in the network, with the observed I² statistic (based on the data from the pairwise meta-analyses) is presented in brackets. Links that do not have any randomised or quasi- randomised controlled trial evidence are indicated in green.

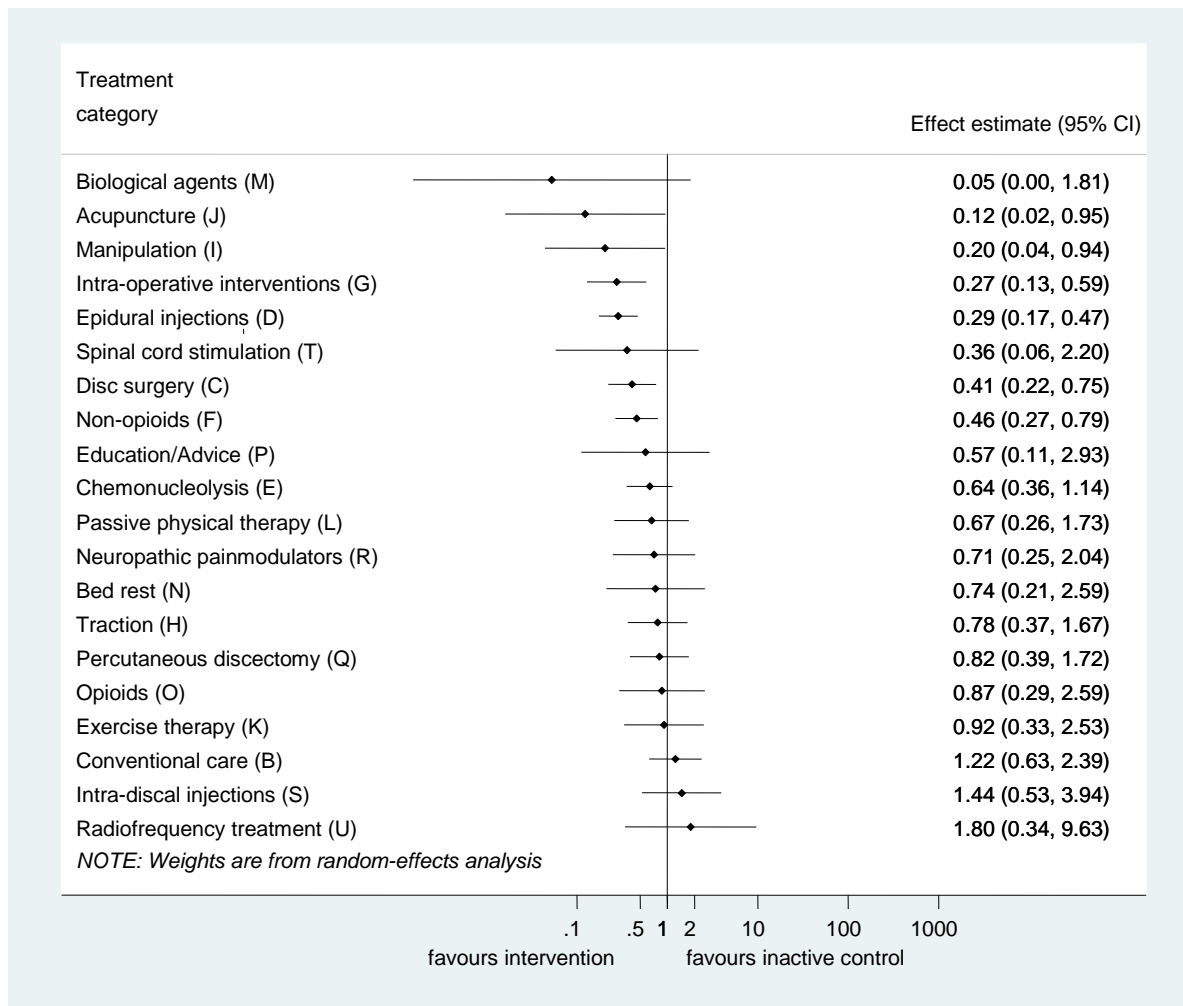
Figure 2.3: Network of treatment strategies for sciatica for comparative studies reporting pain intensity



Note: Each treatment strategy is a node in the network. The links between the nodes are arms in head-to-head (direct) comparisons in eligible studies. The numbers along the link lines indicate the number of studies or pairs of study arms for that link in the network, with the observed I^2 statistic (based on the data from the pairwise meta-analyses) is presented in brackets. Links that do not have any randomised or Quasi- controlled trial evidence are indicated in green.

Summary effect estimates for the comparison of each intervention strategy with inactive control are presented in Figures 2.4 and 2.5. The corresponding confidence intervals provide an indication of the uncertainty surrounding the effect sizes, which needs to be taken into account when interpreting the data (especially the probability of being best). The probabilities for each treatment strategy being best (or most effective) are presented in Appendix Volume II (Appendix A5). The network meta-analyses also provide a full set of comparisons for all treatment strategies, the findings of which are presented in Tables 2.2 and 2.3. The summary effect sizes derived from the network meta-analyses can be directly compared with the summaries of pairwise meta-analyses (derived using Stata 10), which are presented in the same matrices (top right-hand corner); statistically significant findings are indicated by shading. The results of sensitivity analyses restricted to RCTs and Q-RCTs are presented in the Appendix Volume II (Appendix A5-6).

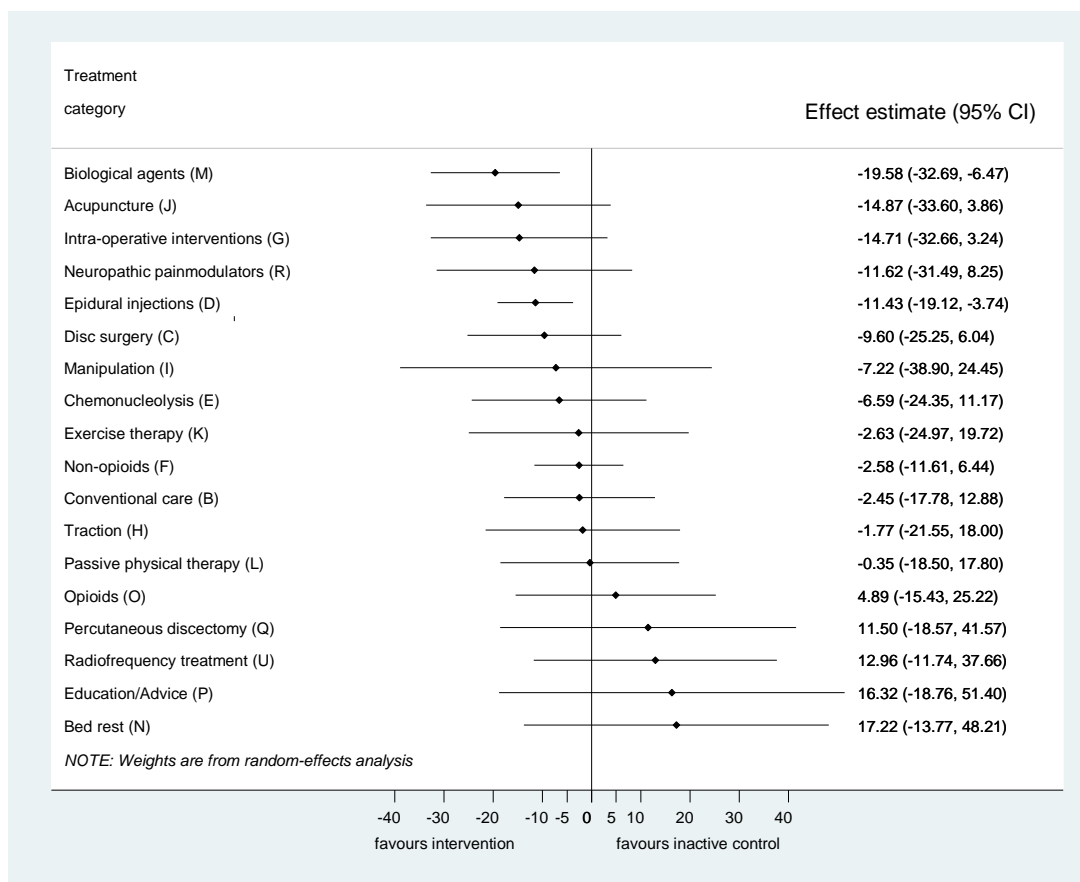
Figure 2.4: Plot of the odds ratios of global effects for different treatment strategies compared with inactive control from the network meta-analysis



Note: The data has been spun around so that effect estimates that favour the intervention are shown on the right hand side. This means that the an odds ratios <1 represents a decrease in the number of patients *not* showing overall improvement in favour of the intervention

Abbreviations: CI confidence interval

Figure 2.5: Plot of the weighted mean difference for pain intensity for different treatment strategies compared with inactive control from the network meta-analysis



Note: A weighted mean difference > 0 represents a reduction in pain intensity in favour of the intervention

Abbreviations: CI confidence interval

2.5.2 Overall response

In terms of overall response or global effect, the following treatment comparisons with inactive control (A) or conventional care (B) were statistically significant at the 5% level: disc surgery (C), epidural injections (D), non-opioids (F), intra-operative interventions (G) such as barrier membranes and steroids used during the surgical procedure, spinal manipulation (I), acupuncture (J), and chemonucleolysis (E). Intra-discal injections (S) were found to be statistically significantly worse than disc surgery (C), epidural injections (D), non-opioids (F), intra-operative interventions (G), manipulation (I), and acupuncture (J). Percutaneous discectomy (Q) was found to be inferior to disc surgery (C), epidural injections (D), and intra-operative interventions (G). Traction (H) and exercise therapy (K) were also found to be inferior to epidural injections and intra-operative interventions. Radio frequency treatment (U) was statistically significantly inferior to disc surgery (C), epidural injections (D), intra-operative interventions (G), and acupuncture (J). Finally, chemonucleolysis (E) was statistically significantly less effective than epidural injections, disc surgery, and intra-operative interventions. The largest treatment effects for the comparison with inactive control were for biological agents and acupuncture, which also had the highest probability of being best (0.57 and 0.26 respectively). The comparison of biological agents with the following treatments also showed large

effect estimates (OR >10), but these were not statistically significant: chemonucleolysis (E), traction (H), exercise therapy (K), passive physical therapy (such as ultrasound and transcutaneous electrical nerve stimulation) (L), bed rest (N), opioid medication (O), percutaneous discectomy (Q), intra-discal injections (S), and radio frequency treatment (U), all of which were associated with very wide confidence intervals. This reflects the limited evidence available for biological agents, which included a small placebo controlled RCT (n=24) that reported a large effect estimate in favour of biological agents (OR 10.0; 95% CI: 0.65, 166.67). See Appendix Volume II: Appendix A5, Table A5.1.

The results of the sensitivity analyses excluding non-randomised studies showed broad agreement with the main analyses. For global effect, the most notable discrepancies occurred with biological agents compared with chemonucleolysis, conventional care, and exercise therapy. A more detailed narrative of the differences between the analyses with and without the non-randomised studies is presented in the Appendix Volume II (Appendix A6).

2.5.3 Pain intensity

In terms of pain intensity, the only treatment comparisons with inactive control that were statistically significant were epidural injections (D) and biological agents (M). Biological agents, which had the highest probability of being best (0.33), were also found to be statistically significantly better at reducing pain than non-opioids (F), bed rest (N), opioids (O) and radio frequency treatment (U); these findings were all associated with wide credible intervals. When considering the magnitude of effect, bed rest (N), education/advice alone (P), percutaneous discectomy (Q), and radiofrequency treatment (U) tended to fare worse when compared with most treatment strategies, with findings showing a non-statistically significant difference of more than 25 points. Acupuncture (J), had the second highest probability of being best (0.19) and resulted in reductions of pain intensity of more than 25 points compared with bed rest, opioids, education/advice alone, percutaneous discectomy and radio frequency treatment, none of which were statistically significant and all had wide credible intervals.

For pain intensity the most notable discrepancies between the network meta-analysis with and without non randomised studies only occurred with biological agents (vs inactive control, conventional care, disc surgery, non-opioids, intraoperative interventions, acupuncture, exercise therapy, opioids, and neuropathic pain modulators). Biological agents no longer had the highest probability of being best (0.03; see Appendix Volume II: Appendix A5, Table A5.4). These discrepancies are likely to be due to the small number of included studies with a limited number of participants evaluating biological agents (2 RCTs n=131; 1 non-randomised controlled trial n=72; and 1 historical cohort study n=10).

Table 2.2: Results (odds ratios, with 95% confidence intervals/credible intervals) of the network meta-analysis for global effect

A			N=9, 2.58 (1.3, 5.3)	N=5, 2.56 (1.6, 4.1)	N=7, 1.68 (1.1, 2.5)		N=2, 1.11 (0.6, 2)	N=1, 4.72 (1.9, 11.4)		N=1, 4.27 (1.5, 12.4)	N=1, 10 (0.7, 166.7)		N=1, 1.37 (0.5, 3.8)		N=2, 2.01 (0.8, 5.2)		N=1, 0.57 (0.2, 1.7)			
0.82 (0.4, 1.6)	B	N=7, 2.7 (1.7, 4.3)	N=3, 5.46 (0.8, 38.5)				N=1, 1.53 (0.6, 4.2)			N=1, 1.46 (0.7, 2)										
2.46 (1.3, 4.5)	3.0 (1.9, 4.9)	C		N=14, 0.5 (0.4, 0.7)	N=1, 6.71 (0.8, 58.8)	N=7, 1.49 (1, 2.2)				N=1, 0.77 (0.18, 3.2)					N=5, 0.67 (0.4, 1.3)		N=1, 1.13 (0.4, 3.6)			
3.48 (2.1, 5.8)	4.26 (2.1, 8.8)	1.42 (0.7, 2.9)	D		N=1, 0.45 (0.2, 1.4)					N=1, 0.2 (0.1, 0.6)			N=1, 0.22 (0.1, 0.9)							
1.57 (0.9, 2.8)	1.92 (1.1, 3.4)	0.64 (0.5, 0.9)	0.45 (0.2, 0.9)	E		N=4, 2.03 (0.8, 5.4)									N=5, 0.65 (0.4, 1.1)		N=4, 0.46 (0.2, 1)			
2.18 (1.3, 3.8)	2.66 (1.2, 6.2)	0.89 (0.4, 1.9)	0.63 (0.3, 1.2)	1.39 (0.7, 3.0)	F				N=1, 3.27 (0.8, 13.9)					N=1, 0.18 (0.1, 0.7)						
3.64 (1.7, 7.8)	4.5 (2.2, 9.0)	1.48 (0.9, 2.5)	1.05 (0.4, 2.4)	2.32 (1.4, 4.0)	1.67 (0.7, 4.1)	G														
1.28 (0.6, 2.7)	1.56 (0.7, 3.7)	0.52 (0.2, 1.2)	0.37 (0.2, 0.8)	0.8 (0.3, 1.9)	0.59 (0.2, 1.5)	0.35 (0.1, 0.9)	H			N=1, 0.87 (0.3, 2.7)	N=1, 0.93 (0.5, 1.9)		N=1, 1 (0.1, 7.1)							
4.88 (1.1, 23)	5.96 (1.1, 32.2)	1.98 (0.4, 10.5)	1.4 (0.3, 7.0)	3.1 (0.6, 16.3)	2.24 (0.4, 11.4)	1.34 (0.2, 7.5)	3.83 (0.7, 21.4)	I												
7.92 (1.1, 67)	9.72 (1.2, 89.2)	3.22 (0.4, 29.2)	2.27 (0.3, 19.9)	5.0 (0.7, 45.5)	3.62 (0.5, 28.4)	2.18 (0.3, 20.7)	6.21 (0.8, 59.0)	1.63 (0.1, 22.3)	J											
1.09 (0.4, 3)	1.34 (0.5, 3.4)	0.44 (0.2, 1.2)	0.31 (0.1, 0.9)	0.7 (0.3, 1.9)	0.5 (0.2, 1.5)	0.30 (0.1, 0.9)	0.86 (0.3, 2.4)	0.22 (0.04, 1.4)	0.14 (0.01, 1.2)	K			N=1, 2.2 (0.6, 7.6)							
1.49 (0.6, 3.9)	1.82 (0.6, 5.5)	0.61 (0.2, 1.8)	0.43 (0.2, 1.1)	0.95 (0.3, 2.8)	0.68 (0.2, 2)	0.41 (0.1, 1.3)	1.17 (0.4, 3.2)	0.3 (0.05, 1.8)	1.37 (0.02, 1.7)	1.37 (0.4, 5.0)	L									
16.83 (0.8, 947)	20.75 (0.9, 1213)	6.89 (0.3, 400)	4.8 (0.2, 278)	10.77 (0.5, 625)	7.71 (0.3, 449)	4.65 (0.2, 277)	13.3 (0.5, 792)	3.48 (0.1, 252)	2.15 (0.1, 185)	15.63 (0.6, 977)	11.41 (0.4, 699)	M								
1.35 (0.4, 4.7)	1.65 (0.5, 6.0)	0.55 (0.2, 2)	0.39 (0.1, 1.3)	0.86 (0.2, 3.2)	0.62 (0.2, 2.4)	0.37 (0.1, 1.5)	1.06 (0.3, 3.8)	0.28 (0.04, 2.0)	0.17 (0.01, 1.7)	1.24 (0.4, 4.3)	0.9 (0.2, 4.0)	0.08 (0.0, 2.3)	N		N=2, 1.31 (0.8, 2.3)					
1.15 (0.4, 3.4)	1.41 (0.4, 5.1)	0.47 (0.1, 1.6)	0.33 (0.1, 1.1)	0.73 (0.2, 2.5)	0.53 (0.2, 1.6)	0.32 (0.1, 1.2)	0.9 (0.2, 3.4)	0.24 (0.04, 1.5)	0.14 (0.01, 1.3)	1.05 (0.2, 4.7)	0.77 (0.2, 3.3)	0.07 (0.0, 1.8)	0.85 (0.2, 4.5)	O		N=2, 0.78 (0.4, 1.7)				
1.75 (0.3, 9.0)	2.14 (0.4, 11.3)	0.71 (0.1, 3.8)	0.50 (0.1, 2.5)	1.12 (0.2, 6.0)	0.80 (0.1, 4.4)	0.48 (0.1, 2.7)	1.37 (0.3, 7.2)	0.36 (0.04, 3.4)	0.22 (0.02, 2.8)	1.6 (0.3, 8.3)	1.17 (0.2, 7.1)	0.1 (0.0, 3.5)	1.30 (0.5, 3.7)	1.52 (0.2, 10.8)	P					
1.23 (0.6, 2.6)	1.50 (0.7, 3.0)	0.5 (0.3, 0.8)	0.35 (0.2, 0.8)	0.78 (0.5, 1.3)	0.56 (0.2, 1.4)	0.34 (0.2, 0.7)	0.96 (0.4, 2.5)	0.25 (0.04, 1.4)	0.15 (0.02, 1.2)	1.13 (0.4, 3.3)	0.82 (0.3, 2.6)	0.07 (0.0, 1.7)	0.91 (0.2, 3.6)	1.07 (0.3, 4)	0.70 (0.1, 3.9)	Q				
1.41 (0.5, 4.0)	1.72 (0.5, 6.0)	0.57 (0.17, 1.9)	0.4 (0.1, 1.3)	0.89 (0.3, 3.0)	0.65 (0.2, 2)	0.39 (0.1, 1.4)	1.1 (0.3, 4.1)	0.29 (0.04, 1.8)	0.18 (0.02, 1.6)	1.29 (0.3, 5.6)	0.94 (0.2, 3.8)	0.08 (0.0, 2.2)	1.04 (0.2, 5.4)	1.23 (0.4, 3.6)	0.8 (0.1, 5.6)	1.15 (0.3, 4.2)	R			
0.7 (0.3, 1.9)	0.85 (0.3, 2.3)	0.28 (0.1, 0.7)	0.2 (0.1, 0.6)	0.44 (0.2, 1)	0.32 (0.1, 0.97)	0.19 (0.1, 0.5)	0.54 (0.2, 1.8)	0.14 (0.02, 0.9)	0.09 (0.01, 0.8)	0.64 (0.2, 2.3)	0.47 (0.1, 1.8)	0.04 (0.0, 1.1)	0.51 (0.1, 2.4)	0.6 (0.1, 2.7)	0.4 (0.1, 2.6)	0.57 (0.2, 1.5)	0.50 (0.1, 2.1)	S		
2.79 (0.4, 17.2)	3.41 (0.6, 20.3)	1.13 (0.2, 6.3)	0.80 (0.1, 5.1)	1.80 (0.3, 10.2)	1.30 (0.2, 8.4)	0.77 (0.1, 4.6)	2.18 (0.3, 14.8)	0.57 (0.1, 6.06)	0.35 (0.02, 5.1)	2.56 (0.4, 18.5)	1.87 (0.2, 14.1)	0.16 (0.0, 6.0)	2.07 (0.2, 17.52)	2.42 (0.3, 20.3)	1.6 (0.1, 18)	2.27 (0.4, 13.8)	2.0 (0.2, 16.3)	4.0 (0.6, 28)	T	
0.56 (0.1, 2.9)	0.68 (0.1, 4.1)	0.23 (0.04, 1.3)	0.16 (0.03, 0.9)	0.35 (0.1, 2.1)	0.25 (0.04, 1.5)	0.15 (0.02, 0.96)	0.43 (0.1, 2.7)	0.11 (0.00, 1.1)	0.07 (0.01, 0.9)	0.51 (0.1, 3.6)	0.37 (0.1, 2.5)	0.03 (0.0, 1.1)	0.41 (0.05, 3.4)	0.48 (0.1, 3.5)	0.32 (0.03, 3.3)	0.45 (0.1, 2.8)	0.40 (0.1, 2.9)	0.80 (0.1, 5.7)	0.20 (0.02, 2.3)	U

Note: Lower triangle includes the findings of the network meta-analysis (posterior median odds ratios ORs plus 95% credible intervals) conducted in the Bayesian statistical package WinBUGS; upper triangle

includes the findings of the direct standard pairwise meta-analyses (OR plus confidence intervals) conducted using STATA

Statistically significant findings have been shaded (significance assessment made on data rounded to decimal places)

OR > 1.0 favours intervention compared with control.

Treatments: A Inactive control; B Conventional care; C Disc surgery; D Epidural injections; E Chemonucleolysis; F Non-opioids; G Intra-operative interventions; H Traction; I Manipulation; J Acupuncture; K Exercise therapy; L Passive physical therapy; M Biological agents; N Bed rest; O Opioids; P Education/Advice; Q Percutaneous discectomy; R Neuropathic pain modulators; S intra-discal injections; T Spinal cord simulations; U Radiofrequency treatment. **Abbreviations:** N number of studies included in conventional pairwise meta-analysis.

Table 2.3: Results (weighted mean difference, with 95% confidence intervals/credible intervals) of the network meta-analysis for pain intensity

			N=7, -8.11 (-19, 3)	N=1, -5.40 (-24, 13)	N=4, -6.26 (-15, 3)		N=1, 3.36 (-15, 21)		N=1, -25.00 (-42, -8)		N=1, -7.00 (-14, -0.4)	N=2, -9.91 (-43, 23)				N=1, -26.66 (-38, -15)	N=1, 13.00 (2, 24)	
-2.44 (-18, 13)	B	N=2, -7.0 (-12, -2)	N=2, -5.32 (-12, 1)							N=1, -2.00 (-12, 8)								
-9.54 (-25, 6)	-7.13 (-21, 6)	C		N=3, 2.36 (-8, 13)	N=1, 1.60 (-8, 11)	N=8, -5.17 (-12, 2)				N=1, 9.00 (-4, 22)								
-11.4 (-19, -4)	-8.96 (-23, 6)	-1.89 (-17, 14)	D		N=2, 18.01 (6, 30)				N=1, 4.00 (-6, 14)		N=1, 35.00 (-25, 95)	N=1, -9.3 (-23, 4.9)						
-6.62 (-25, 11)	-4.17 (-22, 14)	3.02 (-10, 16)	4.85 (-13, 23)	E				N=1, -0.63 (-15, 14)								N=1, 18.00 (8, 28)		
-2.54 (-12, 6)	-0.10 (-16, 16)	7.02 (-9, 23)	8.86 (-1, 19)	4.04 (-15, 22)	F						N=1, -40.50 (-58, -23)		N=1, 22.50 (10, 35)					
-14.64 (-33, 3)	-12.21 (-29, 4)	-5.12 (-14, 4)	-3.27 (-21, 15)	-8.11 (-24, 8)	-12.09 (-30, 6)	G												
-1.81 (-22, 18)	0.68 (-24, 26)	7.79 (-17, 33)	9.61 (-11, 31)	4.79 (-22, 32)	0.69 (-21, 23)	12.89 (-14, 40)	H				N=2, 3.19 (-13, 19)		N=1, 19.00 (8, 30)					
-7.24 (-39, 25)	-4.77 (-37, 27)	2.4 (-27, 32)	4.19 (-28, 36)	-0.63 (-27, 26)	-4.7 (-37, 27)	7.48 (-23, 38)	-5.42 (-43, 32)	I										
-14.82 (-34, 4)	-12.35 (-36, 11)	-5.21 (-29, 18)	-3.42 (-22, 15)	-8.25 (-34, 17)	-12.28 (-33, 8)	-0.13 (-26, 25)	-13.02 (-41, 14)	-7.58 (-44, 29)	J									
-2.62 (-25, 20)	-0.19 (-19, 19)	6.95 (-12, 26)	8.79 (-13, 31)	3.95 (-19, 27)	-0.07 (-23, 23)	12.07 (-9, 33)	-0.75 (-31, 29)	12.19 (-16, 41)	4.58 (-30, 39)	12.19 (-16, 41)	K							
-0.46 (-18, 18)	1.98 (-21, 26)	9.14 (-15, 34)	10.96 (-8, 31)	6.15 (-19, 32)	2.10 (-18, 23)	14.26 (-11, 40)	1.4 (-14, 18)	6.81 (-30, 44)	14.45 (-12, 41)	2.16 (-26, 31)	L							
-19.51 (-33, -6)	-17.03 (-37, 2)	-9.96 (-30, 10)	-8.10 (-22, 6)	-12.95 (-35, 9)	-16.99 (-32, -3)	-4.85 (-27, 17)	-17.72 (-42, 6)	-12.3 (-46, 22)	-4.71 (-27, 18)	-16.93 (-43, 8)	-19.09 (-42, 3)	M						
17.13 (-14, 49)	19.62 (-15, 55)	26.7 (-8, 62)	28.56 (-3, 61)	23.65 (-12, 60)	19.65 (-13, 53)	31.84 (-4, 68)	19.02 (-6, 43)	24.38 (-20, 69)	32.01 (-4, 69)	19.77 (-18, 58)	17.62 (-12, 47)	36.68 (3, 71)	N		N=2, -1.09 (-7, 5)			
4.92 (-16, 25)	7.41 (-18, 32)	14.51 (-10, 40)	16.32 (-5, 38)	11.47 (-15, 38)	7.44 (-13, 27)	19.61 (-7, 46)	6.74 (-22, 35)	12.15 (-25, 49)	19.75 (-8, 47)	7.58 (-22, 37)	5.42 (-23, 32)	24.45 (0.78, 48)	-12.19 (-50, 25)	O		N=2, -8.23 (-18, 2)		
16.22 (-19, 52)	18.71 (-20, 57)	25.83 (-13, 65)	27.61 (-8, 64)	22.8 (-17, 63)	18.77 (-17, 56)	30.99 (-9, 71)	18.11 (-12, 48)	23.39 (-24, 71)	31.07 (-9, 71)	18.9 (-23, 61)	16.69 (-17, 50)	35.76 (-2, 74)	-0.90 (-18, 16)	11.27 (-29, 53)	P			
11.6 (-19, 42)	13.97 (-16, 44)	21.14 (-7, 49)	22.94 (-8, 53)	18.14 (-6, 42)	14.14 (-17, 45)	26.2 (-3, 55)	13.38 (-23, 49)	18.8 (-17, 55)	26.36 (-9, 62)	14.17 (-19, 47)	11.96 (-24, 47)	31.04 (-2, 64)	-5.46 (-49, 37)	6.55 (-29, 43)	-4.62 (-52, 41)	Q		
-11.6 (-32, 8)	-9.13 (-34, 16)	-2.07 (-27, 23)	-0.21 (-21, 21)	-5.07 (-31, 21)	-9.01 (-30, 12)	3.06 (-23, 30)	-9.82 (-38, 18)	-4.37 (-42, 33)	3.21 (-24, 31)	-8.99 (-39, 21)	-11.13 (-39, 16)	7.93 (-16, 32)	-28.78 (-66, 8)	-16.52 (-33, 0.26)	-27.81 (-69, 12)	-23.11 (-59, 13)	R	
12.98 (-12, 38)	15.36 (-14, 45)	22.53 (-7, 52)	24.38 (-2, 51)	19.52 (-11, 50)	15.52 (-11, 42)	27.64 (-3, 58)	14.84 (-17, 46)	20.19 (-20, 61)	27.76 (-3, 59)	15.55 (-18, 49)	13.42 (-18, 44)	32.48 (5, 61)	-4.16 (-44, 35)	8.01 (-24, 40)	-3.28 (-46, 40)	1.4 (-38, 41)	24.55 (-7, 57)	U

Note: Lower triangle includes the findings of the network meta-analysis (posterior median weighted mean differences WMDs plus 95% credible intervals) conducted in the Bayesian statistical package WinBUGS; upper triangle includes the findings of the direct standard pairwise meta-analyses (WMD plus confidence intervals) conducted using STATA

Statistically significant findings have been shaded (significance assessment made on data rounded to decimal places)

WMD > 0 (representing reduction in pain) favours intervention compared with control

Treatments: A Inactive control; B Conventional care; C Disc surgery; D Epidural injections; E Chemonucleolysis; F Non-opioids; G Intra-operative interventions; H Traction; I Manipulation; J Acupuncture; K Exercise therapy; L Passive physical therapy; M Biological agents; N Bed rest; O Opioids; P Education/Advice; Q Percutaneous discectomy; R Neuropathic pain modulators; U Radiofrequency treatment;

Abbreviations: N number of studies included in conventional pairwise meta-analysis.

2.5.4 Between study heterogeneity, model fit and comparison with standard pairwise meta-analyses

Based on the Gelman-Rubin statistic, convergence occurred at around 6 to 8000 iterations for both outcome measures (global effect, pain intensity). The auto-correlation and history plots also showed good convergence. The goodness of fit of the models to the data, measured by the residual deviance, was found to be good for both outcomes (Appendix Volume II: Appendix A7).

The results of the evaluation of between-study heterogeneity showed a moderate to high level⁷ of statistical heterogeneity for many of the pairwise comparisons, as well as across all studies as a whole. The heterogeneity was greater for the analysis of pain intensity than global effect, with an I^2 statistic of less than 75% (i.e. moderate or less) for all but one pairwise comparison (epidural injections vs conventional care). The observed values for I^2 are presented in Figure 2.3. Heterogeneity did not improve when non-randomised studies were removed.

The comparison of the results from the network meta-analyses with that of the conventional pairwise meta-analyses showed broad agreement with slightly more discrepancies for the analyses of pain intensity. These discrepancies were greatest for comparisons that had very little direct evidence, such as biological agents.

2.6 DISCUSSION

This was the first systematic review that included all treatments for sciatica in the same analysis using a network meta-analysis method that includes indirect comparisons. The advantages of such analyses are that they can simultaneously compare more than two treatments in the same coherent analysis; provide relative effect estimates for all treatment comparisons, even those that have not been directly compared in head to head trials; enable the estimation of the probability that each treatment is best; and reduce the uncertainty in the treatment effect estimates.

2.6.1 Summary of results

In terms of overall response or global effect, there was a statistically significant improvement following disc surgery, epidural injections, non-opioid medication, intra-operative interventions, manipulation, and acupuncture when compared with inactive control or conventional care. Epidural injections, disc surgery, and intra-operative interventions were also statistically significantly better than percutaneous discectomy, chemonucleolysis, intra-discal injections, and radiofrequency treatment; with epidural injections, and intra-operative interventions also statistically significantly better than both traction, and exercise therapy. While biological agents and acupuncture had the highest probability of being best and had the largest effect estimates when compared with inactive control, these findings were associated with very wide credible intervals, reflecting the lack of information on these effect estimates.

In terms of pain intensity, there was a statistically significant reduction in pain following epidural injections and biological agents compared with inactive control, but there was no significant difference between disc surgery and inactive control. Biological agents had the highest probability of being best, and were also statistically significantly better than non-opioid medication, opioid medication, bed rest, and radio frequency treatment. However, when the analysis was restricted to RCTs, biological agents no longer had the highest probability of being best and were not found to be statistically better than any other treatments. When considering the magnitude of effect, bed rest, education/advice alone, percutaneous discectomy, and radiofrequency treatment were considerably inferior when compared with most treatment strategies, but these findings were not statistically significant and were associated with wide credible intervals.

Overall, the results of the sensitivity analyses excluding non-randomised studies showed broad agreement with the main analyses, with the findings generally becoming non-statistically significant due to broader credible intervals for the analyses restricted to RCTs and Q-RCTs. The most notable discrepancies occurred with treatment strategies that were associated with a small number of included studies, such as those reporting treatment with biological agents.

2.6.2 Findings of previous reviews

Previous reviews of non-surgical treatments have either found no evidence of effectiveness,^{124 125} conflicting evidence,^{126 127} or have reached different conclusions concerning the effectiveness of epidural steroid injections.^{124 126 128 129 154 155} A Cochrane systematic review of surgical interventions did not combine the results of four RCTs comparing discectomy with non-surgical treatment due to heterogeneity, and concluded that the results showed a temporary benefit of disc surgery at one year follow-up.¹³⁰ In that review the effectiveness of discectomy was justified by using informal indirect comparison of chemonucleolysis with placebo, and chemonucleolysis with disc surgery; chemonucleolysis was more effective than placebo and discectomy more effective than chemonucleolysis, therefore disc surgery was superior to placebo. Using the network meta-analyses, it was possible to make a more robust statement on disc surgery compared with placebo: disc surgery was statistically significantly better than placebo in terms of global effect but not for pain intensity.

2.6.3 Strengths and weaknesses

One of the main strengths of the network meta-analyses is the wide range of treatment strategies used to treat sciatica that were not only considered in the same review, but compared simultaneously in the same analysis. Another strength is that they were based on a systematic and comprehensive search of the literature up (until December 2009) that covered any therapeutic intervention for sciatica. However, it is acknowledged that these searches are not current, and as such, more recent relevant data are likely to have been excluded.

The RCT is widely regarded as the design of choice when assessing the effectiveness of healthcare interventions¹⁵⁶ and we acknowledge the controversy over the inclusion of non-randomised evidence.

Non-randomised studies were included in the search because some treatment approaches may not have been evaluated by RCTs, and also to increase the precision of the findings for interventions evaluated by a limited number of studies. Observational studies can have better external validity than RCTs^{38,39} and provide more generalisable findings. However, observational studies are likely to be affected by selection bias and confounding, and may therefore yield estimates of association that deviate from the true underlying relationship beyond the play of chance.¹⁵⁷ As it happens, most of the RCTs did not report the method of generating the randomisation sequence or allocation concealment, which means that selection bias or confounding might still be present. Excluding the non-randomised studies in a sensitivity analysis did not affect the structure of the network, as the overall findings of both series of network meta-analyses were similar, although less precise for the analyses of RCTs.

Network meta-analysis methods enabled us to go beyond the pairwise comparisons reported in previous systematic reviews. They allowed us to simultaneously compare all the available treatment strategies for sciatica and provided estimates of relative treatment effects for all conceivable comparisons, even those where there was no direct evidence available. However, the small number of relevant studies for some comparisons, statistical heterogeneity (*within pairwise comparisons*), and potential inconsistency (*between different pairwise comparisons*) within the networks means that the encouraging results for interventions, such as biological agents should be interpreted with caution.

In order to answer the question of which is the optimum treatment for sciatica and provide generalisable findings, we were interested in the average treatment effect of each treatment approach (to represent the diversity used in clinical practice). We therefore pooled clinically heterogeneous studies. A random-effects model was used to pool the data, which is based on the assumption that different studies assessed different, yet related, treatment effects. However, included studies also varied in study design and risk of bias (methodological diversity). There was considerable ($I^2 \geq 75\%$)¹⁵² statistically significant between study heterogeneity present for a number of comparisons within the pairwise meta-analyses, especially in the analyses of pain intensity, and it was not possible to ascertain how much was due to clinical or methodological diversity. This needs to be taken into consideration in future work.

The network meta-analyses were based on a single time-point, outcome data closest to six months, which may be considered as a limitation of the analyses. The HTA monograph⁷ included an assessment of each treatment strategy at short (≤ 6 weeks), medium (> 6 weeks to ≤ 6 months) and long (> 6 months) term follow-up, but this evaluation was based on multiple pairwise analyses, with each analysis needing to be interpreted independently. Further research is needed to incorporate multiple time-points within the network meta-analyses in order to incorporate data at different follow up periods.

For the pain intensity outcome, where the standard deviations were missing (and could not be estimated from the published data) these were imputed using the weighted mean standard

deviation^{146 158} for each treatment strategy (11 studies). This is based on the assumption that the variance is similar between studies and the data are not skewed.¹ The median was also used to represent the mean for two studies. We considered that it was better to use these methods in order to incorporate more of the evidence base, as ignoring the findings of these studies may induce bias in the summary effect estimate.¹⁵⁸ Furukawa, et al.¹⁴⁶ have previously shown that it is safe to borrow standard deviations from other studies.

There were insufficient studies to explore the presence of publication or reporting bias for most treatment comparisons. However, a funnel plot of studies comparing surgery and chemonucleolysis showed no evidence of publication bias (see Appendix Volume II: Appendix A2).⁷ The benefit (or effectiveness) of different treatment strategies for sciatica should be considered along with potential harms. Although the present paper does not report adverse effects, they are reported elsewhere.⁷

The network meta-analyses relied on the key assumption that the relative treatment effect of one treatment versus another is the same across the entire set of studies.^{56 132} The use of random-effects models meant that it was assumed that the common distribution of effects was the same across all sets of studies. A further assumption made in the analyses was that the relative efficacy of different treatments is the same at different stages in the care pathway. Pragmatically, sciatica is often treated with a stepped-care approach starting with conservative treatments, such as non-opioid medication, progressing as necessary to more invasive treatments such as epidural injections or surgery. This means that the population of patients treated with conservative treatments was likely to differ from those treated with invasive treatments, resulting in confounding and inconsistency within the network. Although descriptive characteristics were generally poorly reported by included studies, there was a trend for studies evaluating invasive treatments to report a history of previous treatments and include patients with a diagnosis confirmed by imaging, and for studies of conservative treatments to limit inclusion to patients with acute sciatica. Due to the breadth of the review and the novel and speculative use of network meta-analysis methods, we were unable to incorporate stepped-care approaches in the network meta-analyses. The optimum sequence of treatment modalities and what sequence is best for which patients is therefore not yet known and awaits further analysis.

2.7 CONCLUSIONS

The use of network meta-analyses provided new information on the relative effectiveness of treatments for sciatica. This can help clinicians and patients in shared decision-making, as well as providing data for healthcare policy development. The findings provide support for the effectiveness of some common therapies for sciatica such as non-opioid medication, epidural injections and disc surgery. They also suggest that less frequently used treatments such as manipulation and acupuncture, and experimental treatments such as cytokine modulating biological agents, may be considered. The findings of this review do not support the effectiveness of opioid medication, either for pain intensity or global effect. Furthermore, there is no support for the effectiveness of numerous

other interventions such as bed rest, exercise therapy, percutaneous discectomy or traction. The lack of support for education/advice should not be taken to imply that patients should not be given information or advice; rather it is not an effective treatment if delivered alone.

Further research is needed to confirm or refute these findings where we found limited evidence, and to explore the impact of heterogeneity and the range of clinical questions most suited for the use of network meta-analyses. There is also scope to develop more sophisticated methods, such as building on the confidence profile method,³⁹ bias-adjusted results,¹⁵⁹ or Bayesian statistics,³⁸ to incorporate information relating to differences in study design or internal and external validity in the network meta-analyses, as well as data on multiple follow-up periods.

The issue of how best to estimate the effectiveness of treatment approaches according to their order within a sequential treatment pathway remains an important challenge. It is also likely to need new network meta-analytic methods to achieve this.

2.8 IMPLICATIONS OF THE NETWORK META-ANALYSIS OF SCIATICA TREATMENTS FOR METHODS DEVELOPMENT IN EVALUATING TREATMENT SEQUENCES

2.8.1 Implications of the findings in terms of assessing heterogeneity to inform the need to consider treatment sequencing

The review demonstrated that ignoring treatment sequencing is likely to lead to important clinical heterogeneity, as illustrated in Chapter 1, Section 1.5.2. Meta-analysis is based on the fundamental assumption that treatment effects from multiple studies are similar enough to pool the data. The assessment of heterogeneity is therefore, an important part of meta-analysis. Heterogeneity is caused by patient and study characteristics acting as treatment effect modifiers and, although it does not cause bias in a pairwise-meta-analysis, it can mask important differences and render the pooled estimate meaningless.¹⁶⁰ For example, two clinical trials comparing treatment A with B (AB), may be pooled in a conventional pair-wise meta-analysis, where one trial includes participants with late stage disease and the other includes patients with early stage disease. If there is a treatment interaction with disease stage, then the pooled relative treatment effect will exhibit heterogeneity and be difficult to interpret. If two trials included a different comparator, for example AB and BC, and included in an indirect treatment comparison, then the derived AC estimate is also likely to be uninterpretable due to the heterogeneity, referred to here as inconsistency. Heterogeneity across different pair-wise contrasts within a network meta-analysis can cause confounding bias.^{160 161} This occurs when a covariate has an impact on the treatment effect (i.e. it's an effect modifier) and is also associated with the type of treatment comparison.¹⁶⁰ For example treatment B may be a common treatment for early stage disease whilst treatment C may be generally reserved for treating late stage disease, and the clinical trials might be designed to reflect this. The validity of network meta-analysis is based on the underlying assumption that there is no imbalance in the distribution of effect modifiers across different types of direct treatment comparisons.¹⁶⁰

The presence and extent of the heterogeneity in both pair-wise and network meta-analysis of sciatica treatments provides an indication of the potential for previous treatments to be an important effect modifier. It is also possible that the relative efficacy of different treatments may vary at different stages of the care pathway. The validity of combining studies of single treatments used at different points in the treatment pathway is called into question.

Potential sources of heterogeneity within network meta-analysis can be explored using meta-regression or subgroup analysis. I intended to explore the impact of previous treatments using meta-regression analysis. However, the paucity of the data precluded the use of any meaningful subgroup analysis or meta-regression.

The limitation of the evidence base is illustrated in more detail in the Appendix Volume II (Appendix A2), including examples of pairwise meta-analyses and a funnel plot included in the health technology assessment of sciatica treatments. Sciatica resolves spontaneously in many cases, but when it becomes a more chronic condition the likelihood of improvement diminishes with both time and the failure of each successive treatment. The effectiveness of individual treatments may, therefore, depend as much on how early they are utilised after the onset of sciatica, as on the number of prior unsuccessful treatments used. It is also recognised that up to 30% of patients have sciatica that does not respond to treatment.¹¹⁹ Any future evaluation of treatment sequencing would need to consider not only the number and type of previous treatments, but also the timing of treatments and disease duration. Future evaluations would also need to take into account the fact that successive treatments are unlikely to have a 100% success rate due to enrichment with refractory patients. I revisit this in Chapter 6-7 whilst evaluating the methods used to account for these issues within the economic evaluation of sciatica treatments (Fitzsimmons, 2014).¹¹⁵

These issues are likely to be generalisable to other chronic conditions. The assumption of transitivity, i.e. the patient and study characteristics that act as effect modifiers are similarly distributed within and across pair-wise comparisons,^{160 162} may also be implausible where some treatments in clinical practice are available for certain patient populations only. For example, a new treatment may initially only be licenced for those who fail to respond or are intolerant to the current conventional treatment.

2.8.2 Implications relating to the meta-analytical methods used

Network meta-analysis can be performed either under a Bayesian or frequentist framework, and several models have been proposed under both.¹⁶³ The method used for evaluating sciatica treatments, which included a random-effects Bayesian hierarchical linear model, provides some useful information on the advantages and disadvantages of this particular approach. Potential insights for future methodological developments for evaluating treatment sequences were also illuminated.

Network meta-analysis, multivariate meta-analysis and meta-regression represent active areas of multi-parameter meta-analytical research.¹⁶⁴ These methods could provide the basis for a novel

method for developing summary effect estimates that are conditional on positioning in a treatment sequence. What may be important here is how treatment sequences are conceived or evaluated. This could include:

- i. Investigating the clinical effectiveness of whole sequences
- ii. Estimating the effect of individual treatments conditional on positioning in the pathway
- iii. The evaluation of previous treatment as an important effect modifier

The extension of current meta-analytic approaches may not be able to account for more than one specific immediate prior treatment, unless the sequence of previous treatments can be represented as a single estimate, which could lead to the need to make strong simplifying assumptions.

The analyses of sciatica treatments were performed using the free software, WinBUGS, for which various model programme codes and worked examples are available.¹⁶⁵ These are based on a generic model structure developed for both random and fixed-effects syntheses, using linear regression, which can be easily adapted for various data types by changing only the likelihood and link function. Other Bayesian software programmes are also available, and frequentist models developed using generalised lineal mixed models can be implemented using commercial programmes.

The advantage of the Bayesian approach is that it provides a flexible statistical framework to account for the complex data structure of multi-arm trials, and is also straightforward to extend to shared parameter models where studies report outcomes in different formats but from a common underlying model.¹⁶⁵ Both were useful in the sciatica review, where the assessment of global effect included a multi-arm trial, and the assessment of pain required a model for combining both study- (contrast) and arm-level data (Appendix Volume II: Appendix A3: winBUGS codes). The disadvantage of the Bayesian approach is that statistical knowledge is needed in order to: choose the appropriate codes, adjust to account for the type of data available; put prior distributions on all the parameters of the model; decide which nodes to monitor; and know how to interpret the output. Appropriate disparate initial values must be chosen to start the Markov chain Monte Carlo simulation. At the end, convergence, model fit, and consistency must be checked (Appendix Volume II: Appendix A7). However, ongoing work is being done on developing computer software that will automate the process of performing network meta-analysis in both the Bayesian and frequentist framework, so that users can perform the analysis with minimal programming and computational effort.¹⁶³

The random-effects meta-analysis model was used as it allows for the heterogeneity, or unexplained variation, in the effect estimates among studies. The assumptions underpinning the fixed-effects model would not be plausible for investigating the diverse treatments used for sciatica. A fixed-effects meta-analysis model estimates a single effect that is assumed to be common to every study, whilst a random-effects model estimates the mean distribution of effects¹⁶⁶ with a standard deviation that represents the unexplained between-study heterogeneity (τ). However, the random-effects model has the following limitations:

- i. The weights are more evenly distributed in a random-effects model, and as the heterogeneity increases, less weight is given to larger studies, which potentially provide more precise estimates of the true effect (i.e. small and larger studies are given almost equal weight).
- ii. A random-effects model requires an estimate of the between-study variance, which may be imprecise or problematic when the data is sparse. Where the number of studies is low the estimated between-study variance, when using the frequentist framework, has poor precision and is likely to be biased,¹⁶⁶ whilst in a Bayesian analysis with limited data, using a non-informative prior distribution for the between-study standard deviation will give implausible posterior distributions.¹⁶⁷ We used wide uniform prior distribution for tau. An alternative approach would have been to use an informative prior distribution based on data from outside the current set of studies, or expert opinion.
- iii. The findings of the analysis can be difficult to interpret and generalise when there is extensive heterogeneity, especially as the resultant credible or confidence intervals do not generally account for the full extent of the heterogeneity in the variance estimates, above the specified nominal level.

A number of methods have been developed to overcome the limitations of random-effects models, including IVhet method (inverse variance heterogeneity model),¹⁶⁸ the approach developed by Henmi & Copas,¹⁶⁹ and the generalised weighting regression.¹⁷⁰ These need to be taken into account when evaluating treatment sequencing where excessive heterogeneity and sparse data is likely to be an issue. However, the current need to tackle these issues may mean that there is little enthusiasm for further developing a novel method to account for positioning in the treatment sequence.

It is also important to consider that the confidence interval from a random-effects model does not reflect the true extent of the variability between studies, which is provided by a predictive interval. The confidence interval from a random-effects model provides an estimate of the precision of the mean estimate (i.e. it represents only the error in estimating the mean, which is dependent on the number of studies), whilst the prediction interval describes the distribution of the true effect size (and is based on both the error and true variance).¹⁷¹ The routine reporting of prediction intervals is now strongly recommended to help interpret the findings for decision-making, as they reflect the variation in treatment effects over different settings and include what effect is to be expected in future patients.¹⁷² However, the resultant width of prediction intervals will increase with heterogeneity and the pooling of fewer studies,¹⁷³ which are likely to make the overall findings difficult to interpret. When substantial heterogeneity is present, further investigation of clinical heterogeneity using meta-regression and subgroup analyses are likely to be required.

The network meta-analysis was based on the hierarchical modelling approach, which is an extension of meta-regression methods and could potentially be extended to account for treatment sequencing. In hierarchical methods, the information in the meta-analysis stems from different levels, for example studies at the higher level, and participants at the lower level. It is important to note that the

hierarchical modelling approach preserves the clustering of patients within each study, and that a modelling framework which breaks the original randomisation or treatment comparison within each study is inappropriate.¹⁶⁴ I elaborate on this in the Appendix Volume I (Section A), and return to this point in Chapter 5 (Section 5.9.9.2). In terms of future development, the concept of hierarchical modelling could potentially be extended to account for positioning in the treatment sequence, or previous treatments used as another level. These methods have already been extended to incorporate both aggregate and individual participant level data within the same meta-analysis, and to include data from different study designs.^{164 174 175}

The conventional pair-wise meta-analysis is univariate, in that it only considers the one-dimensional treatment effect parameter. Multivariate meta-analysis, on the other hand, enables the inclusion of multiple outcome measures, for example progression-free survival and overall survival in studies of cancer treatments, whilst at the same time allowing for their correlation.¹⁷⁶ These methods could also potentially be used to incorporate multiple treatment lines. The network meta-analysis of sciatica treatments could have been extended to include multivariate analysis, allowing the inclusion of a broader evidence base; for example, including data from multiple follow-up periods, and not just the data reported for the duration closest to 6 months, which may have improved the precision of our findings. However, the potential challenge of any future developments of multivariate meta-analysis is obtaining and estimating within- and between-study correlations. Within-study correlation refers to the association between two parameter estimates in the same study, whilst between-study correlation indicates the strength of association between true parameter values across studies and is caused by differences across studies in patient and study characteristics modifying the true values in a related way.¹⁶⁴

2.9 THE NEXT STEP

Network meta-analysis is becoming an important tool for informing clinical and policy decisions. The rapid growth and popularity of the method is fuelled by the increased availability of multiple treatments for many clinical conditions.¹⁷⁷⁻¹⁷⁹ However, where multiple technologies are available, treatment sequencing is also likely to be an important consideration with patients moving on to the next available option when they do not experience resolution on their current therapy. Evidence is therefore needed to inform the optimal ordering of these interventions in practice.

The next step is to identify what quantitative evidence synthesis methods other researchers have used for assessing the effectiveness of treatment sequences, and what challenges they faced in implementing them. The review of sciatica treatments highlighted that any quantitative evidence synthesis methods for evaluating treatment sequencing are likely to need to take into account:

- i. the number and type of previous treatments,
- ii. the timing of treatments and disease duration, and
- iii. the proportion of patients with refractory disease.

The potential meta-analytic methods used are also likely to depend on how treatment sequences are conceived or evaluated, for example the comparison of whole sequences to develop a summary effect estimate for a treatment conditional on its positioning in a sequence, or to account for previous treatment as an effect modifier (or the cause of heterogeneity).

CHAPTER 3: A REVIEW OF QUANTITATIVE SYNTHESIS METHODS FOR ASSESSING THE EFFECTS OF TREATMENT SEQUENCES

3.1 CHAPTER OVERVIEW

A review of the existing methodology is warranted to inform future practice and research in evidence synthesis in the context of treatment sequencing. This chapter presents the methods used in that review including their context, justification and a detailed description. The results of this review are presented in later chapters.

3.2 INTRODUCTION

Methods for evidence synthesis that produce the least biased estimates of treatment sequencing effects are required to inform clinical and policy decision making. A comprehensive review of quantitative evidence synthesis methods used to date, as reported in the literature, was conducted. These were mainly based on studies evaluating single treatments and incorporated diverse study designs.

The best study design for providing robust estimates of clinical effectiveness of an intervention is the randomised controlled trial (RCT). The reason for this is explored in more detail in the Appendix, Volume I (Section A). However, as discussed in Chapters 1, RCTs tend to focus on single treatments and non-randomised studies may be used as an alternative source of data for sequencing effects. A summary of the advantages and limitations of non-randomised studies, plus a brief overview of the methods used for assessing the risk of bias within individual studies included in the evidence synthesis, and the credibility of the overall body of evidence are provided in Appendix Volume I (Section A). Randomised controlled trials of treatment sequences may be available for some clinical scenarios. Clinical trial designs for evaluating treatment sequences are discussed in the Appendix Volume I (Section B), which includes a summary of an innovative RCT design for evaluating dynamic treatment sequences.

3.3 AIM AND OBJECTIVES

The overall aim of the review of methods was to identify and appraise quantitative evidence synthesis methods used to estimate the treatment effect of whole sequences, or single interventions conditional on position in a treatment sequence.

Specific objectives were to identify, summarise, and appraise:

- i. Studies describing theoretical or proposed novel quantitative evidence synthesis methods to assess the effects of a treatment sequence as a whole, or single treatments used within the context of a treatment sequence

- ii. Systematic reviews or economic evaluations that implement such methods, or illustrate their application

The main aim was not to identify every study that used each approach, but to identify the breadth of approaches developed for evaluating treatment sequencing.

The review of methods was conducted with the intention of developing a framework for undertaking quantitative evidence synthesis of the effectiveness of sequential treatment options within the context of a health technology assessment or similar process, as outlined in Section 1.6. Health technology assessment and the clinical guideline process generally incorporates a clinical and economic evaluation. The review of methods considered both clinical and economic evaluations, however the focus of the thesis and the proposed framework is on clinical effectiveness, and does not consider the impact of treatment sequencing on costs.

Meta-analysis and decision analytic modelling were reviewed as two distinct categories of quantitative evidence synthesis methods. The findings of the review of meta-analytic approaches is presented in Chapter 5 and decision analytic modelling approaches in Chapters 6 and 7. Economic evaluations that implemented decision analytic models of treatment sequences were initially assessed to identify meta-analytic methods used to develop the model parameter estimates of clinical effectiveness. Preliminary findings showed that in the presence of limited or absent data regarding treatment sequencing effects, many modelling studies made simplifying assumptions about the sequencing effects. The analytic judgements or assumptions made by included modelling studies regarding sequencing effects were subsequently reviewed as a separate methodological approach in Chapter 6. The findings of the review of modelling approaches used to evaluate treatment sequences are presented in Chapter 7. An introduction to decision analytic modelling and decision uncertainty is provided in Chapter 6, and an overview of modelling approaches used in health economics in Chapter 7.

3.4 METHODS

A systematic pragmatic review of the literature relevant to quantitative synthesis methods used for evaluating treatment sequencing was conducted.

3.4.1 Literature search

In Cochrane or conventional systematic reviews of the clinical effectiveness of pharmacological interventions, the majority of relevant intervention studies can be identified by searching MEDLINE, Embase and Cochrane Controlled Trials Registry (CCTR).¹⁸⁰⁻¹⁸²

A conventional systematic search was considered insufficient for the current review, given the scarcity of appropriate indexing terms for methodological research^{57 156 183} and the need for prior knowledge of

the terms used by the current methodological literature. Structured electronic database searches have been shown to miss important evidence for systematic reviews of more complex and heterogeneous evidence.^{2 184} Informal approaches and ‘snowball’ techniques are more likely to produce a greater yield of relevant studies.² Furthermore, whilst conventional systematic review searches aim to be as extensive, or sensitive, as possible in order to identify all the relevant studies and reduce the risk of bias, they are usually bound by a narrowly defined research question in order to make the review manageable. However, the more narrow and tightly focused a research question is, the less likely it is to allow an inclusive and iterative approach to the identification and incorporation to evidence.² Reviews that address a broader question, covering multiple and disparate sources are not affected by biases in the same way as effectiveness reviews.^{2 183 184} The aim of such reviews are to identify the breadth of information or evidence that is relevant to the review, and as such the extent of the searches are concerned more with saturation than sensitivity. Searches are deemed to be complete when further efforts to identify information would not add to the analysis.² A similar approach was adopted for the current review, with a more pragmatic and iterative process being used, based on ‘snowballing’ and the use of ‘proximal cue’ or ‘information scent searching’ as outlined by Kaltenthaler *et al.*:¹⁸³

“Following trails of potentially relevant sources can provide an alternative approach to searches that cast the net wide using broad keyword strategies. One form of this technique, sometimes referred to as ‘snowballing’, is used in systematic review searching, whereby trails of cited references are followed prospectively and retrospectively from a single or a series of index sources. However, it is also possible to use any information from the source documents as the starting point of an information trail. As such, a starting point might take the form of an idea or concept, an author or a set of keywords. The starting point or points act as ‘proximal cues’ which can be followed to further, similar, potentially relevant information. In the field of information seeking behaviour this is referred to as following ‘information scents.’”

3.4.2 Sources searched

A number of different approaches were used to identify relevant methodological studies. Initial ‘hand’ searches included the following sources: internet search engines; websites of specific organisations including NICE; electronic journals; agendas of on-line proceedings for conferences or meetings; previously known references of relevant or interesting studies; known author searches; and the use of electronic tracking to forward-track citations. Finally, reference database searches were conducted.

The search process started with key papers identified during an initial scoping search for the project, as well as Internet searches and investigators’ or supervisors’ personal collections. The key words, references and author details from these papers were then used to identify further studies using information scent searching techniques.

Following these initial searches, a search strategy was developed and used to search the following electronic databases:

- MEDLINE
- Embase
- Cochrane Library
 - Cochrane Database of Systematic Reviews
 - Cochrane Methodology Register
 - Database of Abstracts of Reviews of Effects (DARE)
 - Health Technology Assessment (HTA) Database
 - National Health Service Economic Evaluation Database (NHS EED)

The search strategy was designed to be specific rather than sensitive, with published filters used for meta-analysis, economic evaluations, and evidence based decision making process that utilise quantitative evidence synthesis.

The following search strategy was used for MEDLINE and adapted for the remaining databases

1. Meta-analysis as topic/
2. Meta-analysis/
3. meta-analysis.pt
4. (meta-analys* or metaanalys*).sh
5. (meta-analys* or metaanalys*).ti,ab
6. (meta-regression).ti,ab.
7. systematic review.ti,ab.
8. Cochrane.ti,ab.

9. health care costs/
10. cost-benefit analysis/
11. Cost benefit analys*.ti,ab
12. ((cost effectiveness or cost utility or cost consequence*) adj2 analys*).ti,ab.
13. Economic evaluation.ti,ab.
14. sensitivity analys*.ti,ab

15. Health technology assessment.ti,ab.
16. Guideline.pt
17. Practice guideline.pt
18. Clinical guideline*.ti,ab.
19. NICE guid*.ti,ab.
20. Comparative effectiveness research.ti,ab.

21. OR/ 1-20

22. (Sequen* adj2 (treatment* or therap* or drug* or medication* or regimen* or intervention* or biologic*)).ti,ab

21 and 22

The searches were conducted without language restriction, however, articles not published in the English language were only considered on a case by case basis, due to the time and financial constraints of translation. This meant that it was possible to quantify potentially relevant foreign language studies, and then include any that appeared to be highly relevant. The electronic database searches were dated from inception to August 2013.

The use of existing literature reviews to identify further relevant studies

A number of systematic reviews of economic evaluations investigating treatment sequences were identified during the searches, one of which evaluated the use of biological agents for psoriasis,¹⁸⁵ and three investigated biological agents in rheumatology.^{28 186 187} Five other reviews of economic evaluations investigating methods or modelling techniques used to investigate biological agents for rheumatoid arthritis were also identified.¹⁸⁸⁻¹⁹³ These reviews were used to identify any further relevant studies for the current review.

3.4.3 Assessing study relevance

Tightly defined inclusion criteria were not used to assess the relevance of identified studies, as there was no “right” or “wrong” evidence as such, just an interpretation of what was relevant to the scope of the methodological problem.¹⁸³

The review included studies that either applied or developed quantitative evidence synthesis methodology to evaluate the clinical effectiveness of treatment sequencing as part of secondary research. Primary studies evaluating treatment sequences were excluded.

The review included any type of meta-analytic technique incorporating, but not limited to, pair-wise meta-analysis, meta-regression, and network meta-analysis. Studies that used qualitative or narrative evidence synthesis methodology were excluded. Studies, or reviews, that limited inclusion to RCTs of treatment sequences were thought to be unlikely to provide any ‘new’ method of synthesising the evidence and were therefore not reviewed in great depth. This is despite the synthesis of treatment sequencing RCTs being considered the ideal situation, as discussed in the Appendix, Volume I (Section A). It was also considered that there would be no or insufficient relevant sequencing RCTs available for most clinical scenarios.

Decision analytic modelling techniques developed to evaluate treatment sequencing, whether conducted as part of an economic evaluation or not, were also considered relevant and included in the methodology review. Modelling studies aiming to evaluate the effectiveness of single treatments, but incorporated the impact of further downstream treatments were only included if they specifically

modelled sequencing effects, or the individual effect of each subsequent treatment or treatment-line. Decision analytic modelling studies published in abstract form were not considered for inclusion, unless they represented a novel modelling approach, which was not covered by any of the studies published in full. Economic evaluations based on a single RCT were excluded.

The review focused on treatment switching based on a clinical assessment of either lack of effect or adverse reaction (whilst acknowledging these will have a differential effect on the efficacy of the subsequent treatment). Studies evaluating the effectiveness of planned sequential administration of combined therapy, where the addition of treatment is not administered based on clinical assessment, were excluded. The sequential administration of combination therapy is a different type of decision framework as the sequencing is generally fixed; switching is not dependent on the treatment being ineffective, not tolerated or following adverse effects; neither previous treatment effect, nor the passing of time and its differential effect on risk are relevant. Furthermore, reviewing the evidence for combination therapies is less challenging, as they are generally based on sequencing trials. Where sequencing is considered an important factor in the administration of combination therapy, it is generally feasible and straightforward to compare fixed sequences as part of a RCT and conventional meta-analytic techniques can be used to pool the data from multiple trials, for example the administration of triple therapy for H-pylori infection.¹⁹⁴ However, relevant papers that were retrieved in full were scanned to ensure that they did not report methods that could potentially be used to inform future novel evidence synthesis methods for treatment sequencing.

Studies identified via the initial searches that appeared to meet the inclusion criteria were listed in a table. The titles of relevant references identified via the subsequent reference database searches were then cross referenced with these, and any new studies retrieved in full. Iterative searches were continued throughout the review, and cross referenced with the list of included studies until no new methods were identified. This is analogous to reaching the point of 'saturation' in qualitative research.

3.4.4 Study classification and data extraction

Studies were categorised according to whether they investigated clinical- or cost-effectiveness. All included economic evaluations were based on decision analytic modelling.

The type of data extracted for all included studies, where available, included:

- i. Underlying 'methodological' problem or clinical question
- ii. Clinical condition (population) and intervention of interest
- iii. Treatment sequences that were evaluated
- iv. Summary of the method used to evaluate treatment sequencing
- v. Key assumptions used, and whether they were assessed
- vi. Type of outcome measure used
- vii. Type of evidence available or data required

The characteristics of included studies were described and contrasted using summary tables and matrices, and the findings synthesised narratively. A list of key methodological approaches was developed for each section, relating to meta-analytic techniques, simplifying assumptions, and modelling approaches. These are discussed in detail in separate chapters (5, 6 and 7). Each approach was summarised and key findings from the evidence tables discussed. These were subsequently used for developing a framework for the consideration of quantitative synthesis in the context of evaluating treatment sequencing as part of a health technology assessment, systematic review, or guideline development process (Chapter 8).

3.5 QUANTITY OF LITERATURE INCLUDED IN THE REVIEW

The reference database searches, after de-duplication, identified 756 references, of which 94 were retrieved in full. Sixty-two studies met the inclusion criteria, but 23 were published as abstracts only (Figure 3.1). The first author of the only two abstracts that were considered highly relevant were contacted by e-mail to check if the study had since been published in full. One replied, stating that a full report was unavailable,¹⁹⁵ and the methods were still theoretical. This study (Briggs, 2014)¹⁹⁵ is discussed further in the review of modelling approaches (Chapter 7, Section 7.3.4). The study presented in the second abstract¹⁹⁶, was subsequently included in the review (Heeg, 2015) as a full publication (PhD thesis)¹⁹⁷ was later identified via the on-going internet searches. Two studies were unobtainable, and two were presented in Italian but did not appear to contribute sufficient original content to warrant translation, and were therefore not included. Thirty-six studies published in full were included after collating multiple publications, and a further 52 relevant studies were identified by hand. The fact that only 41% of the 88 included studies were identified by the electronic databases confirms the limitations of relying on a conventional search strategy for undertaking a review of methods. However, some studies identified by hand, were published after the end date of the electronic database searches (ie. August 2013). This is discussed further under Section 8.3.

Twenty two studies were included in the evaluation of meta-analytic approaches used for evaluating treatment sequences.^{12 17 20 21 198-215} However, some of these studies were considered relevant in fairly broad terms, such as providing examples of how the limited evidence base precluded the evaluation of treatment sequencing, or representing the use of stratified analysis by line of therapy, which could potentially provide a building block for future methods development. These approaches were initially not considered pertinent to the review but, because of the dearth of relevant methods identified, a post-hoc decision was made to include these studies as examples of simplifying methods. This provided a more comprehensive list of the approaches actually used for evaluating treatment sequencing in general, rather than limited to novel methods for developing sequence-specific summary effect estimates.

Seventy studies were included in the review of decision analytic models.^{17 18 115 199 204 207 216-279} Four studies (NICE CG131; Rodgers, 2011; Connock, 2006; Hind, 2008) were included in both the

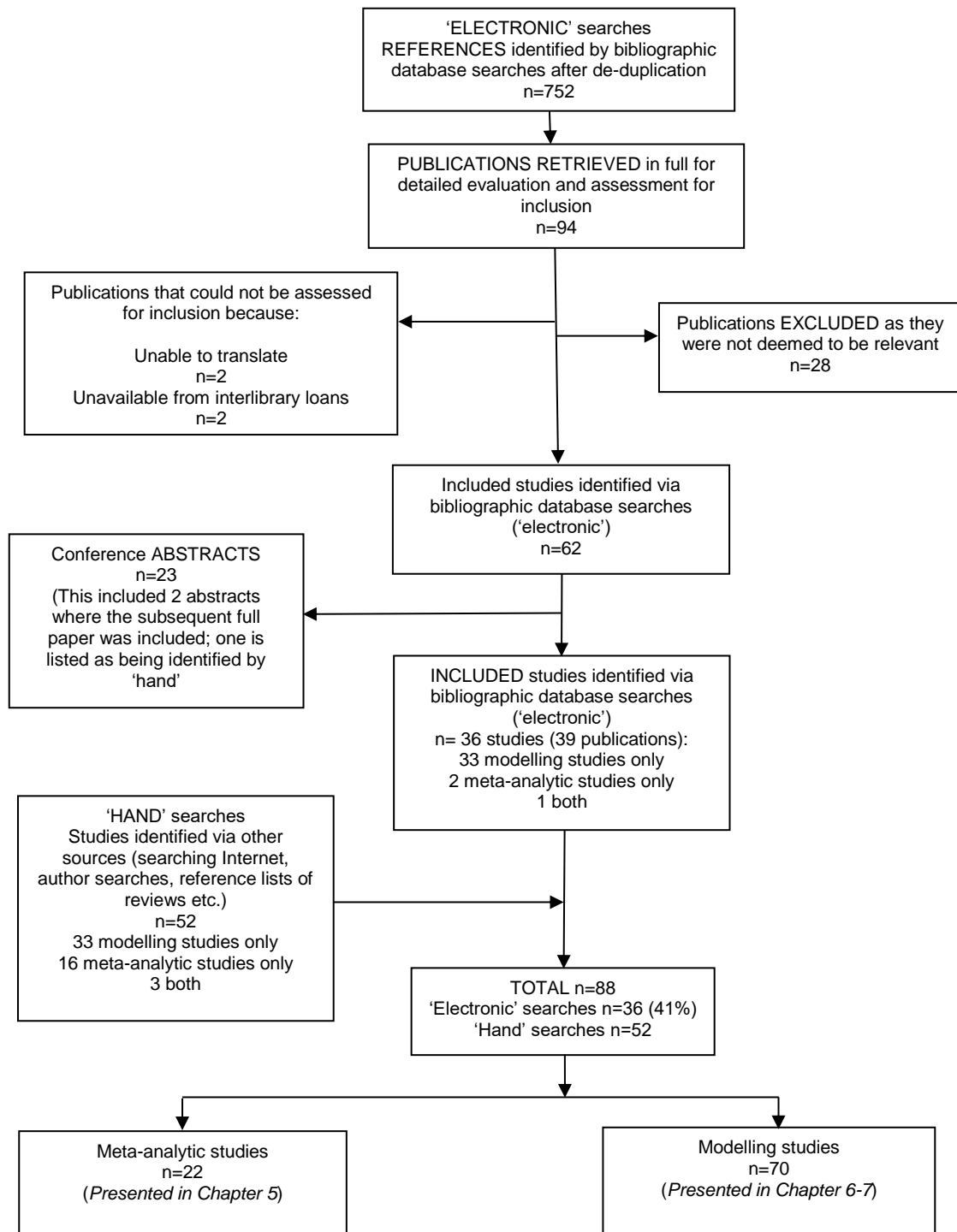
evaluation of meta-analytic techniques and modelling approaches.^{17 199 204 207} All were commissioned by NICE, and are introduced in Chapter 4, which explores the challenges faced by NICE and health technology assessment agencies in terms of the evidence base for evaluating treatment sequences.

Four modelling studies included in the review of modelling approaches were not included in the review of simplifying assumptions. These include two (NICE CG131; Hind, 2008: NICE TA79)^{204 207} that obtained data on clinical effectiveness from sequencing trials, and two (McEwan, 2010, and Launois, 2008),^{249 255} that did not evaluate the clinical effectiveness of treatment sequences, but were included as they provided examples of specific modelling techniques. One was a budget impact study (Launois, 2008) and the other (McEwan, 2010) evaluated the impact of treatment strategies on health related quality of life improvements associated with different hypoglycaemia profiles, rather than efficacy variables. This last study is closely linked to the study by Erhardt *et al.*,²³⁷ which was included in the review of assumptions.

Forty-seven (54%) included studies investigating the use of disease-modifying antirheumatic drugs (DMARDs), including biological agents, for the treatment of inflammatory arthritis including rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. Thirteen (15%) were related to oncology. The remainder addressed epilepsy (5%), psoriasis (5%), depression (3%), glaucoma (2%), schizophrenia (2%), type 2 diabetes mellitus (2%), human immunodeficiency virus (HIV) (2%), neuropathic pain, postherpetic neuralgia, sciatica, fibromyalgia, chronic Hepatitis B infection, Crohn's disease, onychomycosis, and spasticity. Most studies, including all of the rheumatology studies, involved sequences of drug treatments, but some also considered other interventions, for example surgery for sciatica.

The range and type of disease conditions where treatment sequencing is likely to be an important issue is explored in more detail in Chapter 4. A summary of the clinical context and potential limitations of the available evidence base is provided for three of the most commonly evaluated clinical conditions, inflammatory arthritis, advanced or metastatic cancer, and epilepsy, in the Appendix Volume I (Section C). I also re-visit these in Chapter 4 (Section 4.4.4) as a number of NICE technology appraisals or clinical guidelines for these three indications were included in the review of methods. Many of the rheumatology and cancer studies evaluated biological therapies. Definitions and brief descriptions of biological therapies are also provided the appendix (Section C2), as a potential class effect of these drugs appears to be an important consideration for treatment sequencing. The optimal outcome measure for evaluating sequencing for cancer treatments is also explored in Appendix Volume I (Section C4.2).

Figure 3.1: Flow diagram showing the number of references identified, publications retrieved for assessment, and studies included in the methodology review



3.6 CHAPTER SUMMARY AND NEXT STAGE

This chapter outlines the methods used to undertake a review of quantitative evidence synthesis methods used for evaluating treatment sequences. The review is divided into three parts: a summary of the meta-analytic techniques used to develop summary effect estimates of treatment sequences, or effect estimates that are conditional on previous treatment used; the range of simplifying assumptions used as alternatives to conditional effect estimates to inform decision making within modelling studies; and the actual decision analytic approaches used for modelling treatment sequences. The findings of each are presented in the following Chapters 5, 6, and 7, respectively. The next chapter summarises the findings of a review of NICE guidance, which was informed in part by the review of methods.

CHAPTER 4: CLINICAL AND ECONOMIC DECISION-MAKING REGARDING TREATMENT SEQUENCING WITHIN THE NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE'S HEALTH TECHNOLOGY APPRAISAL AND CLINICAL GUIDELINE PROCESS

4.1 CHAPTER OVERVIEW

This thesis aims to develop a framework for utilising quantitative evidence synthesis methods to estimate the effectiveness of treatment sequences for decision-making. In the UK for the National Health Service (NHS), the National Institute for Health and Care Excellence (NICE) is a key decision-maker. NICE base their recommendations on a review of clinical and economic evidence. The National Institute for Health Research (NIHR) commissions Technology Assessment Review (TAR) teams to provide independent research input to inform NICE appraisals. NICE also produces methodological guidelines based on existing methods. The methodology research programme is supported by the Medical Research Council (MRC) and NIHR. The process of developing NICE guidance was used as an exemplar to ensure that the framework would be of direct relevance to practice, and the use of evidence synthesis to inform decision making.

The health technology appraisal and clinical guidelines processes used by NICE to inform policy decision-making and clinical practice in the NHS in England and Wales are described here. A review of NICE guidance documents was also conducted to identify clinical areas where treatment sequencing was likely to be a pertinent issue, and identify some of the challenges associated with the assessment of treatment sequences for decision-making.

4.2 SUMMARY OF THE NICE PROCESS AND HOW IT SERVES THE UK NHS

In the UK, NICE is the body charged with assessing the clinical effectiveness and cost effectiveness of health technologies, and making national recommendations about the introduction of new and established interventions in England and Wales. The Institute's objectives are to ensure that new investment yields maximum health benefit, reduces unwarranted variation in medical practice, and encourages rapid diffusion of high value new technologies.²⁸⁰

The Institute develops recommendations that guide decisions in health, public health and social care. Their guidance takes a number of forms, which include clinical guidelines, technology appraisals guidance, interventional procedures guidance, medical technologies guidance, diagnostics guidance, highly specialised technologies guidance, and medical technologies evaluation programme. The thesis will focus on the following two:

- i. Technology appraisals guidance, which assess the clinical and cost effectiveness of health technologies, such as new pharmaceutical and biopharmaceutical products, but also includes procedures, devices and diagnostic agents. This ensures that all NHS

patients have equitable access to the most clinically and cost-effective treatments available.

- ii. Guidelines, which make evidence-based recommendations on a wide range of topics, from preventing and managing specific conditions to planning broader services and interventions to improve the health of communities.

NICE technology appraisals comprise a clinical evaluation and a cost-effectiveness evaluation. At the time of this review, NICE used two appraisal processes: the multiple technology appraisal (MTA) process that incorporates the comparison of all relevant treatments in the decision framework, and the single technology appraisal (STA) process that is based on an economic evaluation of a single technology conducted by the manufacturer. In the STA process, the evidence submitted by the manufacturer is reviewed by an independent Technology Assessment Review team commissioned by NIHR. This review incorporates the critique, amendment and extension of the manufacturer model. Due to the pressure of providing timely guidance to the NHS at the time of launch, NICE introduced a single technology appraisal process in 2005 in order to “fast track” appraisals of single technologies for single indications.²⁸¹ The drive to develop timely technology appraisals is not unique to the UK or NICE; other health technology assessment agencies and regulators face the same challenge. In the United States there has been an expansion in the use of the comparative effectiveness research process (summarised in Chapter 1, Section 1.6), which is similar to the multiple technology appraisal process. The expansion of, and government research funding for, comparative effectiveness research has also led to federal agencies such as the Agency for Healthcare Research and Quality (AHRQ) and the National Institute of Health (NIH) to re-organise themselves to conduct and disseminate comparative effectiveness research. However, the time available to develop the health technology assessment is still limited.

The NICE clinical guideline development process includes an initial scoping process, used to decide which topics will be covered, which may not necessarily include the whole pathway. Similar to the technology appraisal process, questions of cost-effectiveness are addressed using an initial review of published economic evaluations. If no study is identified that is rated ‘directly applicable’ and with ‘minor limitations’, then a *de novo* economic evaluation is conducted. When only minimal evidence is available, recommendations are based on the Guideline Development Group’s experience and opinion of what constitutes good practice.²⁸²

4.3 NICE METHODOLOGICAL GUIDANCE ON EVALUATING TREATMENT SEQUENCING

NICE regularly reviews its processes and methodology, and in 2013 published an update of its Guide to the Methods of Technology Appraisal.²⁸³ The Decision Support Unit (DSU), a collaboration between a number of universities, is commissioned by NICE to provide a research and training resource to support the Institute's Technology Appraisal Programme (<http://nicedsu.org.uk/>). The DSU has developed a number of briefing papers and reports, which contributed to the 2013 update of the NICE

methods guide. This included a briefing document on reviewing sequential treatments and downstream costs.²⁸⁴ The updated methods guide highlighted the fact that some technology appraisals may need to consider the comparison of treatment sequences. However, neither the updated methods guide, nor the DSU's briefing document provided guidance on evaluating the clinical effectiveness or modelling treatment sequences. This is despite the briefing document highlighting the following key questions:

- i. Under what circumstances is it acceptable to model only individual lines of therapy, rather than treatment sequences?
- ii. How can the methods guide ensure that modelling of treatment sequences is undertaken consistently across appraisals?
- iii. Should explicit guidance on aspects of modelling treatment sequences be given?
 - a. When and how should sequences be identified? Which should be modelled?
 - b. What effectiveness estimates and model parameters can be reasonably used when a treatment is included in different places in different sequences?
 - c. What level of primary and sensitivity analyses should be reasonably expected?

4.4 A REVIEW OF NICE GUIDANCE DOCUMENTS THAT RECORDED TREATMENT SEQUENCING AS A PERTINENT ISSUE

4.4.1 Objectives

A review of NICE guidance documents was undertaken to identify:

- i. Challenges or problems associated with the assessment of treatment sequences for decision-making
- ii. Clinical areas where treatment sequencing was likely to be a pertinent issue

4.4.2 Searches

The review was primarily based on a specific search of the NICE website, however it was also informed by the searches undertaken for the broader review of methods presented in Chapter 3.

The NICE website was searched using the terms 'sequential(ly)', 'sequences' and 'sequencing' in November 2013. The results were scanned to identify NICE guidance or guidelines where treatment sequencing was likely to be a pertinent issue. Diagnostic guidance and diagnostic testing sequences were not considered. NICE guidance or guidelines that related to treatment administration, for example concurrent versus sequential combination therapy, were also excluded. The reason for this is explained in more detail in Chapter 3, Section 3.4.3.

The searches undertaken for the review of methods included the Health Technology Assessment (HTA) Database, which is part of the Cochrane Library. The technology assessment reports for all multiple technology appraisals conducted on behalf of NICE are published as a HTA journal publication, and subsequently listed on the HTA Database. The reference details for NICE technology

appraisals included in the review of methods are, where possible, based on the HTA journal publication.

4.4.3 Overall findings

NICE guidance documents highlighted the relevance and some of the challenges associated with the assessment of treatment sequences for decision-making. Some NICE guidance was identified, using the website search, because of the lack of possibility of the consideration of treatment sequencing. This included reference to a lack of data, inadequate evaluation within the decision models submitted by industry, or the omission of treatment sequencing from the original scope. The main factors relating to treatment sequencing which allowed the NICE guidance documents to be identified by the website's search engine are listed in Table 4.1. Illustrative examples of technology appraisals or clinical guidelines are provided. These are described in more detail in Appendix Volume II (Appendix B).

The technology appraisals and clinical guidelines where treatment sequencing was identified as a potential issue within the decision-making framework or review question are listed in Table 4.2. These are grouped according to the clinical condition, and presented under the headers of inflammatory arthritis, cancer, mental health disorders, hepatitis, and other. The NICE website search did not identify all the NICE technology appraisals included in the review of methods (Chapters 5-7); those that were not identified are listed separately in Table 4.3. This highlights the challenge of identifying relevant methodological studies evaluating treatment sequences when any clinical condition is considered.

Table 4.1: Reasons why NICE guidance were identified by the NICE website search

Reasons why NICE guidance was identified by NICE website search	Related technology appraisals and clinical guidelines*
Treatment sequences were highlighted as an important consideration in the introduction of the guidance document.	<ul style="list-style-type: none"> • Treatment of chronic hepatitis B (CG165).
The committee discussed the cost-effectiveness analysis of the various treatment sequences.	<ul style="list-style-type: none"> • Drug treatment of chronic hepatitis B (TA165). • Chemotherapy for advanced colorectal cancer (TA93).
The final recommendations considered treatment sequencing.	<ul style="list-style-type: none"> • Chemotherapy for advanced breast cancer (CG81) or colorectal cancer (CG131). • Antipsychotic drugs for schizophrenia (CG82).
The committee could not make any recommendations on treatment sequencing due to lack of data.	<ul style="list-style-type: none"> • Sequential TNF-inhibitors for psoriatic arthritis (TA199) or psoriasis (TA180). • Sequential endocrine/hormone therapy for advanced or metastatic breast cancer (TA239 and CG81).
The manufacturer's model did not consider treatment sequencing due to lack of data.	<ul style="list-style-type: none"> • Febuxostat for hyperuricaemia associated with gout (TA164). • Dronedarone for atrial fibrillation (TA197).
The committee could not make any recommendations on sequential treatment strategies as it had not been specified in the scope for the appraisal, and therefore not consider in manufacturer's model.	<ul style="list-style-type: none"> • Telaprevir for hepatitis C (TA252).
The evidence review group criticised the manufacturer's model for not incorporating treatment sequencing, or conducted further exploratory analysis incorporating treatment sequences.	<ul style="list-style-type: none"> • Retigabine for epilepsy (TA232). • Apixaban to prevent stroke and systemic embolism in patients with atrial fibrillation (TA275). • Eltrombopag for thrombocytopenic purpura (TA293).
The manufacturer's model included treatment sequencing.	<ul style="list-style-type: none"> • Golimumab (TA225) or tocilizumab (TA247) for rheumatoid arthritis.

Reasons why NICE guidance was identified by NICE website search	Related technology appraisals and clinical guidelines*
	<ul style="list-style-type: none"> • Drug treatment for attention deficit hyperactivity disorder (TA98), aripiprazole for bipolar disorder (TA292). • Telbivudine (TA154) or tenofovir disoproxil fumarate (TA173) for hepatitis B.
The committee considered the treatment sequences included in the manufacturer's model were unsuitable, or that the model did not consider the full range of relevant sequences.	<ul style="list-style-type: none"> • Entecavir for Hepatitis B (TA153).
The manufacturer's submitted a revised economic model with changes made to treatment sequences.	<ul style="list-style-type: none"> • Aripiprazole for schizophrenia (TA213).
The recommendations for research included treatment sequencing.	<ul style="list-style-type: none"> • Sequential TNF-inhibitors for ankylosing spondylitis (TA143). • Psychological interventions for common mental disorders (CG123), interventions for low back pain (CG88). • Treatment modalities for peritoneal dialysis (CG125).
There was an acknowledgement that ongoing trials existed investigating the optimum sequence of treatment.	<ul style="list-style-type: none"> • Treatment of advanced renal cell carcinoma (TA169 and TA178).
The committee deliberated that overall survival may not be a suitable endpoint for the evaluation of first-line treatment for advanced or metastatic cancer due to the use of sequential treatments used in clinical practice.	<ul style="list-style-type: none"> • TA219 (renal cell carcinoma), TA119 (chronic lymphocytic leukaemia), TA242 (colorectal cancer).
The committee noted the need for the full sequence of treatments to be recorded for all patients in all clinical trials as part of the recommendations for future research.	<ul style="list-style-type: none"> • TA93 chemotherapy treatments for advanced colorectal cancer.

* A full list and description of these are provided in the Appendix Volume II (Appendix B)

Abbreviations: CG clinical guideline; TA technology assessment; TNF Tumour necrosis factor alpha

Table 4.2: Technology appraisals and clinical guidelines identified during a search of the NICE website (ordered by condition and earliest year of publication)

The shading in last column indicates studies that were included in the methodology review of meta-analysis and modelling approaches

Clinical condition and treatment considered in the NICE guidance (technology appraisal or clinical guideline reference number)*	Publication date of NICE guidance	Author, year of associated HTA publication**
INFLAMMATORY ARTHRITIS		
Ankylosing spondylitis - adalimumab, etanercept and infliximab (TA143)	May 2008	McLeod, 2007
Ankylosing spondylitis - golimumab (TA233)	August 2011	
Psoriatic arthritis - etanercept, infliximab and adalimumab (TA199)	August 2010	Rodgers, 2011 (modelling)
Rheumatoid arthritis - adalimumab, etanercept, infliximab, rituximab and abatacept (after the failure of a TNF inhibitor) (TA195) (Updates TA126 and TA141 and partially TA36)	August 2010	Malottki, 2011 (modelling)
Rheumatoid arthritis (after the failure of previous anti-rheumatic drugs) - golimumab (TA225)	June 2011	
Rheumatoid arthritis - tocilizumab (rapid review TA198) (TA247)	February 2012	
Rheumatoid arthritis - abatacept (2nd line) (rapid review of TA234) (TA280) [WITHDRAWN; replaced by TA375]	April 2013	
Rheumatoid arthritis (not previously treated with DMARDs or after conventional DMARDs only have failed) - Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept (TA375; previously ID537, 2013. (Updates of TA130, TA186, TA224, TA234, TA225, and TA247)	January 2016***	Stevenson, 2016***
CANCER		
Breast cancer (advanced) (CG81)	February 2009	(modelling)
Breast cancer - bevacizumab (in combination with a taxane) (TA214)	February 2011	
Breast cancer (metastatic) - fulvestrant (TA239)	December 2011	
Colorectal cancer (advanced) - irinotecan, oxaliplatin and raltitrexed (TA93) (replaced by CG131)	August 2005	Hind, 2008
Colorectal cancer (metastatic) - bevacizumab and cetuximab (TA118) (partially updated by TA242)	January 2007	Tappendern, 2007
Colorectal cancer (CG131)	November 2011	(modelling and meta-analysis)
Leukaemia (chronic lymphocytic) - fludarabine (TA119)	February 2007	
Leukaemia (chronic myeloid, first-line) - dasatinib, nilotinib and standard-dose imatinib (TA251) [WITHDRAWN]	April 2012	Pavey, 2012
Multiple myeloma - lenalidomide (TA171)	June 2009	
Ovarian cancer (advanced) - paclitaxel, pegylated liposomal doxorubicin hydrochloride and topotecan (TA91) [WITHDRAWN]	May 2005	Main, 2006
Renal cell carcinoma - sunitinib (TA169)	March 2009	
Renal cell carcinoma (advanced and/or metastatic) - bevacizumab (first-line), sorafenib (first- and second-line), sunitinib (second-line) and temsirolimus (first-line) (TA178)	August 2009	Thompson Coon, 2010

Clinical condition and treatment considered in the NCE guidance (technology appraisal or clinical guideline reference number)*	Publication date of NICE guidance	Author, year of associated HTA publication**
Renal cell carcinoma (advanced, second-line treatment - everolimus (TA219)	April 2011	
MENTAL HEALTH DISORDERS		
Antisocial personality disorder (CG77)	January 2009	
Attention deficit hyperactivity disorder (ADHD) - methylphenidate, atomoxetine and dexamfetamine (review of TA13) (TA98)	March 2006	
Bipolar disorder (adolescents) - aripiprazole (TA292)	July 2013	
Common mental health disorders (CG123)	May 2011	
Depression in children and young people (CG28)	September 2005	
Depression in adults (update) (CG90)	October 2009	
Obsessive compulsive disorder (OCD) and body dysmorphic disorder (BDD) (CG31)	November 2005	
Psychosis with coexisting substance misuse (CG120)	March 2011	
Schizophrenia (update) (CG82)	March 2009	
Schizophrenia - aripiprazole (TA213)	January 2011	
HEPATITIS		
Hepatitis B (chronic) - adefovir dipivoxil and pegylated interferon alfa-2a (TA96) [partially updated by CG165]	February 2006	Shepherd, 2006 (modelling)
Hepatitis B - entecavir (TA153)	August 2008	
Hepatitis B - telbivudine (TA154)	August 2008	
Hepatitis B - tenofovir disoproxil fumarate (TA173)	July 2009	
Hepatitis C (genotype 1) - telaprevir (TA252)	April 2012	
Hepatitis B (chronic) (CG165)	June 2013	
OTHER		
Atrial fibrillation - dronedarone (TA197)	August 2010	
Chronic kidney disease [CKD] (stage 4 or 5): management of hyperphosphataemia (CG157)	March 2013	
Cystic fibrosis (pseudomonas lung infection) - colistimethate sodium and tobramycin (TA276)	March 2013	Tappendern, 2013
Epilepsy (partial onset seizures) - Retigabine as adjunctive treatment (TA232)	July 2011	
Epilepsies: diagnosis and management (CG137) [<i>id. In website search but no sequencing related terms or references identified in guidance document when retrieved in full</i>]	January 2012	(modelling)
Gout (Hyperuricaemia) - febuxostat (TA164)	December 2008	
Gout (tophaceous, severe debilitating, chronic) - pegloticase (TA291)	June 2013	
Low back pain - Early management of persistent non-specific low back pain (CG88) [WITHDRAWN – updated and replaced by NG59]	May 2009	
Neuropathic pain in adults: pharmacological management in non-specialist settings (CG96) [WITHDRAWN – updated and replaced by CG173]	March 2010	
Neuropathic pain - pharmacological management (CG173)	November 2013	
Osteoporotic fractures - denosumab (TA204)	October 2010	
Peritoneal dialysis (CG125)	July 2011	
Psoriasis - ustekinumab (TA180)	September 2009	
Psoriasis (CG153)	October 2012	Sawyer, 2013 (modelling)
Stroke and systemic embolism (prevention, non-valvular atrial fibrillation) - apixaban (TA275)	February 2013	
Thrombocytopenic purpura - eltrombopag (TA205) [WITHDRAWN replaced by TA293]	October 2010	
Thrombocytopenic purpura - eltrombopag (TA293)	July 2013	
Thrombocytopenic purpura - romiplostim (TA221)	April 2011	
Type 2 diabetes: newer agents (Short CG87) [REPLACED by NG28]	May 2009	
Type 2 diabetes - Dapagliflozin combination therapy (TA288)	June 2013	
Type 2 diabetes in adults: management (NG28) [REPLACES CG87]	December 2015	
Urinary incontinence in women: management (CG171)	September 2013	

*Multiple technology appraisals are reported here using the same format as single technology appraisals on the NICE website, using the condition first then treatment.

**The information in the parenthesis indicates whether the study was included in the review of meta-analytic methods (Chapter 5) or modelling approaches (Chapters 6-7).

Id identified

*** This review of modelling studies was available as a proposal when the methodology review for the thesis was conducted in 2013

Abbreviations: CG clinical guideline; DMARD disease modifying anti-rheumatic drugs; TA technology assessment

Table 4.3: Technology appraisals included in the methodology review that were not identified during the NICE website search (ordered by condition and earliest year of publication)

Shading used in last column to indicate studies that were included in the methodology review of meta-analysis and modelling approaches

Clinical condition and treatment (technology appraisal or clinical guideline reference number)*	Publication date	Author, year of HTA reference used in methods review*
RHEUMATOLOGY		
Rheumatoid arthritis and juvenile poly-articular idiopathic arthritis - etanercept and infliximab (after failure of previous conventional DMARDs) (TA36) [WITHDRAWN Replaced by TA130 and TA195]	March 2002	Jobanputra, 2002 (modelling)
As above (TA36)		Barton, 2004 (modelling)
Rheumatoid arthritis – anakinra (TA72) [WITHDRAWN Replaced by CG79]	November 2003	Clark, 2004 (modelling)
Rheumatoid arthritis - adalimumab, etanercept and infliximab (TA130) [Updates TA36. WITHDRAWN updated and replaced by TA375]	October 2007	Chen, 2006 (modelling)
OTHER		
Epilepsy in adults - newer drugs (TA76) [WITHDRAWN replaced by CG137]	March 2004	Wilby, 2005 (modelling)
Epilepsy in children - newer drugs (TA79) [WITHDRAWN replaced by CG137]	April 2004	Connock, 2006 (modelling)
Psoriasis (severe) in adults - etanercept and efalizumab (TA103) [WITHDRAWN replaced by CG175]	July 2006	Woolacott, 2006 (modelling)
Crohn's disease: management (CG152)	October 2012	(modelling)
Depressive illness: electroconvulsive therapy (TA59). <i>The NICE TA covered electroconvulsive therapy (ECT) for treating depressive illness, schizophrenia, catatonia and mania, but treatment sequencing was only considered for major depressive disorder (MDD)</i>	April 2003	Greenhalgh, 2005 (modelling)

*The information in the parenthesis indicates whether the study was included in the review of meta-analytic methods (Chapter 5) or modelling approaches (Chapters 6-7).

Abbreviations: CG clinical guideline; DMARD disease modifying anti-rheumatic drugs; TA technology assessment

4.4.4 Clinical conditions for which treatment sequencing was an important consideration for NICE decision-making

The findings from the NICE website search showed a breadth of clinical conditions and decision-making contexts where treatment sequencing was considered relevant within the NICE process. The clinical conditions most frequently covered by NICE technology appraisals and clinical guidelines identified during the website search (Table 4.2) or included in the methodology review (Table 4.3), were rheumatoid arthritis (8 TAs/CGs), hepatitis B (8 TAs/CGs), epilepsy (4 TAs/CGs), renal cell cancer (3 TAs/CGs), breast cancer (3 TAs/CGs), colorectal cancer (3 TAs/CGs), type 2 diabetes mellitus (3 TAs/CGs), and psoriasis (3 TAs/CGs). Several technology appraisals and clinical guidelines identified also covered mental health disorders, such as schizophrenia, bipolar disorder, and depression.

The importance of treatment sequencing within the decision framework may differ according to both overall treatment goal and clinical context. Treatment sequencing was identified as an important issue for treatments which aimed to:

- i. Retard the disease process, with potential benefit extending beyond the treatment period, such as biological agents for inflammatory arthritis
- ii. Prolong life, such as systemic therapies for advanced cancer. This was identified as particularly relevant in that the overall survival benefit from first line treatment could be confounded by subsequent treatments (Table 4.1)

- iii. Alleviate symptoms, or where the benefits may be restricted to the treatment period only, for example, treatments for epilepsy, neuropathic pain, or low back pain

These three scenarios, and the corresponding clinical context, are explored in more detail in the Appendix Volume I (Section C). Three clinical indications that were identified as particularly relevant to NICE decision-making include:

- i. The introduction and sequential use of biological agents for retarding disease progression in inflammatory arthritis, which is illustrated in more detail using a rheumatoid arthritis case study
- ii. The introduction of novel biological (targeted) therapies for prolonging life in advanced or metastatic cancer, which is illustrated in more detail here using renal cell carcinoma as a case study
- iii. The optimal sequencing of new antiepileptic drugs

4.4.5 The evaluation of treatments sequences within clinical guidelines

NICE decisions or recommendations on treatment sequencing were often based on the deliberations of the guideline development groups, for example the clinical guideline for neuropathic pain (CG96), type 2 diabetes mellitus (CG88), and psoriasis (CG153). They were often informed by a narrative synthesis of the evidence of clinical effectiveness, or ordered according to the cost effectiveness of individual treatments. However, the clinical guidelines for breast cancer (CG81), colorectal cancer (CG131), and epilepsy (CG137) included treatment sequencing within the *de novo* model conducted as part of the economic evaluation, and all three are therefore included in the broader review of methods presented in Chapters 5-7. Due to the limitations of the evidence base, treatment sequencing effect estimates were based on the use of simplifying assumptions (CG131) and the expert clinical opinion of the guideline development group (CG81). The simplifying assumptions used are reviewed in more depth in Chapter 6. In the epilepsy clinical guideline (CG137), a reduction factor was applied to the treatment effect estimates, which was informed by an anchoring study. This is reviewed in more detail in Chapter 5, Section 5.6.2. A summary of some of the issues pertaining to the limitations in the evidence base for treatment sequences in advanced or metastatic cancer and epilepsy is provided in the Appendix Volume I (Sections C4 and C5).

4.4.6 The evaluation of treatments sequences within the technology appraisals guidance

Treatment sequencing within the technology appraisals was generally evaluated as part of the economic evaluation and analysed using decision analytic modelling, but not considered as part of the clinical evaluation. However, one technology appraisal of chemotherapies for advanced colorectal cancer (TA93) evaluated the clinical evidence of both treatment sequences and individual lines of treatment. For the economic evaluation they limited inclusion to sequencing trials and identified only two relevant randomised controlled trials (RCTs). This is reviewed in more detail in Chapter 5, Section 5.3. The NICE committee's deliberations included making indirect treatment comparisons between treatment sequences used in one of these trials and those 'currently recommended by NICE'. A

second technology appraisal of biological agents for psoriatic arthritis (TA199; Rodgers, 2011) included an economic evaluation of sequential TNF-inhibitors as part of their sensitivity analyses. This was informed by a supplementary review of observational studies to inform the effectiveness of treatment sequences, which is reviewed in more detail in Chapter 5, Section 5.6.2. The remaining technology appraisals that investigated sequencing as part of the economic evaluation were mainly based on the application of simplifying assumptions, which are reviewed in depth in Chapter 6. These related to the use of biological agents for rheumatoid arthritis (TA36; TA72; TA195; TA130), antiviral drugs for chronic hepatitis B (TA96), biological agents for psoriasis (TA103), antiepileptic drugs (TA76, TA79), and the use of electroconvulsive therapy for treating a major depressive disorder (TA59).

4.5 THE OMISSION OF NICE SINGLE TECHNOLOGY APPRAISALS AS A POTENTIAL LIMITATION OF THE REVIEW OF MODELLING STUDIES

The literature search for the review of methods presented in Chapter 3 did not include an extensive search of NICE single technology assessments, and therefore did not include unpublished manufacture submission models. However, a review by Zhen *et al.* (2017) addressed this potential limitation.¹⁰⁴ The authors conducted a review of economic models capturing treatment sequences published by NICE. Technology appraisals published as of 24 October 2014 (search date) were identified via the NICE website. The review included economic models that were developed by the manufacturer or by the assessment group. The authors evaluated the key features of the decision problem that necessitate modelling treatment sequences and the methodology and data requirements.

The review identified 40 treatment sequencing models in the following disease areas: oncology (13), autoimmune (7), cardiovascular (6), neurology/mental health (4), infectious disease (2), diabetes (2), and other (6).¹⁰⁴ The most common rationale for modelling treatment sequences was to reflect either clinical practice or clinical trial design. In other cases it was used to assess where in a treatment sequence a new treatment should be placed, to evaluate the addition of more efficacious treatment options to a current treatment sequence, or because of disease-specific rationale. Examples of disease-specific rationale included the need to reflect the treatment algorithm required by disease progression and aging in diabetes, and the need to track patient-resistance to treatment and treatment history in infections.

The review conducted Zheng *et al.* did not identify any modelling approaches or simplifying assumptions not identified in the review of modelling studies presented in Chapters 6 and 7.¹⁰⁴ I revisit the comparison of the findings of the two reviews in Chapter 6 (Section 6.7.2) and Chapter 9 (Section 9.3). The disease areas and range of rationale for evaluating treatment sequences identified by Zheng *et al.* were also comparable to that of the modelling studies summarised in Chapters 6-7.¹⁰⁴

4.6 CHAPTER SUMMARY AND WHAT NEXT

The findings of the review of NICE guidance showed that treatment sequencing was often identified as a potential issue in the deliberation of NICE decisions, but not always considered as part of the evidence review. They also showed that when treatment sequences were evaluated as part of the NICE clinical guideline or technology appraisal process, they were usually only considered within the economic evaluation. No guidance was provided on methods for evaluating treatment sequences. Decisions on treatment sequencing were frequently made based on the deliberations of the NICE Committee and a very limited evidence base. A more objective assessment of the evidence and the uncertainty in both the sequencing specific effect estimates and decision-making are required. The review of the methods used to synthesise the data on clinical effectiveness of treatment sequences can inform this process, the findings of which are presented in the next Chapter.

The review of NICE guidance identified some of the challenges faced by NICE when considering treatment sequences. It highlighted the need to consider whether treatment sequencing is an importance issue from the outset, i.e. during the scoping stage for NICE guidance. The review identified examples the Committee's inability to make decisions due to the economic models not considering the relevant sequences.

Another challenge is that the number and order of sequences used in clinical practice continually changes with the introduction of new treatments, for example in advanced cancer. The lack of data on treatment sequences was often a stumbling block, and the evidence for new treatments was frequently limited to the licencing trial. The review also highlighted that overall survival is not a suitable endpoint for evaluating first line treatments for advanced or metastatic cancer.

The review was based on the NICE guidance and not an in-depth evaluation of the evidence review group reports. This was in line with the purpose of identifying clinical scenarios where treatment sequencing was a potential issue, as well as issues that may need to be considered in the framework presented in Chapter 8. The review was also used to identify technology appraisals and clinical guidelines for inclusion in the review of methods, the findings of which are presented over the next three chapters.

CHAPTER 5: META-ANALYTIC METHODS USED FOR EVALUATING TREATMENT SEQUENCES

5.1 CHAPTER OVERVIEW

This chapter presents the findings of the review of meta-analytic techniques used to evaluate treatment sequences. The term 'meta-analytic technique' is used here in its broadest sense, and incorporates conventional pair-wise meta-analysis, meta-regression, and network meta-analysis. Each method employed for evaluating the clinical effectiveness of treatment sequences, or single treatments conditional on their positioning in the pathway, is summarised along with an assessment of its strengths and applicability for future practice. Due to the dearth of methods identified, a broad remit was iteratively applied for identifying relevant approaches, including simplistic approaches such as stratifying the analysis according to treatment history and restricting inclusion to comparative studies of treatment sequences. The implications and limitations of each approach is discussed, and the challenges of using observational studies to inform treatment-sequencing effects is summarised. The discussion section explores methodological development in this field, which can potentially contribute to the framework for the conduct of meta-analytic methodology to evaluate treatment sequences presented in Chapter 8. The chapter ends with a summary of the implications for future practice and methodological development.

5.2 OVERALL FINDINGS

The literature searches and the methods used to identify and review the relevant meta-analytic studies are described in Chapter 3.

The studies discussed in this section are listed in Tables 5.1 and 5.2, which summarises their overall aim and the methodological approach to which they contribute, respectively. Of the 22 included meta-analytic studies, 11 specifically aimed to evaluate the clinical effectiveness of treatment sequences.¹² 17 20 199-201 203 207 209 214 215 The remaining studies are included as they contribute to developing a broader list of approaches used for evaluating the effect of previous treatment or positioning in the treatment pathway.^{21 198 202 204-206 208 210-213} A summary of their aim and why they were selected for inclusion is provided in Table 5.1. These studies represent examples, rather than provide an exhaustive list of studies that used each particular approach.

The breadth of meta-analytic approaches considered relevant to treatment sequencing was expanded due to the scarcity of methods identified to include the following approaches that could potentially answer the following questions, which are relevant for informing decision making:

- i. Do treatment-sequencing effects need to be considered?
- ii. Does the number of prior treatments influence treatment effects?
- iii. Does the class of prior treatments influence treatment effects?
- iv. Does the reason for discontinuing prior treatments influence treatment effects?

- v. Is it feasible to limiting inclusion to sequencing studies?
- vi. Is it feasible to develop a conditional effect estimate, considering the available evidence?

The 22 included studies identified five main approaches used to assess treatment-sequencing effects:

- i. Restricting inclusion to comparative studies evaluating treatment sequences.
- ii. Subgroup analyses to assess whether treatment effects varied according to the treatment patients had previously received. This includes stratifying meta-analyses according to line of treatment, an approach that does not provide a means of developing summary effects that allows for treatment sequencing, but is frequently used to inform clinical practice where treatments are being used as part of a sequence.
- iii. Meta-regression to estimate the contributing effect of, or adjust for, the number and type of previous treatments.
- iv. Applying a modifying factor to individual treatment effects in order to represent their use at a different point in the treatment pathway.
- v. Ranking individual treatments according to their absolute effect estimates in order to inform the effects of whole sequences. This is an example of a naïve method, which assumes no sequencing effect.

These approaches, were all implemented within the framework of meta-analytic techniques, including conventional pairwise meta-analyses, and network meta-analyses, covering indirect treatment comparisons or multiple treatment comparisons. The use of network meta-analysis, as noted in Chapter 2, is becoming increasingly popular. None of the included studies incorporated further adaptations of network meta-analysis methods in order to specifically accommodate or evaluate treatment sequencing. No novel evidence synthesis or meta-analytic techniques were identified. None were directly aimed at developing a conditional summary estimate of effect. No meta-analysis of sequential multiple assignment randomised trials (SMART) or N-of-1 randomised controlled trials (RCTs) for investigating treatment sequences were identified.

Most included studies investigated treatment sequencing for either inflammatory arthritis (mainly biological agents for rheumatoid arthritis) or advanced cancers. Each of the approaches identified by the review is expanded upon in the following sections.

Table 5.1: Primary aim of the included studies evaluating meta-analytic approaches for evaluating treatment sequences

(Studies ordered according to whether they aimed to evaluate treatment sequencing, and then alphabetical)

Author, year	Study aim included treatment sequencing	Aim of the study	Notes or reason why study included	Approach to which they contribute
NICE CG131 (2011)	YES (economic evaluation)	To assess the effectiveness and cost-effectiveness of chemotherapy sequences for advanced colorectal cancer.	The economic evaluation included analysis of chemotherapy sequences, which required treatment-sequencing effect estimates. It was also acknowledged that overall survival of first-line treatments are confounded by subsequent treatments.	<i>Stratified analysis; Sequencing studies</i>
Connock, 2006 (NICE TA79)	YES (economic evaluation)	To examine the clinical effectiveness and cost-effectiveness of newer antiepileptic drugs (AEDs) for epilepsy in children. For the newly, or recently, diagnosed population, the key question for the newer drugs is how soon they should be tried.	Economic model included treatment sequencing, which required treatment-sequencing effect estimates.	Modifying factor
Cooper, 2011	YES	To systematically review studies of the management of treatment-refractory depression in older people, covering pharmacological, physical, and psychological interventions.	The searches specifically included treatment sequencing; 'sequential treatment or trial' used as a search term.	<i>Stratified analysis; Sequencing studies</i>
Heng, 2014	YES	To systematically summarise and interpret the published real-world evidence comparing sequential treatment for metastatic renal cell carcinoma.	Set out to evaluate clinical effectiveness of treatment sequences.	<i>Sequencing studies</i>
Lloyd, 2011	YES	To evaluate the effectiveness of TNF-inhibitors, for the treatment of RA, when used sequentially	Set out to evaluate clinical effectiveness of treatment sequences.	<i>Meta-regression; Subgroup analysis</i>
Rendas-Baum, 2011	YES	To evaluate the relationship between the clinical response to biologics and the number of previous treatments with TNF-inhibitors for the treatment of RA.	Included patients who had failed at least one TNF-inhibitor.	<i>Subgroup analysis</i>
Rodgers, 2011 (NICE TA199)	YES (economic evaluation)	To determine the clinical effectiveness, safety and cost-effectiveness of TNF-inhibitors in the treatment of active and progressive PsA.	Economic evaluation included investigating the cost effectiveness of using sequential TNFs as part of a sensitivity analysis. The model required treatment-sequencing effect estimates.	<i>Modifying factor</i>
Ruhe, 2006	YES	To systematically review the evidence for switching pharmacotherapy after a first SSRI for major depressive disorder.	Set out to evaluate clinical effectiveness of treatment sequences.	<i>Stratified analysis; Sequencing studies</i>
Stenner, 2012	YES	To evaluate the optimal sequence for the tyrosine kinase inhibitors sorafenib and sunitinib in metastatic renal cell carcinoma.	Highlights the potential contrasting methods and findings of reviews that limit inclusion to sequencing studies.	<i>Sequencing studies</i>
Suarez-Almazor, 2007 (CADTH)	YES	To review the evidence on the TNF-inhibitors, INF and ETA, regarding the timing of therapeutic introduction, dose escalation, and switching.	One of the research questions included: Do patients with RA who fail treatment with one TNF agent respond to therapy with a different one? Timing of therapeutic introduction included an evaluation of potential differences in clinical effectiveness according to timing of therapy (as initial therapy or after failure with other drug therapies) and disease duration.	<i>Subgroup analysis; Stratified analysis</i>
Finnerup, 2005	YES	To develop up-to-date calculation of NNT and NNH in neuropathic pain as the basis of a proposal for an evidence-based treatment algorithm.	Conducted an updated review of placebo-controlled RCTs to support an evidence-based algorithm (sequence of treatments) to treat neuropathic pain conditions. Evaluation based on naive assumptions regarding treatment-sequencing effects.	<i>Ranking absolute effects</i>
Grothey, 2004	PARTIAL	To evaluate the importance of the availability of all 3 active cytotoxic agents, FU-LV, irinotecan, and oxaliplatin, on overall survival (OS) in	Data on the percentage of patients receiving 2 nd -line therapy and the percentage of patients receiving all 3 agents were correlated with the reported median OS, using a weighted	Meta-regression

Author, year	Study aim included treatment sequencing	Aim of the study	Notes or reason why study included	Approach to which they contribute
		advanced CRC. FU-LV plus irinotecan or oxaliplatin were the standard 1 st -line therapies.	analysis. The analysis did not address the question of the best treatment sequence, i.e. whether the irinotecan or oxaliplatin-based protocol should be used first.	
<i>Hind, 2008 (NICE TA93)</i>	PARTIAL	To evaluate the cost-effectiveness of irinotecan, oxaliplatin and raltitrexed as first-line treatments in the management of advanced colorectal cancer.	Each line of therapy was evaluated separately in the clinical evaluations, with the caveat that in trials of 1 st -line treatment over half of participants, in all but 2 included trials, received unplanned 2 nd -line treatment. Existing economic models were considered weak due to the use of unplanned 2 nd -line therapies, and were either limited to using progression-free survival (PFS) as a surrogate outcome, or subject to confounding. An improved model was therefore implemented by the authors using data from the 2 trials that planned treatment sequences. The model was used to compare the 7 chemotherapy sequences included in these 2 trials.	<i>Sequencing studies Stratified analysis</i>
<i>Anderson, 2000</i>	NO	To identify factors predicting response to second-line treatment, with conventional DMARDs or devices, in rheumatoid arthritis based on use of individual patient data.	Evaluating treatments used at a single point in the treatment pathway, but the factors they considered included both previous DMARD use and disease duration.	<i>Meta-regression</i>
<i>Christensen, 2015</i>	NO	To determine if variations in trial eligibility criteria and patient baseline characteristics could be considered effect modifiers of the treatment response when testing targeted therapies (biological agents and targeted synthetic DMARDs for RA).	Evaluating treatments used at a single point in the treatment pathway, but the factors they considered included both previous DMARD use and disease duration.	<i>Meta-regression and subgroup analysis</i>
<i>Kanters, 2014</i>	NO	To explore which clinical factors and patient characteristics are associated with the magnitude of comparative efficacy between biologics vs MTX in RA patients with inadequate response to MTX.	Evaluating treatments used at a single point in the treatment pathway, but demonstrates the challenges of including previous treatments as a covariate in meta-regression due the poor reporting of primary studies.	<i>Meta-regression</i>
<i>Mandema, 2011</i>	NO	To compare the dose-response relationship for the efficiency of biologics for the treatment of RA. Two of the objectives included: Are TNF-inhibitors different in patients with an inadequate response to MTX compared to those who are MTX-naïve; Are TNF-inhibitors more efficacious than MTX in MTX-naïve patients.	Evaluating treatments used at a single point in the treatment pathway, but provides an example of evaluating effect of previous treatment in a dose-response meta-regression analyses.	<i>Meta-regression; Stratified analysis</i>
<i>Nixon, 2007</i>	NO	To compare the efficacy of four biological agents, three of which were TNF-inhibitors for the treatment of RA.	Provides an example of using meta-regression to account for disease duration in network-meta-analysis. Also represents the first study to combine the techniques of mixed treatment comparisons and meta-regression to adjust or study level covariables.	<i>Meta-regression; Subgroup analysis</i>
<i>Salliot, 2011</i>	NO	To compare efficacy of biologics for the treatment of RA in 2 clinical situations: i) active disease despite MTX; ii) after inadequate response to TNF-inhibitor.	Evaluating treatments used at a single point in the treatment pathway, but provides an example of evaluating effect of previous treatment in subgroup analysis.	<i>Subgroup analysis; Stratified analysis</i>
<i>Schmitz, 2012</i>	NO	To compare efficacy of TNF-inhibitors for the treatment of RA in patients with inadequate response to MTX.	Evaluating treatments used at a single point in the treatment pathway, but provides an example of the challenges of including previous treatments as a covariate in meta-regression due the poor reporting of primary studies. (Builds on methods used by Nixon, 2007.)	<i>Meta-regression</i>

Author, year	Study aim included treatment sequencing	Aim of the study	Notes or reason why study included	Approach to which they contribute
Schoels, 2012	NO	To compare efficacy and safety of biologics, for the treatment of RA, after inadequate response to TNF-inhibitors.	Provided an example of the limitations of examining differences in treatment effects according to the number of previous biologics used.	Subgroup analysis
Singh, 2009	NO	To compare efficacy and safety of biologics for the treatment of RA.	Included planned subgroup analyses for: TNF failure vs none; and biologic failure vs conventional DMARD failure vs none.	Subgroup analysis

Abbreviations: CADTH Canadian Agency for Drugs and Technologies in Health; DMARD disease-modifying antirheumatic drug; CG clinical guidelines; CRC colorectal cancer; ETA etanercept; EULAR European League Against Rheumatism; FU-LV fluorouracil-leucovorin; HTA health technology assessment; INF infliximab; MTX methotrexate; NNH numbers needed to harm; NNT numbers needed to treat; PsA psoriatic arthritis; RA rheumatoid arthritis; TNF-inhibitors tumour necrosis factor-alpha inhibitors; SSRI Selective Serotonin Reuptake Inhibitor; vs versus

Table 5.2: Overview of the meta-analytic approaches used by included studies

(Studies are ordered according to the methodological approach used)

Study	Condition	Decision point	Study aim included sequential treatments	Sequencing studies	stratified analysis (single point in pathway)	Subgroup analyses	Meta-regression	Modifying factor	Ranking absolute effects	Available evidence base*
Heng, 2014	Metastatic renal cell carcinoma	2 nd line targeted therapies	YES	X						Observational studies (sequencing studies)
Stenner, 2012	Metastatic renal cell carcinoma	Sequential targeted therapy (2 lines)	YES	X						Observational studies (sequencing studies)
Hind, 2008	Advanced colorectal cancer	1 st and 2 nd -line chemotherapies	PARTIAL	X	X					RCTs
NICE CG131	Advanced colorectal cancer	1 st and 2 nd -line chemotherapies	YES	X	X					1 st -line: RCTs; 2 nd -line prospective sequencing trials (n=3, 2 were included as quasi-sequencing)***
Ruhe, 2006	Depression	2 nd -line treatments (after nonresponse to SSRI)	YES	X	X					RCTs (one sequencing) and observational studies; only RCTs pooled due to heterogeneity
Cooper, 2011	Depression	2 nd and subsequent-line treatments	YES	X	X					RCTs and uncontrolled open-label trials (one non-RCT sequencing study); pooled breaking randomisation.
Lloyd, 2011	Rheumatoid arthritis	Sequential TNF-inhibitors (1 st -line vs 2 nd /3 rd -line TNF-inhibitors)	YES			X	X			Observational studies (uncontrolled and sequencing studies)
Rendas-Baum, 2011	Rheumatoid arthritis	1 st , 2 nd or subsequent-line biologics	YES			X				RCTs and observational studies; pooled breaking randomisation

Study	Condition	Decision point	Study aim included sequential treatments	Sequencing studies	stratified analysis (single point in pathway)	Subgroup analyses	Meta-regression	Modifying factor	Ranking absolute effects	Available evidence base*
Suarez-Almazor, 2007	Rheumatoid arthritis	1 st and 2 nd -line TNF-inhibitors	YES		X	X				1 st -line: RCTs; 2 nd -line: observational studies only (not pooled)
Schoels, 2012	Rheumatoid arthritis	2 nd -line 'new' biologic (after multiple TNF-inhibitors)	NO			X				RCTs
Singh, 2009	Rheumatoid arthritis	1 st and 2 nd line biologics	NO			X				RCTs
Salliot, 2011	Rheumatoid arthritis	1 st and 2 nd biologics	NO		X	X				RCTs
Nixon, 2007	Rheumatoid arthritis	1 st -line biological agents (in early and late stage disease)	NO				X			RCTs
Schmitz, 2012	Rheumatoid arthritis	1 st -line TNF-inhibitors	NO				X			RCTs
Christensen, 2015	Rheumatoid arthritis	1 st and subsequent-line biologics	NO				X			RCT
Kanters, 2014	Rheumatoid arthritis	1 st -line biologics	NO				X			RCTs
Anderson, 2000	Rheumatoid arthritis	2 nd -line conventional DMARDs	NO				X			RCTs (individual patient-level data)
Mandema, 2011	Rheumatoid arthritis	1 st -line TNF-inhibitors	NO		X		X			RCTs
Grothey, 2004	Advanced colorectal cancer	1 st and 2 nd -line therapies	PARTIAL				X			RCTs (one sequencing)**
Rodgers, 2011	Psoriatic arthritis	2 nd line TNF-inhibitors	YES					X		Observational studies
Connock, 2006	Epilepsy	1 st and subsequent 'new' antiepileptic drugs	YES					X		RCTs
Finnerup, 2005	Neuropathic pain	Treatment algorithms	PARTIAL						X	RCTs

Abbreviations: CG clinical guidelines; DMARD disease-modifying antirheumatic drug; MTX methotrexate; RCTs Randomised controlled trials; TNF-inhibitors tumour necrosis factor-alpha inhibitors; vs versus

*Unless otherwise stated, RCTs relate to the evaluation of individual treatments; 'placebo RCTs' included a placebo control, whilst 'RCTs' included either an active or placebo control.

**Included published RCTs that reported the number of patients receiving second-line therapies made by the authors of the trials

***Quasi-sequencing trials: RCTs of 1st-line treatment with subsequent treatment predefined in protocol, or high proportion of patients went on to receive the same 2nd line treatment.

5.3 RESTRICTING INCLUSION TO TREATMENT SEQUENCING STUDIES

5.3.1 Number of studies

Six studies tried to identify sequencing studies, to evaluate the evidence on whole treatment pathways.^{200 203 204 207 209 214} This would enable the primary studies to be pooled using standard meta-analytical techniques. However, very few RCTs or prospective sequencing trials were available; those that existed were generally confined to a limited number of treatment lines, and few addressed the review questions completely. Five studies (NICE CG131²⁰⁷; Ruhe, 2006²⁰⁹; Cooper, 2011²⁰⁰; Hind, 2008²⁰⁴) limited inclusion to RCTs, whilst two (Heng, 2014;²⁰³ Stenner, 2012²¹⁴) included observational studies.

5.3.2 Limiting inclusion to studies of predefined sequences

The following section provides examples of reviews that limited inclusion to studies of predefined or fully formed sequences to develop summary effect estimates of treatment sequences. The reviews illustrate how limiting the inclusion to sequencing studies can curtail the ability of the meta-analysis to adequately inform clinical decision-making.

Two studies that evaluated chemotherapy sequences for advanced colorectal cancer limited inclusion to sequencing studies when reviewing the evidence to inform the economic evaluation. One was a technology appraisal (NICE TA93)²⁰⁴ and the other was a clinical guideline (NICE CG131),²⁰⁷ both of which are listed in Chapter 4, Table 4.2. The technology appraisal (Hind, 2008; TA93) included stratified analysis by line of therapy for the clinical evaluation. For the economic evaluation, the authors considered the findings of existing economic models to be weak because of the use of unplanned second-line therapies in their trial data, which meant that survival benefit could not be uniquely attributed to the allocated therapy. The data used to inform the clinical-effectiveness parameters of their own *de novo* economic model were therefore based solely on sequencing RCTs. The study (Hind, 2008) identified two relevant trials, FOCUS trial²⁸⁵ and GERCOR trial,²⁸⁶ which are described in the Appendix Volume I, Section B1. The review only considered treatment sequences for which RCTs were available, with inclusion subsequently limited to the seven chemotherapy sequences evaluated in these two trials. Both trials included two lines of therapy, but the FOCUS trial also included subsequent salvage chemotherapy. The NICE Clinical Guideline (NICE CG131) for colorectal cancer, in contrast, included the evaluation of 10 predefined chemotherapy sequences of up to three lines of treatment.²⁰⁷ This was underpinned by a systematic review of clinical effectiveness based on separate network meta-analyses for each line of treatment. Preliminary analyses confirmed that the overall survival for second-line treatment was dependent on first-line treatment. The authors noted that the potential influence of first-line treatments on survival had not been adequately addressed in studies of second-line treatment, which provided limited data on prior treatments. The RCTs of first-line treatments also provided limited data on subsequent treatment, with the majority of participants going on to receive a mix of second-line treatments. When the second-line treatments were reported, medians or hazard ratios for overall survival were not reported separately for each line of treatment. Data on second-line treatment for the economic model were therefore only taken from

prospectively sequenced studies. However, only two RCTs (GREGOR,²⁸⁶ CAIRO²⁸⁷) that evaluated the specific sequences of interest were identified. One further RCT comparing first-line treatments was included as a quasi-sequencing trial. The study protocol pre-specified that patients progressing on first-line treatment should only receive irinotecan second-line, of which more than 75% in both arms did. Unlike the analysis of first-line treatment, which included 23 RCTs, the limited number of available studies for second-line treatment did not form a complete network, so treatments had to be grouped by mode of action, allowing the comparison of three treatment sequences. This necessitated using the assumption that treatment effects, informed by the clinical guideline development group, were the same within each mode of action. In order to implement the analysis, assumptions were also made regarding the data from the quasi-sequencing studies. Data on median progression-free survival were not available for this study, but the median duration of second-line treatment was reported to be the same in both arms. It was assumed that mean duration of treatment was highly correlated with progression-free survival, and therefore the hazard ratio for progression-free survival for the second-line treatment was imputed as 1, with a standard error of 0.14, based on relationship between the study sample size and standard errors for all other log hazard ratios from the first and second-line treatment studies. This review highlights how better reporting of previous and subsequent treatments by RCTs of single treatments, or the availability of individual patient data (Section 1.4.1.2), would greatly enhance the ability to undertake evidence synthesis to evaluate the effects of treatment sequences. It also identified the challenge of maintaining a closed network of trials when evaluating treatment sequences.

In a comparative effectiveness review investigating the use of second-line targeted therapies for metastatic renal cell carcinoma, Heng *et al.*, included observational studies.²⁰³ They noted that the data from the available RCTs were insufficient for informing the optimal sequencing of targeted therapies (specific VEGF TKI or mTORi or their combination; see footnote¹), and that up to three lines of treatment were often used in practice. The review included only observational studies that met the following three criteria indicative of better quality:

- i. a retrospective cohort design that imposed inclusion criteria only up to initiation of second-line treatment, and then followed all included patients as long as possible for outcomes;
- ii. adjustment for patient characteristics; and
- iii. the use of data from multiple centres.

Most of the observational studies reported only class level treatment groups and most patients received a VEGF TKI in the first-line setting, so the review focused on class effect, investigating sequential treatments with VEGF TKI followed by mTORi versus VEGF TKI followed by VEGF TKI. Many studies did not adequately represent third-line treatment. They identified 12 retrospective observational studies with 2,686 patients. Seven studies were based on medical records or chart reviews, and one on a national register. Ultimately, only four studies were included in the meta-analysis for the outcome overall survival and three studies for progression-free survival, due to clinical heterogeneity. Unexplained statistical heterogeneity remained in all meta-analyses other than one limited to better quality studies for overall survival. The high level of heterogeneity precluded the

authors' ability to draw a single comparative conclusion, despite the review potentially including a large evidence base.¹

The review by Heng *et al.* identified three studies that departed from the preferred retrospective cohort design by requiring patients to receive a third-line therapy after the initiation of second-line treatment.²⁰³ This resulted in the exclusion of a large proportion of second-line patients who did not receive third-line treatment due to loss to follow up, continuation of second treatment at the time of the chart review, death during second-line treatment, or other reasons, resulting in immortal time bias for the effects of second-line treatment. Immortal time refers to an interval in the observation or follow-up period of a cohort during which the outcome under study could not have occurred.²⁸⁸ Although the authors did not include third-line treatment, they did acknowledge the need to investigate this. They recommended that an appropriate retrospective cohort design for comparing third-line treatment outcomes would follow patients after initiation of third-line treatment, and would adjust for patient characteristics available at the time initiation, including treatments received in the first and second-line. In other words, this approach would provide the summary effect of treatments used as third line, which are adjusted for the previous treatments used. However, this is still dependent on the primary study not only implementing this, but also having access to the data on previous treatments used.

An alternative review and pooled analysis of treatment sequences for metastatic renal cell carcinoma is provided by Stenner *et al.*, which focused on a narrower decision problem. They evaluated the optimum sequencing of two VEGF TKIs, sorafenib followed by sunitinib versus sunitinib followed by sorafenib.²¹⁴ This review was also based on observational studies, and included 11 retrospective case series and two prospective cohort studies with a total of 853 patients. The review does not provide anything new in terms of evidence synthesis methodology but does highlight the contrary methods and findings provided by these two reviews. This makes it challenging for decision makers to interpret the evidence and shows the need for guidance on reviewing observational studies of treatment sequencing. No quality assessment of included studies was conducted in this review. The data on progression-free survival were pooled for each sequence using weighted linear regression, with data for both sequences taken from each study but considered separately. The importance of maintaining the individual treatment comparison within each study is discussed in the Appendix Volume I (Section A) and Section 5.9.9.2. Separate analyses were conducted for progression-free survival on first-line treatment, on second-line treatment, and in total. Regression analysis was used to examine the impact of age, gender and study design. Sensitivity analyses were conducted, producing pooled

¹ *Clinical guidelines recommend initial treatment with a vascular endothelial growth factor tyrosine kinase inhibitors (VEGF TKI) for most patients, with the current standard of care generally including subsequent VEGF TKI or mammalian target of rapamycin inhibitors (mTORi). Seven targeted therapies were available as second-line treatment, four of which were VEGF TKI. The data from the RCTs were considered insufficient as only three were available in the second-line setting: one comparing an mTORi with placebo; one comparing a VEGF TKI with an mTORi; and one comparing two VEGF TKIs. The RCT comparing a VEGF TKI with an mTORi did not report data on subsequent treatments that were off-protocol, which might have influenced the results. The authors also noted that a different mTORi is used by the majority of patients in practice.*

estimates without studies that i) included patients who had previously received prior chemotherapy or anti-angiogenic compounds, ii) studies where the reason for treatment discontinuation was not clear, and iii) studies where discontinuation due to toxicity or intolerance were part of the definition of progression.

The review by Stenner *et al.* also highlighted potential limitations of using progression-free survival for evaluating treatment sequences, and showed that it is important to report how this endpoint is defined and used.²¹⁴ The optimum endpoint for evaluating sequencing of cancer treatments is discussed in the Appendix Volume I (Section C4.2). The results showed that combined progression-free survival was longer for the sorafenib-sunitinib sequence than the same therapies used in reverse, but this gain was primarily due to the progression-free survival on second-line treatment. There was no significant difference between the two sequences for progression-free survival on first-line treatment, but progression-free survival on second-line treatment was significantly longer for sunitinib than sorafenib. The authors noted that some included studies found the sorafenib-sunitinib sequence superior to sunitinib-sorafenib in terms of progression-free survival, whilst others, including the largest study (n=260) found no difference. They proposed that this discrepancy may be explained by the definition of combined progression-free survival used.²¹⁴ The authors calculated the sum of the progression-free survival period for each treatment line without considering the treatment-free interval. This combined endpoint is referred to as 'duration of disease control' by Chibaudel *et al.*²⁸⁹ and illustrated in Figure C2 in the Appendix Volume I, Section C4.2. The combined endpoint that incorporates both the progression-free period on each treatment and the interval between relapse and start of the next treatment is referred to as 'time to failure of strategy', also illustrated in Figure C2 (Appendix Volume I).²⁸⁹ The largest study in the review by Stenner *et al.* defined overall progression-free survival from the start of the first VEGF TKIs to progression on the second, including the treatment-free interval. When the progression-free intervals were analysed separately, without considering the treatment-free period, a different result with sunitinib achieving a longer interval in second-line compared with sorafenib was found.²¹⁴ Stenner *et al.* also noted that the impact of treatment-free interval on progression-free survival may be explained by a more pronounced carry-over effect after cessation of sunitinib compared to sorafenib. Patients who received first-line sorafenib may also be more willing or fit to receive second-line treatment due to less adverse effects of the drug.²¹⁴ The quality and speed of progression and number of sites involved may have also influenced the clinician's decision to start the second therapy.²¹⁴ The authors decided to omit the interval and only compare active treatment times of sequential treatments as the duration of the pause could not be adequately addressed in the pooled analysis.

The review by Stenner *et al.* was accompanied by a retrospective analysis of patients with metastatic renal cell carcinoma who were treated with either the sorafenib-sunitinib sequence or sunitinib-sorafenib at five Swiss centres.²¹⁴ This study highlighted the potential limitation of using observational studies for comparing older versus newer drugs. The authors noted that the individual treatment decision regarding the treatment sequence was driven by the registration status of the two drugs.

Patients were initially treated with sorafenib, the first drug registered for use in metastatic renal cell carcinoma, and when sunitinib became available all subsequent patients were treated with it. This could potentially lead to channelling bias, a form of confounding that occurs when a drug is preferentially prescribed to patients with different baseline characteristics.²⁹⁰ I return to this when reflecting on the issues of using real-world data from Lloyd *et al* in the section on meta-regression (Section 5.5.2)

5.3.3 The inclusion of adaptive treatment sequencing trials

The following section illustrates the challenge of reviewing treatment sequences, even when an RCT using the innovative clinical trial design for adaptive treatments, such as SMART, is available. The SMART design is summarised in the Appendix Volume I, Section B2. The studies summarised below showed that the availability of SMART on its own was insufficient to inform clinical practice as it did not cover all the relevant treatments and excluded important subgroups such as elderly patients. However, the SMART design still provides better data than RCTs of individual treatments that are vague about the previous treatments used. The studies also highlight the complexity of the decision problem, and illustrate how limitations of the primary data reduced the feasibility of conducting a meaningful quantitative synthesis to inform treatment sequences.

One of the examples of the use of the SMART design given in the Appendix Volume I, Section B2 was the STAR*D trial, conducted to determine the effectiveness of treatments for people with major depression who have not responded to initial treatment with an antidepressant.^{291 292}

The systematic review conducted by Ruhe *et al.* of the evidence for switching antidepressants in major depressive disorder included the STAR*D trial.²⁰⁹ The authors noted that several national guidelines recommended selective serotonin reuptake inhibitors (SSRIs) as first-line treatment for major depressive disorder. In the case of nonresponse, the guidelines recommended three major strategies, which included dose escalation, switching to another antidepressant, and augmenting the antidepressant with another a drug that by itself is not an antidepressant. The review investigated whether the available evidence justified distinct recommendations for next-step strategies after nonresponse to a first SSRI. The review included studies of pharmacological switching strategies for adults with major depressive disorder nonresponsive to SSRIs. Randomised, non-randomised, and uncontrolled studies were eligible, provided at least 50% of the participants had either used an SSRI in the current depressive episode or, in the case of treatment resistant depression, had well documented prior use of an SSRI. Studies describing switching from tricyclic antidepressants to SSRIs were excluded. Included studies were grouped by class of antidepressant, and where possible, risk differences and numbers needed to treat for benefit or harm were calculated for each study. The review included 21 controlled and 20 uncontrolled studies. Inadequate response or intolerance to an SSRI was only determined prospectively in seven studies, and several failed to treat non-responders promptly after stopping the unsuccessful drug. There was marked clinical and methodological heterogeneity between included studies, and therefore only three RCTs, including the STAR*D study,

were pooled in a meta-analysis, comparing switching to venlafaxine, a novel dual-acting agent, versus a second SSRI. The authors concluded that the available evidence for switching strategies did not justify distinct recommendations, and only allowed general recommendations. They also noted that the findings carried the risk of ecological fallacy, with lower remission rates due to the inclusion of patients who were more chronically depressed, from lower socioeconomic status, and suffered from comorbidities. Ecological bias, also referred to as aggregate bias, relates to the failure of aggregate-level associations to properly reflect individual-level associations.²⁹³ It was unclear how many participants had received more than one prior antidepressant, or prior dose escalation.

An alternative review of treatments for refractory depression in older people, which also considered treatment sequences, shows the inability of the STAR*D study, due to its limited inclusion criteria, to provide evidence to inform clinical decision making.²⁰⁰ The authors specifically noted that the STAR*D study did not provide suitable evidence for a population of older people. The optimum sequencing for older people was likely to differ due to higher rates of physical and cognitive comorbidities, differing social circumstances, greater likelihood of polypharmacy, and age-related pharmacodynamics and pharmacokinetic changes. The review included 14 studies, three of which were RCTs and ten uncontrolled open-label trials. The review included 'sequential treatment or trial' as search terms, but only identified two studies of treatment sequences, both were small open-label trials. One provided the highest response rate out of all included studies, and the other did not report non-completers and non-responders separately for the subsequent treatments and therefore only data for the first treatment could be used. In summary, the review failed to evaluate treatment sequences for refractory depression in older people due to the limitation of included studies.

5.3.4 Lessons about studying treatment sequences that have emerged from restricting inclusion to sequencing studies

The ideal scenario to inform evidence-based treatment algorithms would be to undertake meta-analyses of sequencing RCTs but, even before conducting the review of methods, it was clear that it is not feasible to undertake RCTs of all conceivable sequences. The findings also demonstrated that, when using the endpoint progression-free survival to evaluate whole sequences based on data from individual treatment lines, it is important to report how it is defined and utilised within the synthesis. It is essential to note whether the sum of the progression-free period for each treatment line also incorporates the interval between relapse and start of the next treatment or not. I re-visit this in Chapter 7, Sections 7.3.2 and 7.3.4.

In summary, relying solely on treatment sequencing studies, either clinical trials or observational studies, is unlikely to provide sufficient evidence to inform clinical decision making. This means that in order to make the most of the available evidence base, there is a need to use data from comparative studies of single treatments. However, the summary effect of treatments would need to be adjusted for the previous treatment used. The next sections explore methods that could be used to achieve this.

5.3.5 Lessons learned about the use of observational data

The review findings showed that observational studies based on registry data may be able to provide data on the effectiveness of whole sequences or a specific line of treatment, which is adjusted for patient characteristics at the time of initiating the treatment, including the treatments received in previous lines. However, the use of such studies in a meta-analysis needs to include adjustment for any potential bias.

A number of limitations of using observational data for evaluating treatment sequences were identified. These are in addition to the usual concerns about using observational studies to inform treatment effects, which are discussed in more detail in the Appendix Volume I (Section A). The additional limitations relating to treatment sequencing include the following:²⁰³

- i. The potential for immortal time bias affecting the assessment of, for example, second-line treatment in retrospective studies that include patients receiving third-line treatment, as this will exclude patients who do not reach third-line, are continuing second-line treatment, lost to follow up, or other reasons;
- ii. The comparison between drug classes may be confounded by differences in the type of patients treated with each class;
- iii. The potential for missing or inaccurate data obtained from real-world practice; and
- iv. Having to pool across treatments at class level as most studies do not report drug level data when there is evidence that individual drug effects can vary within class.
- v. The choice of treatment may be driven by the registration status of the individual drugs, which in turn could also influence the clinician's decision to switch treatments rather than adjust the current treatment.²¹⁴

5.4 SUBGROUP AND STRATIFIED ANALYSES

5.4.1 Number of studies

Subgroup analysis is generally used to determine whether the treatment effect is modified by a value of another variable.²⁹⁴ Three studies were identified that used subgroup analysis to investigate treatment-sequencing effects, all of which included patients with rheumatoid arthritis.^{12 20 215} Three studies, comparing the efficacy of biologics used at a single point in the treatment pathway, are included here as further examples.^{210 212 213} These did not specifically aim to evaluate treatment sequencing as such, but did explore the effect of previous treatments using subgroup analysis. This type of subgroup analysis is common in company submission to recent NICE technology appraisals, as it is required by the manufacturers submission template.^{283 295} A summary of the included studies is presented in Table 5.3 and Appendix Volume II (Appendix C).

There are two ways of applying subgroup analysis within meta-analysis.

- i) The subgroups are defined by splitting all studies into two or more groups (e.g. randomised vs non randomised study design)

- ii) The subgroups are defined by taking partial data from studies, which means the same study can appear in both subgroups (e.g. male only and female only analysis)

The former does not require any more data than the overall analysis and is referred to here as stratified analysis, whilst the latter requires additional data on specific covariates. The method used by included studies is provided in Table 5.3. Dividing the all studies into groups can also be done using a categorical covariate in meta-regression,²⁹⁶ which is summarised in the next section (Section 5.5).

The correct way to analyse subgroups is by using a statistical test of subgroup treatment-effect interaction.²⁹⁷ When comparing two estimates of the same quantity from separate analyses, it is not enough to note that the intervention was statistically significant in one group and not in another.²⁹⁸ Significant benefit, or harm, is likely to be absent in small subgroups or populations that are underrepresented in RCTs, which could lead to the mistaken finding that an intervention is ineffective in a subgroup. Only one included study (Lloyd, 2010) used statistical tests for interaction, which also examined the statistical significance of various covariates using regression analysis.²⁰

5.4.2 Subgroup and stratified analysis

The use of subgroup analysis can be interpreted as evaluating treatment sequences in a piecemeal fashion. Two studies examined differences in treatment effects according to the number of previous biologics used,^{12 212} and four studies stratified analysis according to the type of treatments previously failed.^{20 210 213 215} Of note, the distinction between 'used' and 'failed' was not always clear and may in fact be the same thing.

Systematic reviews frequently include separate meta-analyses for interventions used in different treatment lines as illustrated, for example, in the study by Salliot *et al.* which undertook separate meta-analyses for first- and second-line biological agents.²¹⁰ The studies that did this, or conducted a meta-analysis of treatments used at a single point in the pathway (e.g. first-line) are referred to in Tables 5.1, 5.2 and 5.3 as having used stratified analysis. This is generally done to limit clinical heterogeneity. Another approach used is to 'lump' all treatment lines in a single meta-analysis, and then use subgroup analysis to explore heterogeneity. An example of this is provided by an overview of Cochrane reviews of all biological agents for rheumatoid arthritis by Singh *et al.*²¹³ Both Salliot *et al.* and Singh *et al.* defined subgroups by splitting all included studies to two or more groups. Two included studies used subgroup analysis to investigate the influence of disease duration,^{213 215} which was accounted for in the accompanying meta-regression analyses in one study (Lloyd, 2010).²⁰ Two studies used subgroup analyses to explore the influence of the reasons for switching treatment.^{12 20} and two studies provide examples of where this was precluded due to poor reporting by primary studies.^{210 212}

5.4.2.1 Poor reporting of primary studies limiting the analysis

Three studies (Lloyd, 2009²⁰; Schoels, 2012;²¹² Rendas-Baum, 2011¹²) encountered problems with poor reporting of the data by primary studies, which limited the extent and level of analyses that could

be implemented. However, these studies also failed to report their findings on sequencing effects in detail.

Poor reporting of primary studies limiting the analysis of specific previous treatments and reason for switching

The first study (Lloyd, 2009) explored differences in treatment effects by using subgroups based on the type of TNF-inhibitor switched to and from, and the reason for switching from one biologic to another.²⁰ They noted that of 20 included studies, three did not specify the biologic agent used, and one did not specify the agent that was switched to. Eleven studies did not differentiate between primary (lack of response) and secondary (loss of response) inefficiency, and only three studies reported on all four relevant outcome measures. However, Lloyd *et al.* themselves did not report their actual results, only a narrative stating that the subgroups had no significant effect on pooled estimates.

Poor reporting of primary studies limiting analysis of the exact number of previous treatments

A second study (Rendas-Baum, 2011) examined the relationship between the clinical response to any biological agent and the number of previous TNF-inhibitors received. Pooled weighted averages of treatment effects, based on sample size, were calculated for each number of up to four previous biologics used. However, weighted averages were also estimated for a '2+' category, because a number of studies did not report results disaggregated by number of previous biological agents and only gave results for the biologic under evaluation in patients with inadequate response to at least one previous biological agent. Efficacy was evaluated for four subgroups, including type of biological agent and reason for discontinuing previous biologic treatment. However, the authors themselves only presented their results as bar graphs, for visual comparison of trends. No statistical tests of interaction were conducted, nor were estimates of the variance of the pooled treatment effects provided.

Limited availability of primary studies precluded analysis of the number of previous treatments used

The third study (Schoels, 2012) also examined differences in treatment effects according to the number of previous biologics used, and highlights the limitation of this approach for investigating new treatments.²¹² They compared the efficacy of 'new' biological agents in patients with an inadequate response to previous TNF-inhibitors. They also compared the efficacy for each biological agent after only one previous TNF-inhibitor with outcomes after multiple TNF-inhibitor drug failures. Although this method can potentially provide useful information on the impact of sequential TNF-inhibitors or the number of previous treatments, the review was limited by the lack of available studies. As is often the case for the evaluation of new treatments, only one RCT for each new drug was available. Again, due to the lack of significant findings, only limited results were reported, which were presented only narratively, noting that efficacy rates did not differ with a history of one or multiple previous TNF-inhibitor failures. Stratified analyses were also presented for the indirect comparison of golimumab versus tocilizumab according to the number of previous TNF-inhibitors at baseline, 1, 2, or 3, for three

separate response outcomes. Subgroup numbers were small, and the proportion of participants who discontinued previous biologic agents due to inadequate response varied considerably between the two new biologic agents.

5.4.2.2 Different approaches used for incorporating observational studies, which were included to account for the limited primary studies

The evaluation of sequential TNF-inhibitors required the use of observational studies due to the lack of RCTs evaluating a second or subsequent line TNF-inhibitor. The included studies used different approaches for incorporating observational studies, and none tried to account for potential bias in the observational studies. Suraez-Almazor *et al.*, who only compared two TNF-inhibitors, investigated differences in efficacy according to the timing of starting treatment, and subsequent switching between agents.²¹⁵ The review included both RCTs and observational studies. However, only observational studies were identified for the TNF-inhibitors used as second-line treatment, or switching between agents. No meta-analyses were conducted due to clinical heterogeneity. Rendeas-Baum *et al.*, on the other hand, included both observational studies and RCTs in their meta-analysis by using unadjusted data from the treatment arms from the biologic RCTs, thus breaking randomisation (as discussed in the Appendix Volume I, Section A).¹² Lloyd *et al.*, conducted both meta-analysis and meta-regression using observational studies only, and is also included in Section 5.5.²⁰

5.4.2.3 Poor reporting or primary studies problematic when conducting stratified meta-analysis

Poor reporting of prior treatments by primary studies was still a problem even for systematic reviews that investigated treatment sequences in a piecemeal fashion, or by 'splitting' studies according to line of therapy. Salliot, *et al.* compared the clinical efficacy of biological therapies within two separate analyses according to the study patient populations, representing first or subsequent use of biological agents.²¹⁰ Only non TNF-inhibitors were evaluated in RCTs of patients with an inadequate response to a previous biological agent. Despite stratifying the analyses, there was still clinical heterogeneity in the analysis of biological agents used as subsequent treatment, which was thought to be due to variation in the number of previous biological agents and reasons for treatment failure. The authors noted that poor reporting of primary studies, and inability to ascertain precise numbers for these variables were limitations for their analyses.

5.4.2.4 Limitations of using subgroup analysis for evaluating treatment sequencing

The Cochrane overview (Schoels, 2012) which 'lumped' different lines of therapy in the same analysis, demonstrates why doing a series of subgroup analyses may not be helpful for evaluating treatment sequencing, and highlights the problem of subgroup analyses having limited power.²¹³ The main analysis included studies of patients receiving biological agents as their first, second, or subsequent line biological therapy. A series of planned subgroup analyses then evaluated the effect of several factors including the type of previous treatments failed, classified as none, conventional

DMARDs, or biological agents. They also investigated whether the previous failed treatment was a TNF-inhibitor or not, whether the current treatment was a TNF-inhibitor or not, and whether disease duration was classified as early, established or late. Biological agents were evaluated as single agents as well as by class for the TNF-inhibitors, but only differences in the class effect were examined in the subgroup analysis. The results were presented in terms of relative risk of treatment response plus 95% confidence interval for each individual subgroup, with no statistical test for interaction. As only one variable was considered at a time, the resulting number of individual subgroups analysed made it difficult to interpret the overall findings. As noted by the authors, the analyses were also susceptible to type II error due to the small number of studies within the subgroups. A type II error is a false negative, or a failure to reject a false null hypothesis, due to inadequate power.^{299 300} A number of reviews evaluated multiple outcome measures as well as numerous subgroups, which increases the likelihood of type I error.^{299 300} A type I error is a false positive, or rejecting a correct null hypothesis (i.e., falsely inferring the existence of something that is not there), due to multiple comparisons resulting in chance findings. For example Rendas-Baum *et al.*, evaluated efficacy using several different response measures over four subgroups.¹² They also had a limited number of studies within each subgroup. Many included studies failed to report the actual data for subgroup analyses that did not show statistically significant results.

The study by Suraez-Almazor *et al.*, which used subgroup analyses to investigate the optimum timing of introducing TNF-inhibitors, demonstrates how not to interpret the findings of subgroup analysis and the importance of using a statistical test for interaction.²¹⁵ Subgroup analyses frequently have limited power, and a small number of events or studies included in one of the two subgroups could make the estimates and p values appear very different by chance, or result in wide confidence intervals. Suraez-Almazor *et al.* meta-analysed RCTs of TNF-inhibitors used as first-line treatment, and used subgroup analysis to assess differences in effects according to whether the patients had previously received methotrexate or not, and whether they had early disease duration, of less than two years, or longer. The results for the subgroups were compared using informal indirect analyses, with statistical significance established by examining the point estimates of the treatment effect and whether the 95% confidence intervals overlapped, rather than the correct approach using a test of interaction for the statistical comparison between two estimates of the same quantity from separate analyses.²⁹⁸ The statistical test should target the question at hand, and it is not sufficient to compare the p-values or confidence intervals from separate analyses.²⁹⁸

A methodology review was conducted by Thorlund *et al.* to identify issues that can explain the discrepancies in the findings of network meta-analysis of biological agents for rheumatoid arthritis.³⁰¹ The methodological items that were assessed included objectives, eligibility criteria, databases searched, completeness of trial inclusion, bias risk assessment, effect measures, statistical methods and additional analysis. The comparison of statistical analysis included, among other items, the methods employed for dealing with DMARD-naïve and DMARD-experienced patients. The review identified 13 published network meta-analyses, which were shown to have major discrepancies. Six

studies included DMARD-naïve patients, patients with inadequate response to TNF-inhibitors (TNF-IR), or both, in addition to patients with inadequate response to previous conventional DMARD (DMARD-IR). Of these, two produced stratified analyses by patient groups, one developed an interaction analysis including the group term in the regression model, and three inappropriately lumped data across groups without controlling or separating (two studies lumped DMARD-naïve and DMARD-IR, and one lumped DMARD-IR and TNF-IR).³⁰¹ Three of the six studies were also included in the current review: Nixon, 2007²⁰⁸ that lumped studies; Mandema, 2011²⁰⁶ and Salliot, 2011,²¹⁰ which used stratified analysis; and Singh, 2009,²¹³ which performed an interaction analysis.

5.4.3 Lessons about studying treatment sequences that have emerged from studies using subgroup or stratified analysis

The review has identified that, whilst subgroup analysis can be used to answer the question of whether the efficacy of an individual treatment is different in patients with an inadequate response to a specific previous treatment compared to those who had never used it, this approach can be limited by the poor reporting of previous treatment and a potential small number of studies. Similarly, subgroup analysis would benefit from having access to individual patient-level data. Overall, the findings of the reviewed subgroup analyses were generally poorly reported and statistical tests of interaction were rarely conducted. The review identified the following drawbacks of using subgroup analysis as a way of exploring sequences:

- i. Subgroup analysis does not enable treatment sequencing to be evaluated in great depth, or provide conditional or adjusted effect estimates.
- ii. Subgroup analysis can only test for an interaction (whether it is zero) between the treatment effect and a single covariate representing treatment history, and cannot estimate the extent of this interaction.
- iii. Subgroup analysis is based on categorical data, and included studies used a simplified, dichotomised or categorised summary of previous treatments, for example number of treatments failed or a pre-specified treatment graded as yes/no. Although a more elaborate remodelling of discrete variables could have been adopted, for example using coding dummy variables for first-line, second-line, and third-line treatment etc.
- iv. Subgroup analysis can produce misleading results due to confounding and aggregation biases (ecological fallacy).
- v. Subgroup analysis investigate only one factor at a time, and cannot account for the simultaneous influence of other important factors. They are unable to demonstrate the independent effect of previous treatments.
- vi. With subgroup analysis as the number of outcomes and covariates being evaluated increases, the likelihood of a type I error, or producing the wrong results due to chance, increases.
- vii. Non-significant results of subgroup analysis are likely to be common due to lack of power (type II error), especially when previous treatments are poorly reported, or the evidence base is limited.

- viii. Subgroup analysis are frequently hampered by poor reporting of previous treatments, which also means that the influence of important factors such as reasons for switching treatments cannot be evaluated.
- ix. It is difficult to interpret the findings of a series of multiple subgroup analyses.

Some of the limitations of subgroup analysis can be overcome by using meta-regression, which represent extensions of subgroup analysis that allow the effect of continuous or categorical variables. For example, meta-regression also allows the extent of the interaction to be investigated, not just the statistical significance, as well as the evaluation of more than one factor at a time. The next section explores the use of meta-regression.

Table 5.3: Summary of included rheumatoid arthritis studies using subgroup-analysis

Study	Decision point – treatments of interest	Patient population - treatment history	Outcome measures	Covariates related to sequencing included in analysis [no. of studies if 10 or below]	Available evidence base*	Results
Lloyd, 2011	1 st -line or 2 nd /3 rd -line TNF inhibitors (<i>Sequential TNF-inhibitors</i>)	TNF-naive vs TNF-IR	Achieving ACR50 and EULAR response Mean improvement in DAS-28 and HAQ scores	Subgroup analysis: <ul style="list-style-type: none"> Type of TNF switched from and to [ADA, ETA, INF] Reason for switching (intolerance, primary inefficacy; secondary inefficacy; either inefficacy [<i>most studies did not differentiate between primary and secondary</i>]) Meta-regression (included 3 covariates): <ul style="list-style-type: none"> Mean no. of previous cDMARDs [n=10] Disease duration Duration of TNF 	Observational studies (uncontrolled and sequencing studies)	Subgroup analysis: No significant difference identified (<i>no results presented</i>)
Rendas-Baum, 2011	1 st , 2 nd or subsequent-line biologics	TNF-IR	Achieving ACR (20/50/70), EULAR (good/moderate) and DAS-28 ($\geq 3.2/\leq 2.6$) response criteria (7 outcomes)	<i>Data analysis based on individual patients with efficacy ('clinical response' rate) estimated for groups of patients according to no. of previous TNFs received (1, 2, 2+, 3, or 4). Pooled weighted averages, based on sample size (of each group), were developed for each outcome measure (n=7).</i> Estimates also evaluated for the subgroups: <ul style="list-style-type: none"> Type of biologic (TNF vs other) Reason for switching to 2nd TNF (intolerance, lack of efficacy, loss of efficacy) Type of study (RCT vs observational) duration of follow-up Evaluation of reason for discontinuation: Response to a 2 nd TNF stratified by reason for discontinuing 1 st TNF	RCTs and observational studies; analysis based on individual patient data breaking randomisation	<i>No formal statistical inference was undertaken.</i> There was a trend of lower efficacy rates with increased number of previous TNF-inhibitors used. The magnitude of the decline may depend on the type of biologic used. Patients who discontinued treatment due to adverse effects were more likely to achieve a response to a 2 nd TNF than patients who discontinued due to efficacy-related reasons. (<i>The available data precluded comparison of patients who switched to a 2nd biologic that was not within the TNF class</i>)
Suarez-Almazor, 2007	1 st - and 2 nd -line TNF-inhibitors (switching between INF and ETA)	1 st -line TNF: cDMARD[MTX]-naive or cDMARD[MTX]-IR. and 2 nd -line TNF: TNF[INF/ETA]-IR	Achieving ACR50 and ACR70 response	<i>Stratified MA conducted for 1st- and 2nd- line TNF inhibitors</i> Subgroup analysis for 1st-line TNF inhibitors, with RCTs categorised according to: <ul style="list-style-type: none"> Type of previous treatment failed (patient population: MTX-naive [n=2] vs MTX-IR [n=1-2]) Disease duration (early [n=2-3] vs established/late stage [n=1]) Multiple MAs conducted over different outcomes using fixed effect and random effects models <i>narrative synthesis for 2nd-line TNF-inhibitors (due to heterogeneity)</i>	1 st -line: RCTs; 2 nd -line (switching between INF and ETA); observational studies only (not pooled)	<i>Statistical significance established by comparing the overlap between 95% confidence intervals for individual treatment effects.</i> There was statistically significant differences favouring TNF-inhibitors (plus MTX vs MTX alone) in patients with longer disease duration or patients who had failed MTX. <i>RCTs often compared INF or ETA to MTX in patients who had already failed MTX. When the trials included a true MTX control group of naive patients, the results showed less benefit with INF or ETA (compared to MTX) than in RCTs that included patients who had failed MTX.</i> 2nd-line TNF-inhibitors: All but one study reported that most patients who failed one TNF-inhibitor can respond to another one after switching.

Schoels, 2012	2 nd or subsequent-line biologics ('new' biologic after multiple TNF-inhibitors)	TNF-IR	Achieving ACR20, ACR50 and ACR70 response	<p>Subgroup analysis only conducted for 2/4 included biologics (within an indirect treatment comparison MA of TOC vs GOL):</p> <ul style="list-style-type: none"> No. of previous TNFs (1 vs multiple previous TNFs) [n=2] No. of previous TNFs (1, 2, or 3 previous TNFs) [n=2] <p>(Only 1 RCT available for each biologic; analysis based on partial data from trials)</p>	Placebo RCTs ('new biologics')	<p>Efficacy after one vs multiple TNF-inhibitors did not differ significantly. Response rates for GOL and TOC did not differ significantly when assessing separately those who had failed 2 or 3 previous TNFs, but failure of 3 TNFs occurred in few patients and confidence intervals were wide.</p> <p>Only 2/4 RCTs reported the proportion switching treatment due to inefficacy, ranging from 58% (GOL) to 95% (TOC).</p>
Singh, 2009	1 st , 2 nd , or subsequent-line biologics	cDMARD-naïve, cDMARD-IR, TNF-naïve, or TNF-IR	Achieving ACR50 response	<p>Overview of Cochrane reviews. Individual drugs compared with placebo, and then as a single group (biologics vs placebo); ITC between biologics also conducted. Planned subgroup analyses conducted for 'biologics' (as a single group) vs placebo. Analyses performed using a generalised linear mixed model (accounting for heterogeneity between the drugs due to study and that due to study x drug interaction).</p> <p>Subgroup analysis (7 in total) with RCTs categorised according to:</p> <ul style="list-style-type: none"> Type of previous treatment failed (none vs cDMARDs vs biologics) [n=27, but only 2 for 'none'] Failed previous TNF (yes vs no) [n=27, but only 5 for 'yes'] Biologic = TNF (yes vs no) Disease duration (early vs establishes vs late stage) [n=27, but only 5 for 'early'] 	Placebo RCTs	Biologics were similarly effective regardless of stages of disease, type of drug previously failed, whether previous TNF treatment had failed, or whether the biology used was a TNF or 'other'. Limited number of studies for some subgroups increased susceptible to type II error. There was heterogeneity in placebo group (due to concomitant cDMARDs/MTX in some studies).
Salliot, 2011	1 st and subsequent-line biologics	1 st -line biologic: cDMARD [MTX]-IR; subsequent biologic (only non-TNFs): TNF-IR	Achieving ACR50 response	<p>Stratified MA for 1st and 'subsequent' biologic</p> <p>No subgroup-analysis considered for 1st-line biologics (comparing TNFs as a group vs non-TNFs as a group and individual agents)</p> <p>Planned subgroup analysis precluded for subsequent-line biologics (only included non-TNFs) due to limited number of studies [n=5] and poor reporting:</p> <ul style="list-style-type: none"> No. of previous TNFs Reason for switching [n=2] Duration of prior TNF-inhibitor 	Placebo RCTs	Significant clinical heterogeneity in MA of subsequent biologic, reported as being likely due to the number of previous biologics and reasons for treatment failure. Limited number of studies precluded a-priori planned subgroup analysis.

Abbreviations: ADA adalimumab; ARC American College of Rheumatology score; cDMARD conventional disease modifying antirheumatic drug; DAS disease activity score; EULAR European League Against Rheumatism classification; ETA etanercept; GOL golimumab; HAQ Health assessment questionnaire score; INF infliximab; IR inadequate response; MA meta-analysis; MTX methotrexate; no. number; RCT randomised controlled trial; TNFs tumour necrosis factor-inhibitors; TOC tocilizumab.

5.5 THE USE OF META-REGRESSION

5.5.1 Number of studies

Meta-regression is generally used to explore heterogeneity in meta-analysis and to explain the difference between intervention effects in a collection of studies.²⁹⁶ Summary effect sizes (e.g risk ratio) that are adjusted for study-level covariates are produced by regressing the effect size from each study onto the covariate of interest within each study.²⁰⁸ The regression coefficient obtained from the meta-regression describes how the treatment effect (dependent variable) changes between subgroups of studies, in the case of a categorical explanatory variable, or with one-unit increase in a continuous explanatory variable.

The review included eight studies on meta-regression relevant to treatment sequencing.^{20 21 198 202 205 206 208 211} However, only one of these specifically aimed at evaluating treatment-sequencing effects, which included the study by Lloyd *et al.*, that evaluated the sequential use of biological agents for rheumatoid arthritis.²⁰ Five further studies exemplify the use meta-regression techniques to explore the effect of prior treatments or disease duration. These studies build on the findings of study by Lloyd *et al.*, and also focus on the evidence syntheses of biological agents for rheumatoid arthritis.^{21 205 206 208 211} One additional study, by Grothey *et al.*, demonstrates the use of meta-regression to investigate the influence of subsequent treatments on overall survival for advanced cancer, with a partial aim concerned with evaluating treatment sequencing: it used linear regression to correlate the percentage of patients receiving second-line therapy with the percentage of patients who had received three active chemotherapies for advanced colorectal cancer.²⁰² The methods used in the included studies are summarised in Table 5.4.

5.5.2 Meta-regression used for evaluating treatment sequences

An overview of the clinical scenario, and the available evidence base, is provided in the Appendix Volume I (Section C3).

The study by Lloyd *et al.* was a systematic review and meta-analysis of the effectiveness of sequential TNF-inhibitors to ascertain whether a second or a third TNF-inhibitor is still effective.²⁰ The only relevant studies identified were observational, and mostly uncontrolled. Conventional pair-wise meta-analyses were used to estimate the treatment effect of TNF-inhibitors used as a class. These were initially based on single-arm data with univariate meta-regression used to try to explain the heterogeneity using study level demographic covariates, including disease duration, previous conventional DMARDs, and duration of biologic treatments. Subgroup analyses and meta-regression were then used to determine if there were differences in outcomes according to the type and sequence of TNF-inhibitors received, or reasons for switching. Separate meta-analyses were conducted for developing comparative effects of sequential TNF-inhibitors compared with first-line use. These were based on four studies that compared outcomes for patients receiving TNF-inhibitors as second or third-line treatment with other patients taking their first TNF-inhibitor, which may or may not include the group that subsequently switched. The results of the meta-regression analyses were

not presented for demographic variables that were not found to be statistically significant. Disease duration was the only covariate that was found to have a significant effect, on only one outcome measure, change from baseline response in disease activity score (DAS-28). The results showed that there was an extra reduction in the DAS-28 score, representing a worsening of the condition, of -0.16 (95% confidence interval: -0.24 to -0.07) per additional year. However, disease duration is likely to be correlated with the number of previous treatments used, as the likelihood of failing prior treatments will increase with increasing disease duration. The meta-regression was based on univariate analysis, so the potential effect of other covariates was not accounted for.²⁰ However, there is a need to try to disentangle whether long standing disease *per se* is associated with a poor response to treatment, or whether prior failure on previous treatments predicts response to subsequent treatment.³⁰² This may provide justification for including previous treatment as a covariate in multivariate meta-regression analyses, even when it is not found to be statistically significant in univariate analyses. The results of the single arm meta-analyses (based on patient change from baseline) showed that there was considerable heterogeneity for all four outcome measures used.²⁰ This heterogeneity was not explained by either the meta-regressions or subgroup analyses. The results of the comparative meta-analyses showed superior responses in patients receiving TNF-inhibitors for the first time compared to subsequent use.²⁰ Based on their findings, the authors concluded that sequential use of TNF-inhibitors was likely to have a beneficial effect, but the probability of achieving response, and the magnitude of that response, is lower with subsequent use of biologics than first-line use.²⁰ They were unable to make any conclusions about the specific TNF-inhibitor used or reasons for discontinuing prior TNF-inhibitors, due to the limitation of the available evidence.

5.5.2.1 Reflections on the use of real-world data

The study by Lloyd *et al.* highlights some of the challenges and limitations of using observational data, such as that obtained from registry studies.²⁰ The study used observational studies as no relevant RCTs were identified. However, the extent of the bias in the observational studies was not estimated or taken into account in the analyses, which is equivalent to assuming that it does not influence treatment-sequencing effects. The results represent the comparison of TNF-inhibitors used during an earlier versus a later part of the treatment pathway, ignoring the likely effect of disease trajectory, changes in pathophysiology and pharmacokinetics with time, as well as other confounding factors. It also did not account for the potentially increasing proportion of patients who do not respond to any TNF-inhibitor (discussed in the Appendix Volume I, Section C3).³⁰³⁻³⁰⁵ The authors did, however, acknowledge the likelihood of selection bias being present in the available comparative data (4 studies), which was based on cohorts where the patients were receiving biologic treatment for the first time, and that of patients who fail biological agents and therefore have a worse prognosis and are likely to show limited responses to all treatments.²⁰ Another potential source of bias in the observational data, is a phenomenon known as regression to the mean. This is because patients tend to be treated with a subsequent TNF-inhibitor at the height of their disease activity, where there is a greater than 50-50 likelihood that the disease activity will start improving after the intervention purely by chance.^{302 306 307} In a narrative review conducted by Rubbert-Roth *et al.*, which included the same

observational studies as Lloyd *et al.*, the authors concluded that the decline in efficacy after switching to a second TNF-inhibitor was likely due to both a class effect and a channelling bias, favouring patients with more severe disease.³⁰³ These same issues are likely to arise in other chronic conditions, where new treatments are being introduced. The review also highlights the limitation of the research question comparing first-line versus subsequent use TNF-inhibitors to inform treatment-sequencing effects, as the 'true' population treatment effect will be different in the two settings or patient populations, with the latter equating to a more severe and chronic disease condition. In other words, an estimate of the 'absolute sequencing effect' cannot be obtained in the same way as that of an individual treatment would be obtained (as discussed previously in Chapter 1, Sections 1.3 and 1.5).

5.5.3 Meta-regression used to account for the number of previous treatments or disease duration

The studies summarised in this next section provide examples of the challenges of including previous treatments as a covariate in meta-regression, primarily due the poor reporting of primary studies.^{20 21 205 206 208 211} There are also examples of methodological approaches that may provide useful groundwork for developing a novel method for evaluating treatment sequences in the future, including the combined use of network meta-analysis and meta-regression and model based meta-analysis.

The comparative effectiveness of biological agents for rheumatoid arthritis is sometimes assessed using meta-analyses that pool across studies, representing their use at different points in the treatment pathway with meta-regressions used to explore between the study heterogeneity. The inclusion of covariates, such as the number of previous treatments, in the final meta-regression is generally dependent on significant findings in prior univariate analysis.^{20 205} The results of non-significant preliminary analysis are rarely presented, and the lack of statistical significance may be due to limited power.^{20 211} Insufficient power to detect an association is one of the main challenges for using meta-regression. Poor reporting of previous treatments is likely to contribute to this, as is the need to limit inclusion to studies reporting a specific type of outcome data. For example, the primary analysis in the study by Lloyd *et al.* was based on change from baseline in the Health assessment Questionnaire (HAQ) score, with studies that reported proportion of patients achieving a clinically meaningful improvement in HAQ score, or dichotomous outcomes, being excluded.²⁰

The first study to combine the use of meta-regression with network meta-analysis in order to develop models that allow for the simultaneous comparison of multiple competing treatments while adjusting for study level covariates was conducted by Nixon *et al.* The study compared the efficacy of four biological agents, three of which were TNF-inhibitors, used in early or late stage disease.²⁰⁸ The review included biological agents used as monotherapy or combination therapy with methotrexate, and studies that included patients who had not previously received any DMARDs and those who had had an inadequate response to these drugs were lumped together. There was significant heterogeneity across baseline characteristics of included studies, for example, the mean number of

previous DMARDs ranged from 0 to 4. Four RCTs included patients who had not previously tried methotrexate, the results of which showed consistently lower odds of response than those where patients had previously used methotrexate. In order to avoid problems with insufficient power, only two study-level covariates were incorporated in the meta-regression, a measure of the baseline disability using HAQ, and disease duration. Efficacy was measured using the log odds ratio of achieving the American College of Rheumatology (ACR) 50 responder status at six months. The modelling was done gradually, increasing the complexity over a series of four models, outlined below. Models 1 and 3 were network meta-analyses; 2 and 4 were meta-regression models, which augmented models 1 and 3, respectively. A random-effects meta-regression model was fitted that adjusted the log odds ratio for the two study-level prognostic factors. A different random effect distribution on the log odds ratios was allowed for each different treatment. The odds ratio was found as a function of the prognostic factors for each treatment.

Model 1 was a network meta-analysis with univariate random effects. The model assumed exchangeability between treatment arms (both within a study and between studies), and that the effect of methotrexate was the same in each study.

Model 2 was a network meta-analysis with bivariate random effects. The model assumed exchangeability between treatment arms (both within a study and between studies), and that the effect of methotrexate was the same in each study. Here the study level characteristics, average baseline disease duration and average baseline HAQ, were included as treatment-disease duration and treatment-HAQ interaction effects to assess how they affect the effect estimate.

Model 3 was a network meta-analysis with bivariate random effects, including meta-regression coefficients. Here methotrexate was treated in a similar way to a biologic, and allowed to be exchangeable between studies.

Model 4 included adding different random effects for TNF-inhibitors and anakinra. The average baseline disease duration and average baseline HAQ of patients were included as meta-regression co-variables in the same way as model 2, and with the same random effects structure as model 3. The results showed that disease duration increased the odds ratio advantage of biologics by a factor of 1.13 per year.

The methods presented by Nixon *et al.*, were based on a binary outcome measure of achieving treatment response.²⁰⁸ Schmitz *et al.*, further extended these methods of combining the use of network meta-analysis with multivariate meta-regressions, in order to incorporate continuous outcome measures for disease severity.^{211 308} This is important as it can increase the power to detect statistical significance. Schmitz *et al.* compared the results of network meta-analyses of the same RCTs using binary versus continuous outcome data.^{211 308} Two popular response measures for rheumatoid arthritis were either translated from a binary to a continuous scale, or the other way around. The

former included the binary ACR response criterion, and the latter the HAQ score. The results of the analyses based on both binary and continuous response measures were compared in terms of power to detect differences between treatments. The findings showed that information is lost when continuous variables are dichotomised into a binary outcome measure, which results in a loss of power to detect differences between treatments in a network meta-analysis.³⁰⁸

The study by Schmitz *et al.* compared the efficacy of first-line TNF-inhibitors in patients with an inadequate response to methotrexate, i.e. all patients had established or late stage disease.³⁰⁸ There was heterogeneity between included studies, in particular in relation to severity of the disease at baseline, and dose of methotrexate. The statistical models were therefore extended to meta-regression to evaluate the effects of the number of previous DMARDs, disease duration, and disease severity, but as the findings were not statistically significant these covariates were not included in the final analyses. No details were reported of the regression analysis, and the number of participants and events for each grouping was not stated. Despite the inclusion of a sufficient number of studies for evaluating treatment efficacy, this may not have been the case for previous treatments as they were not always reported in the primary studies. Sixteen placebo-controlled trials were included in the review, with the number of previous DMARDs ranging from one to three in ten trials, but 'not assessed' for the remaining studies. A more recent systematic review by Kanters *et al.*, which aimed to explore which clinical features and characteristics are associated with the magnitude of comparative efficacy between biologics and methotrexate, in patients with inadequate response to methotrexate, also experienced a similar problem.²⁰⁵ Univariate meta-regression analyses were performed only for covariates that were reported in at least 10 studies. As only eight of the 22 included studies reported the number of previous DMARDs failed, this was not included in the analysis.

Evidence corroborating the presence of a correlation between disease duration and number of previous treatments used was identified in a more recent meta-epidemiological study of all trials evaluating a targeted therapy (biological or the newer targeted synthetic DMARDs) approved by regulatory authorities for treating rheumatoid arthritis.¹⁹⁸ The study (Christensen, 2015) investigated whether variations in trial eligibility criteria and baseline characteristics modify treatment effects using both subgroup analysis and meta-regression.¹⁹⁸ The study was limited to trials investigating the use of targeted therapies as add-on therapy. Included studies were pooled using conventional pairwise meta-analysis to investigate the level of heterogeneity, and effect modifiers were then identified in meta-regression analyses. The included studies were also stratified according to the DMARD history of included patients, and grouped into one of the following categories:

- i. DMARD-naïve (patients were either conventional synthetic DMARD naïve or had not exhausted the treatment potential of at least one conventional DMARD)
- ii. DMARD-inadequate responders (where patients had exhausted at least one conventional DMARD option and had inadequate response)

- iii. Targeted therapy-inadequate responders (patients had experienced an inadequate response to at least one previous targeted therapy)

Stratified meta-analysis according to DMARD history resulted in a reduction in the overall between-study heterogeneity, and there was a statistically significant interaction between the groupings. The overall findings showed that treatment history and baseline disease duration modified the added effect of the targeted therapies. The added benefit of targeted therapies was lower in studies that included patients who were DMARD-naïve. The authors concluded that researchers conducting meta-analyses assessing the efficacy of targeted therapies for rheumatoid arthritis should include the patients' treatment history as a covariate in order to improve the precision of estimates. They also recommended that future trials should report adequately on the number of previously failed conventional DMARDs and biological agents for each patient, as well as the trial definition of treatment failure for each class of drugs. However, they also noted that their findings may be influenced by ecological fallacy, and therefore recommend that the findings should be confirmed using individual patient data.

The study by Christensen *et al.* also demonstrates the practical limitations of using meta-regression. Sixty-two trials were included in the analysis, but only 13 reported data on DMARD history and disease duration (Table 5.4).¹⁹⁸

The study presented by Mandema, *et al.*, demonstrates a *model-based meta-analysis*, which could potentially be further developed to allow for treatment sequencing.²⁰⁶ Model-based meta-analysis techniques allow for controlling and measuring variability in treatment response that comes from differences in dose, time under treatment, and baseline characteristics. Mandema, *et al.*, used a dose-response meta-regression analysis to enable the inclusion of data from varying dosages and outcome measures. This represented an alternative to previous methods such as lumping different doses together, ignoring differences in efficacy, or using information from only one particular dose.²⁰⁶ The main analyses combined data from studies that differed according to patients' population characteristics, line of therapy, and treatment histories. The analyses included 11 treatment arms comprising placebo, methotrexate, five TNF-inhibitors, and four 'newer' biologics using a different mode of action. The data were pooled across different patient populations, including those who were methotrexate-naïve, those with an inadequate response to methotrexate, and those who had an inadequate response to previous TNF-inhibitors. The potential heterogeneity in treatment effects was addressed by including placebo response, study design, and patient characteristics as parameters in the dose-response analyses, with a random-effects model used to account for the remaining unexplained between study heterogeneity. Study specific model parameters included background treatment and failed prior treatments for the study population. These were grouped as conventional DMARD other than methotrexate, conventional DMARD including methotrexate, and TNF-inhibitors. Treatment-arm specific covariates relating to patient characteristics included, among others, disease duration. The authors reported that there was no significant impact of differences in background treatment, disease duration, or other covariates included in the regression analysis, but did not report

the actual data for these findings. Regression analyses were also used to characterise the interaction coefficient for TNF-inhibitor and methotrexate initial combination therapy, which was -0.32 (-0.37 to -0.27), showing that the effect of initial combination was less than the sum of the effects of each component. The study also included stratified meta-analyses according to whether the trial population included patients with an inadequate response to prior methotrexate, or were methotrexate naïve. The analyses of TNF-inhibitors in the methotrexate naïve population showed that combination therapy with methotrexate provided a significant increase in the response rate as compared with methotrexate monotherapy.

The review conducted by Anderson *et al.*, demonstrates how individual patient-level data were used to inform the regression analysis.²¹ Individual patient data meta-analyses have much higher power if patient-level covariates are of interest.^{309 310} The use of individual patient data can also address to some extent the potential problem of ecological fallacy (referred to in Section 5.3.3). In terms of treatment-sequencing effects, ecological fallacy would be to assume that every patient within a subgroup has received the same previous treatments. The use of individual patient data, i.e the actual 'raw' data for all the patients included in the primary study, would overcome this limitation, but is unlikely to be available for all relevant studies. The Anderson review aimed to identify which patient or disease activity factors predict response to second-line treatment with drugs or devices for rheumatoid arthritis.²¹ All but one included RCT investigated the use of conventional DMARDs (mainly methotrexate). The review did not consider biological agents. Only RCTs for which individual patient data were available on baseline and outcome variables were included. Despite this, the authors used a crude estimate of previous conventional DMARDs used, which was classified as either yes or no. The authors were able to undertake multivariate logistic regression analysis, which showed that prior DMARD use was associated with a lower rate of treatment response. This effect was independent of disease duration or other factors. However, the study did not consider response to individual treatments, and the findings relate any active treatment compared to placebo. Treatment was controlled for in the analyses with an indicator variable for each non-placebo treatment, whilst the 'study' was not included as an effect because some active treatments were studied in only one RCT.

5.5.4 Meta-regression used for evaluating the influence of subsequent treatments on overall survival

A non-systematic review, by Grothey *et al.*, investigated the impact of subsequent active salvage treatments on median overall survival for studies of first-line treatment, using regression analysis. The authors noted that the median overall survival from RCTs of chemotherapy regimens used as first-line treatment for advanced colorectal cancer showed substantial variation.²⁰² The large number of participants enrolled, and the short time frame over which they were conducted meant that this heterogeneity was unlikely to be due to patient selection or factors unrelated to the actual treatment strategies used.²⁰² It was considered that this could be due to the impact of subsequent salvage treatments, but treatment sequences were only investigated in one RCT, the GERCOR trial, which was included in some of the reviews discussed in Section 5.3.2. Grothey *et al.*, therefore investigated

the influence on the reported median overall survival of the availability of all three active cytotoxic agents, fluorouracil-leucovorin, irinotecan, and oxaliplatin, in the course of the treatment. They included seven recently-published phase III RCTs in advanced colorectal cancer, which reported the number of patients receiving second-line therapies.²⁰² Only treatment arms containing an irinotecan- or oxaliplatin-based combination therapy as first-line treatment were included, from which they recorded the percentage of patients who received any subsequent treatment and the percentage of patients who were treated with all three cytotoxic agents during the course of their disease. The order in which these drugs were used was not specified, and patients could have been exposed to any agent as either first or second-line treatment. The analysis did not distinguish between whether the missing active drug was used as second or third-line treatment. Data on patient median overall survival from each trial was also recorded. The goal of the analysis was to correlate both the percentage of patients receiving second-line therapy and the percentage of patients receiving all three agents with the reported median overall survival. For the analysis, the Spearman rank correlation test was supplemented by simple linear regression. As a sensitivity analysis, weighted linear regression was used with weights proportional to the trial's sample size. The reported median overall survival was significantly correlated with the percentage of patients receiving all three drugs in the course of their disease ($p=0.0008$), but not the percentage of patients who received any second-line therapy. Studies in which a high percentage of patients had access to all three active cytotoxic drugs in the course of their disease showed the longest overall survival.

This represented a simple analysis that assumed homogenous trial populations. The authors recognised this as a potential bias in their analysis. Specifically, that patients who live longer have a greater opportunity to receive all three drugs, and that patients with poorer performance status, and, thus, shorter life expectancy might have been excluded from irinotecan- or oxaliplatin-based second-line treatments. A high percentage of patients in the included studies, ranging from 52 to 81%, lived long enough to have received second-line treatment of any kind. However, at the time the studies were conducted, not all drugs were available to all patients enrolled in the studies. A similar issue was also identified in Section 5.3.2 relating to channelling bias and illustrated using the study conducted by Stenner *et al.* The authors concluded that the findings showed that any second-line treatment does not impact survival but using all three drugs during the course of the disease does, suggesting that specific second-line treatment impacts survival.

The analysis did not address the question of which is the best sequence of treatment options, and whether an irinotecan or oxaliplatin based regimen should be used first. The findings however, did show that there is a need to take subsequent treatments into account when analysing overall survival for first-line treatments in advanced colorectal cancer. This is also likely to have an important impact on economic evaluations which often use data on overall survival. I return to this as part of the review of modelling studies in Chapter 7 (Sections 7.3.2, 7.3.3, and 7.3.4), it is discussed further in the Appendix Volume I, Section C4.2, and the issue of switching treatments within an RCT of individual treatments in Section A4. The findings also provide useful information for decision-making and

potential future research. The authors concluded that for maximum survival benefit, patients who received, for example, first-line oxaliplatin-based monotherapy should receive irinotecan-based combination therapy, where the effective salvage therapy might compensate for the less active first-line therapy.

5.5.5 Lessons about studying treatment sequences that have emerged from studies using meta-regression

Meta-regression has not been used extensively in the evaluation of treatment sequences. However, it has been used to explore whether the number of previous treatments or disease duration (which was discussed in Chapter 2 as a potential surrogate for previous treatments) has an impact on treatment effect. It has also been used to investigate the impact of subsequent treatments on overall survival. Disease duration and the number of previous treatments were identified as potentially important effect modifiers, but they are also correlated, and the effect of either one may have been confounded by the other. This may provide a justification for including both covariates in future meta-regression to obtain the effect of each one controlling for the other, e.g. the number of previous treatments adjusted for disease duration. There is also a need to ascertain whether long-standing disease *per se* is predictive of treatment response, or whether prior failure on previous treatments predicts response to subsequent treatment. This was also identified as an important issue in Chapter 2, and the health technology assessment of sciatica treatments. However, the main challenges of using this approach are the poor reporting of previous treatment, the susceptibility to type II error due to a small number of studies, and the potential for ecological fallacy. The combined use of RCT data and observational studies, and the availability of individual patient-level data may help to overcome some of these limitations, as would methods that incorporate different outcome measures.

Table 5.4: Summary of included rheumatoid arthritis studies using regression methods

Study	Decision point – treatments of interest	Patient population - treatment history	Outcome measures	Covariates related to sequencing included in analysis [no of studies if 10 or below]	Available evidence base*	Results
Lloyd, 2011	1 st -line or 2 nd /3 rd -line TNF inhibitors for RA (<i>Sequential TNF-inhibitors</i>)	TNF-naive vs TNF-IR	Achieving ACR50 and EULAR response Mean improvement in DAS-28 and HAQ scores	Meta-regression (included 3 covariates): <ul style="list-style-type: none"> • Mean no. of previous cDMARDs [n=10] • Disease duration • Duration of TNF Subgroup analysis: <ul style="list-style-type: none"> • Type of TNF switched from and to (ADA, ETA, INF) • Reason for switching (intolerance, primary inefficacy; secondary inefficacy; either inefficacy [most studies did not differentiate between primary and secondary]) 	Observational studies (uncontrolled and sequencing studies)	Meta-regression: Disease duration was the only covariate that was found to have a significant effect. There was an extra reduction in the DAS-28 score, representing a worsening of the condition, of -0.16 (95% confidence interval: -0.24 to -0.07) per additional year. Subgroup analysis: No significant difference identified (<i>no results presented</i>)
Nixon, 2007	1 st -line biological agents for RA (in early and late stage disease)	DMARD-naïve, or cDMARD-IR	Achieving ACR50 response	Meta-regression (included 2 covariates): <ul style="list-style-type: none"> • Disease duration 	Placebo or MTX controlled RCTs	For every additional year of disease the expected OR of an ACR50 event with biologics was 1.13 times longer. <i>Absolute effectiveness of biologics did not improve with longer mean disease duration.</i>
Schmitz, 2012	1 st -line TNF-inhibitors for RA	cDMARD[MTX]-IR	Achieving ACR20, ACR50, and ACR70 response Mean improvement in HAQ score	Meta-regression (included 3 covariates): <ul style="list-style-type: none"> • No. of previous cDMARDs [n=9] (<i>no further details</i>) • Disease duration 	Placebo RCTs	Results were not statistically significant, therefore covariates not included in final NMA.
Kanters, 2014	1 st -line biologics for RA (in late stage disease)	cDMARD[MTX]-IR (excluded patients who were MTX-naïve, or had IR to biologics)	Achieving ACR20 and ACR50 response	Meta-regression (10 covariates analysed using univariate analysis): <ul style="list-style-type: none"> • Disease duration <i>Precluded due to lack of studies:</i> <ul style="list-style-type: none"> • No. of previous cDMARDs [n=8] 	Placebo RCTs	The relative increase in the OR of an ACR50 response was 1.6 (95% CI: 1.06 to 1.27) per 1-year increment Disease duration was not associated with magnitude of ACR20 response.
Christensen, 2015	targeted therapies (biological agents and targeted synthetic DMARDs) used as add-on therapy for RA	DMARD-naïve; cDMARD-IR; or targeted therapy-IR	Achieving ACR20 (primary outcome); DAS28-remission	Meta-regression (9 covariates relating to trial eligibility criteria): <ul style="list-style-type: none"> • DMARD history [n=62; 49 in 'not reported' group] • Maximum disease duration at inclusion (early/established (>2yrs)/not reported) [n=62; 49 in 'not reported' group] 	Placebo or MTX controlled RCTs	The added benefit of targeted therapies was lower in trials including "DMARD-naïve" patients compared with trials including "cDMARD-IR" (ratio of ORs for ACR20 response = 0.45, 95% CI 0.31 to 0.66) and trials including "targeted therapy-IR" (0.50, 95% CI 0.29 to 0.87). Longer mean disease duration was associated with a higher likelihood of responding to treatment ($\beta = 1.05$, 95% CI 1.00 to 1.11 OR's per year; $p = 0.03$). Analyses conducted using DAS28-remission as the outcome supported the above-mentioned findings.
Anderson, 2000*	2 nd -line conventional DMARDs for RA	cDMARD-IR or cDMARD-naïve	Achieving ACR20 response	Regression analysis (included 8 covariates): <ul style="list-style-type: none"> • Previous cDMARD use (yes/no) • Disease duration (divided into 5 categories) Analysis based on individual patient-level data**; only factors that were significant in univariate analysis included).	RCTs of MTX (vs placebo or alternative cDMARD)	Prior cDMARD use was associated with a lower rate of response (Adj OR for yes = 0.62, $p=0.008$). Disease duration had a strong effect on likelihood of patient response (Adj OR per year = 0.98, $p=0.028$).
Mandema, 2011	1 st -line TNF-inhibitors for RA	cDMARD[MTX]-naïve, cDMARD[MTX]-IR, and/or TNF-IR	Achieving ACR20, ACR50 and ACR70 response	Meta-regression (included 7 covariates): <ul style="list-style-type: none"> • Failed prior treatment (for study population) • Disease duration (for each treatment-arm) 	Placebo or MTX controlled RCTs	Differences in failed prior treatments and disease duration did not have a significant impact on treatment effect (<i>actual results not presented</i>).

				<i>Stratified MA also conducted for patient population with IR to previous MTX and MTX-naïve.</i>		
Grothey, 2004	1 st , 2 nd -, and 3 rd -line chemotherapy treatment (CTX) for advanced colorectal cancer The impact of treatments subsequent to 1 st -line for advanced colorectal cancer	IR to previous CTX. Data from RCTs of 1 st -irinotecan or oxaliplatin based CTX.	Median overall survival	<i>Weighted linear regression (with weights proportional to the trial's sample size) based on arm level proportions.</i> Regression analysis: <ul style="list-style-type: none"> • %age of patients receiving all three active agents (FU-LV, irinotecan, or oxaliplatin) during the course of their disease • %age of patients receiving any (i.e. not necessarily irinotecan- or oxaliplatin-based) second-line treatment 	10 treatment arms from 7 RCTs of 1 st -line CTX	The median OS was significantly correlated with the %age of patients who received all three drugs in the course of their disease (p=0.0008; Spearman rank correlation test) but not with the %age of patients who received any 2 nd -line therapy, therapy (p=0.19; Spearman rank correlation test)

Abbreviations: ADA adalimumab; Adj adjusted; ARC American College of Rheumatology score; cDMARD conventional disease modifying antirheumatic drug; CI confidence interval; CTX chemotherapy; DAS disease activity score; ETA etanercept; EULAR European League Against Rheumatism classification; HAQ Health assessment questionnaire score; IR inadequate response; INF infliximab; MA meta-analysis; MTX methotrexate; NMA Network meta-analysis; OR odds ratio; OS overall survival; RA rheumatoid arthritis; RCT randomised controlled trial; TNFs tumour necrosis factor-inhibitors; %age percentage.

5.6 USING A MODIFYING FACTOR TO ADJUST THE EFFECT OF SINGLE TREATMENTS

5.6.1 Number of studies

Two studies developed a reduction, or multiplication, factor which, could then be applied to the treatment effects obtained from RCTs of first-line use to represent their use later in the treatment pathway. Both studies used this approach when faced with a limited evidence base to develop sequencing effect estimates to inform the parameters of their economic model. The studies were undertaken as part of the NICE technology appraisal process, and were introduced in Chapter 4 (Section 4.4).^{17 199} Treatment sequencing was not considered within the clinical evaluation of either technology appraisal. One study included an appraisal of three TNF-inhibitors for psoriatic arthritis (NICE TA199),¹⁷ and the second evaluated the use of newer anti-epileptic drugs for the treatment of epilepsy in children (NICE TA79).¹⁹⁹ Both studies were included in the review of modelling studies, the findings of which are presented in the next two chapters.

Seven further studies included in the review of modelling studies also applied a reduction factor to individual treatment effects to represent their use at a later point in the treatment pathway, but did not report the methods used for developing the reduction factor.^{18 229 234 244 261 264 272} These studies reported using estimates based on the available evidence, the choice of which was frequently not justified and it was unclear how the reduction or multiplication factor was actually developed. They are described in more detail in the review of modelling studies presented in Chapters 6-7, but the approach used to develop and implement the reduction factors are summarised briefly here. The application of the assumption that treatment effect is reduced when used at a later point in the pathway, in order to implement the decision analytic model is reviewed in more detail in Chapter 6. Most of these studies evaluated treatment sequences for inflammatory arthritis,^{229 234 261 264 272} whilst one study included antiviral therapy for HIV,²⁴⁴ and another antiepileptic drugs (NICE CG137).¹⁸

5.6.2 Methods used to develop a reduction factor

5.6.2.1 Based on the comparison of the same drug class used as first-line and subsequent treatment from observational studies

The economic impact of a second-line TNF-inhibitor was explored in sensitivity analysis, as part of the technology appraisal presented by Rodgers *et al.* (NICE TA199)¹⁷ As there were no RCT data available beyond first-line use, the treatment effect of the second TNF-inhibitor was obtained by reducing or increasing the treatment effect of the first-line inhibitor by a set amount, depending on whether the initial biologic was discontinued due to inefficacy or adverse effects. The treatment effects of TNF-inhibitors were based on placebo controlled trials of single treatments, and the multiplication factors on an observational study of data from the British Society of Rheumatology Register (BSRBR) of biological agents used for rheumatoid arthritis (introduced in the Appendix Volume I, Section C3). The economic model considered two subgroups: patients who discontinued their first biological agent due to adverse effects, and patients who discontinued due to lack of efficacy. No distinction was made between those who had an initial lack of response and those who had secondary loss of treatment efficacy. A supplementary literature review of patient registry studies

was conducted to find the response or withdrawal rates from TNF-inhibitors for patients with psoriatic arthritis or rheumatoid arthritis. However, only one study that reported the appropriate data was identified, which included patients with rheumatoid arthritis. The study distinguished between outcomes for the patients who started a second TNF-inhibitor after adverse events from the first, and patients who started a second course of TNF-inhibitors following lack of efficacy of the first TNF-inhibitor. The hazard ratio for withdrawal from second-line treatment due to inefficacy, compared with the discontinuation rate for first-line treatment due to inefficacy was 2.7 (95% confidence interval: 2.1 to 3.4). The findings were not statistically significant for those who discontinued their first-line treatment due to adverse effects. The hazard ratio for withdrawal from the second biologic due to adverse events, compared to withdrawing from the first course due to adverse effects was 2.3 (95% confidence interval: 1.9 to 2.9). The findings were not statistically significant for those who discontinued their first-line treatment due to inefficacy. For the economic model, the treatment effect of the second TNF-inhibitor was assumed to be equal to that of the first TNF-inhibitor multiplied by the relative risk of failing second-line treatment compared to first-line treatment. The hazard ratios for failing the second TNF-inhibitor compared to the first was assumed the same for all three biological agents, due to the lack of data. For initial treatment response, it was assumed that if the first TNF-inhibitor was discontinued due to inefficacy, the odds of achieving an initial response with the second biologic would be *reduced* by a factor of 2.7, and unchanged if it was discontinued due to an adverse event. For treatment withdrawal after the first three months, it was assumed that if the first biologic agent was discontinued due to inefficacy, the risk of withdrawal due to inefficacy *increased* 2.7-fold, and unchanged if it was discontinued due to an adverse event. It was also assumed that if the first biologic was discontinued due to adverse events, the risk of withdrawal of the second biologic due to adverse events would be increased by an average of 2.3, and the odds of withdrawal due to inefficacy unchanged.

5.6.2.2 Based on the comparison of the same drug used as first-line and last resort from RCTs

The economic evaluation of the technology appraisal presented by Connock *et al.* (NICE TA79)¹⁹⁹ included the comparison of pre-defined sequences of up to four treatment lines, that contained either only older antiepileptic drugs, or a combination of both 'older' and 'newer' antiepileptic drugs. The newer antiepileptic drugs were generally licensed, and evaluated in placebo-controlled trials as add-on therapies for patients with more or less refractory epilepsy, but one drug was available as both initial monotherapy and add-on therapy. For the economic model, the proportion of patients achieving seizure freedom with each line of therapy were reduced by a factor of 0.4. This was based on the data available for the one new antiepileptic drug for which there was an RCT of its use at two different time points, first-line monotherapy and later as an add-on therapy. The RCT data of antiepileptic drugs used as add-on therapy were assumed to be representative of fourth-line therapy, and were also used as 'anchor points'. The proportion discontinuing due to adverse effects or lack of efficacy was kept constant, based on the same trial data. No sensitivity analyses were conducted to assess the impact of the reduction factor. A de novo economic evaluation of drug sequences for epilepsy was also conducted for NICE clinical guideline number 137.¹⁸ It included the comparison of antiepileptic drugs

used for both newly diagnosed, and refractory epilepsy, analysed separately. The model included fixed sequences of three treatments: first-line monotherapy, second-line monotherapy, and third-line adjuvant therapy. Patients failing first-line due to inadequate seizure control were assumed to be 75% (risk ratio 0.25) less likely to achieve remission with second-line monotherapy. This was informed by an observational study, and the figure was varied in one-way sensitivity analyses. For patients who failed first-line treatment due to intolerable side effects it was assumed that response to the second-line monotherapy was independent of response to first-line antiepileptic drug. The observational study showed that the probability of achieving seizure freedom with a second monotherapy, after failure of the first due to lack of efficacy, was much lower than that of treatment naïve patients (11% vs 47%; risk ratio 0.24), whilst the probability of response after failure due to intolerable adverse effects was similar (45% vs 47%).³¹¹ The probabilities of treatment failure for each antiepileptic drug were derived from a published network meta-analysis of monotherapies, based on RCTs included in eight Cochrane systematic reviews and the SAND I trial.³¹² Trial participants included children or adults with a new diagnosis of epilepsy, or relapsed following antiepileptic withdrawal, or who had failed on other therapies. The network meta-analyses were performed using individual patient data from each trial, which allowed the standardisation of outcome definition and enabled the assessment of the outcome for patients with either partial or generalized onset seizures. The analyses were conducted using a Cox proportional hazards model stratified by trial, and adjusted for treatment and the known prognostic factors, epilepsy type and number of seizures before randomisation. They did not, however, consider treatment sequencing.

5.6.2.3 Based on selected published data sources

Five rheumatology studies applied different reduction factors derived from various sources of evidence. One modelling study (Diamantpoulus, 2012), due to the lack of evidence about the efficacy of a second TNF-inhibitor in rheumatoid arthritis, reduced the treatment effect of adalimumab, used as a TNF-inhibitor in the sequence, by 30%.²³⁴ They cited NICE Technology appraisal 130 (Chen, 2006)²²⁷ as the source, but provided no further explanation. Another study (Clark, 2004; NICE TA72) that also included treatment sequences containing sequential TNF-inhibitors, explored the potential sequencing effect of a second TNF-inhibitor as part of their sensitivity analysis.²²⁹ A HAQ score of 0.5 was used for both TNF-inhibitors in the base-case analysis, and in the sensitivity analysis this was changed to 0.625 for the first TNF-inhibitor, etanercept, and 0.25 for the second, infliximab. The 0.625 estimate was based on a previous published model by the same research group (Jobanputra, 2002; TA36).²⁴⁵ Another modelling study (Russell, 2009), which also included the use of a third TNF-inhibitor, used a 10% reduction in effectiveness after each switch.²⁶¹ The assumption that switching TNF-inhibitors is associated with lower efficacy was reported to have been based on clinical experts and published observational studies of patient registries, but no further details were given. One modelling study (Tran-Duy, 2011), which evaluated treatments strategies incorporating a second TNF-inhibitor for ankylosing spondylitis, used a reduction factor of 0.65, applied to the mean decrease from baseline in the disease activity score.²⁷² The reduction factor was based on data obtained from a registry study, but it was not stated how it was developed. One further modelling study (Schadlich,

2005) of treatment sequences based on conventional DMARDs for rheumatoid arthritis reduced the estimates of clinical effectiveness by 25% for four DMARDs, when they were used as second or subsequent-line treatments.²⁶⁴ This was based on an observational study that showed there was a 50% reduction in clinical effectiveness between first-line and second-line DMARD, but the difference between second and subsequent lines was not statistically significant. I re-visit the use of a reduction factor to RCTs of first-line TNF-inhibitors in Chapter 6, Section 6.6.2.2.

One modelling study, which investigated the cost-effectiveness of antiretroviral therapy strategies for single-dose nevirapine exposed HIV women, used a multiplication factor to reduce the efficacy data used to inform the first-line treatment in order to represent its use as a second-line treatment.²⁴⁴ The study included two treatment sequences, starting with either a nevirapine based treatment strategy followed by a lopinavir-ritonavir based strategy, or vice versa. Both strategies were triple therapies, which included two nucleoside reverse transcriptase inhibitors (NRTIs) that differed between regimens but their efficacy were assumed to be equal. Treatment effects for the nevirapine based strategy was taken from an uncontrolled follow-up study (of a previous RCT of intrapartum nevirapine vs placebo), and the treatment effects for the lopinavir-ritonavir was derived from an RCT of initial treatment for HIV. The efficacy of the treatments used as second-line were estimated to be 90% of their efficacy as an initial therapy, because of nucleoside reverse transcriptase inhibitor resistance resulting from first-line treatment failure. This multiplication factor was reported to have been based on the findings of a previous economic model developed by the authors to evaluate the cost-effectiveness of genotype resistance testing in treatment naïve patients. However, it was not stated how the multiplication factor was developed.

5.6.3 Lessons about studying treatment sequences that have emerged from studies using a modifying factor

Treatment sequences are generally represented within a decision model as a series of individual treatments, each requiring a summary treatment effect conditional on their positioning in the treatment pathway. The use of data from studies of single treatments to parametrise such models necessitates adjusting the individual effect estimates conditional on positioning in the sequence. This approach is one way of doing this. However, more research is needed to identify the best method of estimating and testing the modifying factors. The included studies showed that treatment switching due to lack or insufficient effect appeared to have a greater impact on the efficacy of subsequent treatments than discontinuing treatment due to intolerance or adverse effects.

The evidence available for developing the modifying factors is likely to be limited. In most cases, they were informed by a single observational study and, as such, the data on which they are based are likely to have been affected by some of the limitations listed in Sections 5.4 and those discussed in the Appendix Volume I (Section C). One study (Connock, 2006) used data from RCTs evaluating the same treatment used as first-line and fourth-line, but the data to inform the modifying factor would have been observational in nature. One of the potential limitations of using RCT data, as opposed to

patient registry data, highlighted by this study is that it may not be possible to differentiate between the impact of discontinuing previous treatment due to lack of efficacy or adverse effects. Furthermore, the RCT of fourth-line treatment in this study did not just account for the immediate prior treatment, and as such this type of data may not have been so useful here or easy to interpret. The included studies also assumed that the effect of various treatments from the same class (e.g. TNF-inhibitors, new antiepileptic drugs) are the same.

5.7 RANKING TREATMENTS ACCORDING TO THEIR ABSOLUTE EFFECT ESTIMATES

5.7.1 Number of studies and description of approach

This approach represents a fairly simplistic method, based on identifying the optimum sequence by ranking individual treatments according to their effectiveness, which is generally based on the evidence of their use at a single point in the treatment pathway, or worse, data pooled across different treatment lines. In essence, this method ignores any potential sequencing effects, and is based on the assumption of treatment independence. An example of this approach is provided by Finnerup *et al.*, who ranked treatments according to their absolute effect estimates in order to develop the optimum evidence-based algorithm for the treatment of neuropathic pain.²⁰¹

Finnerup *et al.*, noted that the choice of treatment for neuropathic pain was generally based on the following six criteria, which would need to be considered when developing an evidence-based treatment algorithm: consistent outcomes from RCTs; high degree of pain relief, superior to existing treatment; persistent pain relief; limited side effects; effect on quality of life; and low cost. They limited inclusion to double-blind placebo-controlled randomised trials of single treatments, and identified 105 relevant studies, 59 of which were small crossover studies. The data were pooled for each treatment using numbers needed to treat and numbers needed to harm, which are the reciprocal of the absolute risk difference. Numbers needed to harm were based on dropouts due to adverse effects. Treatment algorithms were then obtained by first ranking studies according to the numbers needed to treat and then taking into account potential side effects. The authors listed some of the limitations to their study, which included having to dichotomise the outcome data, and the methodological complexity of pooling data from both small crossover and large parallel group trials. The findings from this review (Finnerup, 2005)²⁰¹ were subsequently used as the source of clinical effectiveness data for an economic model of sequential medication strategies for postherpetic neuralgia reported by Smith *et al.*, which was based on a Markov Cohort modelling approach making the simplifying assumption of treatment independence.²⁶⁸

An important cause of treatment non-independence is the progressive change in responsiveness of the disease to therapies as the disease progresses. However, this deviation may not be relevant for treatment sequences for neuropathic pain, as the objective of the treatment is to reduce the symptom of pain, rather than alter the underlying cause; when the treatment is stopped the pain is expected to return. Another potential deviation from treatment independence is that patients who are responsive

to one treatment are likely to respond to others, and likewise those that are non-responsive to one treatment are less likely to respond to other treatments.²¹⁶ The included studies did not report previous treatments, but it was noted that some trials of gabapentin and pregabalin excluded patients who failed to respond to previous treatments with gabapentin. In other words, patients who failed one treatment were likely to be less responsive to the next treatment. A third potential deviation occurs if there is a subset of patients that are resistant to all treatments.

5.7.2 Lessons about studying treatment sequences that have emerged from studies that rank treatments according to absolute effects

This represents a naïve approach for developing evidence-based treatment algorithms, and is only suitable if treatment effectiveness is not dependent on positioning in the sequence or disease duration. Providing evidence to show that this is in fact the case may not be straightforward.

5.8 A CASE STUDY FROM THE GetReal PROJECT: USING REAL-WORLD DATA TO LINK FIRST AND SECOND-LINE TREATMENTS INCLUDED IN SAME NETWORK META-ANALYSIS

The GetReal project (<http://www.imi-getreal.eu/>) represents a recent programme of work about incorporating real-life data into drug development. The three year project was funded by the Innovative Medicines Initiative (IMI), an EU public/private consortium consisting of pharmaceutical companies, academia, Health Technology Assessment agencies and regulators (e.g., NICE, EMA and ZIN), patient organisations and Small and Medium Enterprises (SMEs). The project was launched in 2013 and was only identified when the review of methods was completed.

Work package one of the GetReal project included two case studies in different disease areas, which were conducted to explore how real-world data, from patient registries, can be used to help demonstrate the relative effectiveness of new medicines. One of these case studies, in rheumatoid arthritis, is directly relevant to the review of methods as it also considered the evaluation of treatment sequences. This represents a meta-analytic approach not covered by the included studies, where individual treatment lines, in this case first and second-line, were included in the same network meta-analysis as separate treatment nodes and, in order to achieve this, real-world data was used to link first and second-line use of the same treatments. The methods used can be viewed, in some respect, as building on those presented by Nixon *et al.*²⁰⁸ and included in Section 5.5.3 under meta-regression. The case study also contributes to developing methods to address having an incomplete network, which was identified in Section 5.3.2, relating to the inclusion of predefined sequences.

The relevant case study evaluated methods to incorporate real-world data in the evidence synthesis of second-line biologics in rheumatoid arthritis. Data from two national patient registries were used to supplement randomised evidence to address two key issues:

- i. How to connect disconnected networks of evidence in order to conduct network meta-analyses?

- ii. How to optimise an evidence base using first-line evidence to inform second-line effectiveness estimates?

The two national patient registries included the Swiss Clinical Quality Management in Rheumatic Diseases (SCQM) database, and British Society for Rheumatology Biologics Register (BSRBR) database, both of which were introduced in the Appendix Volume I (Section C3.3). The authors had access to individual patient data (IPD) from the two registries and for five RCTs, two of which investigated second-line treatment.

The base case analysis included two Bayesian network meta-analyses, which were performed using only the aggregate results from RCTs for Disease Activity Score 28 (DAS-28) remission at six months for first-line and second-line biologics, respectively. Several additional analyses were then undertaken using the registry data extracted from the two registries, adding RCT individual patient data, and combining the first and second-line networks. The results of the base case network meta-analysis for second-line biologics was used as a reference against which all other analyses and results were compared. Analyses 1-4 related to developing the best method to connect disconnected networks, whilst analyses 5-6 were regarding the use of first-line data to inform the second-line treatment, and are therefore summarised below.

Analysis 5 was a univariate network meta-analysis using data from the two registries to bridge the evidence gaps between the first and second-line networks. 'Real-world evidence' from the BSRBR register was used to incorporate relative effect estimates for first-line versus second-line treatments, which allowed the two networks of evidence to be connected and for the treatment comparisons, for example, for drug A in first-line versus drug A in second-line to be obtained. The bridging was performed by using the BSRBR as one large study, but the registry data could have also been split into multiple pairwise comparisons (i.e. studies) to be used as small multiple bridges.

Analysis 6 represents an alternative approach proposed for utilising the real-world data, which was to use multivariate analysis to model separate outcomes simultaneously, using the correlation to borrow information across multiple outcomes or time points. Here, the treatment effect in first-line was modelled as outcome one and the treatment effects in second-line as the second outcome. The correlation referred to the between-treatment (lines of therapy) correlation, and not between or within study.

Analysis 6a was based on a bivariate network meta-analysis, where the two treatment lines were assumed to be correlated, and a correlation estimate was used as a prior distribution. The data from both registries were used to obtain a correlation estimate between treatment effects, in the first and second-line of treatment, by splitting the registry data into first and second-line response and a pairwise meta-analysis performed whilst monitoring the correlation. The network of evidence was based on data from RCTs of first-line and RCTs of second-line treatment.

In Analysis 6b the data from both registries were used as part of the network by splitting into multiple pairwise studies), providing treatment effect estimates in the first and second-line of therapy. This allowed for modelling between-studies correlation between the lines of therapy. The network of evidence was based on registry data for first and second-line treatment, rather than the RCTs.

Analysis 6c used data from the registries, reporting treatment effect estimates on both lines, which allow the assumption of exchangeability on the average level to be relaxed.

The univariate analysis included the registry data as data, whereas the bivariate network meta-analysis used the registry data to inform the prior distribution for the correlation parameter between first and second-line of therapies. The bivariate network meta-analysis did not directly compare first versus second-line treatment but provided estimates for all treatments in first and second-line, by using the correlation between them to predict estimates in second-line (or first-line) where these estimates did not exist previously. However, the authors noted that the results from both the univariate and bivariate network meta-analyses may be more difficult to interpret in a decision making context for a new treatment in a specific line of therapy.

The interpretation of the overall findings was limited by a large amount of uncertainty in the results for analysis 5 and 6, as reflected in very wide credible intervals. There was also large variation in the point estimates, in some instances, suggesting inconsistencies within the evidence base and across networks. This may have been influenced by the choice of outcome, as only a small number of patients will achieve full remission with second-line treatment, leading to a low event rate. The methods proposed were quite complex, and obtaining the necessary individual patient data from the patient registries was not straightforward and very time consuming. There was a lot of programming involved in obtaining the necessary data, and missing information restricted the datasets from both registries.

The wider GetReal project included the production of a series of literature reviews of methods covering network meta-analysis,⁵⁹ combining randomised and non-randomised evidence in network meta-analysis,³¹³ meta-analysis of individual participant data,¹⁷⁵ and mathematical modelling.³¹⁴ These have been incorporated in the framework, which is presented in Chapter 8.

5.9 DISCUSSION

5.9.1 Summary of the findings for the review of meta-analytic techniques

The review of meta-analytic techniques used in the context of treatment sequencing identified that there are two key approaches for synthesising the data on treatment sequences: one where the 'treatments' included in the meta-analysis represent whole sequences; and one where individual treatments effects are conditional on their positioning in the sequence. When pooling data on individual treatments used early in the sequence, the use of subsequent treatments is likely to impact

long-term outcomes. When synthesising treatments used later in the treatment pathway, previous treatments are likely to represent an important effect modifier. An in-depth evaluation of between-study heterogeneity when pooling data from studies of treatments used at any point in the treatment pathway may provide useful information about the importance and the need to consider sequencing effects. In terms of the clinical evaluation within a health technology assessment, included studies generally approached the issue of sequencing by limiting the complexity of the problem, and adopting a well-defined or narrow review question, evaluating isolated points in the treatment pathway.

The review of methods for analysing the effectiveness of treatment sequences identified that, to date, most studies use standard quantitative evidence synthesis methods. This included restricting inclusion to comparative studies of complete sequences, and the use of regression analysis and investigation of subgroups to account for previous treatments used. In certain indications where treatment effectiveness is not impacted by previous treatments, it might be acceptable to select the order of individual treatments within a sequence based on the best absolute effects of individual treatments, but this is a naive approach and should be adopted cautiously and with reservations. No novel evidence synthesis or meta-analytic techniques were identified that were developed for evaluating treatment sequences, or developing summary treatment effect estimates conditional on the effect of previous treatments. This is likely to be due to the poor reporting of prior and subsequent treatments by primary studies, as it is difficult to gauge the extent of their impact on the current treatment effect or overall outcomes measure without such information.

5.9.2 Relying on sequencing trials

The review findings showed that relying solely on treatment sequencing studies, either clinical trials or observational studies, is unlikely to provide sufficient evidence to inform clinical decision making. The included studies that undertook meta-analysis of complete treatment sequences generally had to limit inclusion to two treatment lines for which prospective treatment sequencing studies were available. This was especially the case for studies of advanced cancer. Establishing a connected evidence network for complete sequences, as required for implementing network meta-analysis, can also be problematic. For example, in the NICE CG131 study, chemotherapy treatments for advanced colorectal cancer had to be grouped by mode of action in order to develop a closed network of three treatment sequences.²⁰⁷ The potential limitations of grouping interventions, such as class effect bias and aggregate data collection are discussed in Section 5.9.9. Even when pragmatic, gold standard adaptive treatment sequencing RCTs, such as the STAR*D trial in refractory depression,^{291 292} were available these did not always match the review question. However, as discussed in the Appendix Volume I, Section B2 it is acknowledged that the SMART design does not represent a confirmatory trial.

5.9.3 The use of subgroup and stratified analysis

A number of included studies explored the influence of previous treatments using subgroup analysis. Some stratified analysis according to the type of previous treatments used, whilst others examined the number of previous treatments used in different subgroups. Four studies used subgroup analysis to explore the implications of different reasons for switching treatments, but the poor reporting by primary studies precluded doing this in much depth. Presenting the findings relating to treatment sequencing as separate subgroups, especially when presented graphically, provided some useful information on potential trends. It is also easier and simpler to implement than meta-regression. However, on the whole, the use of subgroup analysis for evaluating treatment sequencing was fairly limited, primarily because it only provided a means of comparing two subgroups at a time, with or without one covariate. This means that for each subgroup the information over all other covariates are pooled, and each analysis is therefore confounded by other variables.³¹⁵ Subgroup analysis can also only test for the presence of an interaction and cannot estimate the extent of it. Only one study used appropriate statistical tests of interaction to compare subgroups, but there was frequently insufficient data available to conduct such analyses. When a series of subgroup analyses were implemented in order to examine a number of different variables, it was difficult to interpret the overall findings, especially when low power (type I error) produced insignificant findings. The evaluation of numerous variables, resulting in a number of statistical tests, can also lead to type II error or false positive results. The findings of the subgroup analyses were themselves also generally poorly reported by included studies, with most studies only noting that the results were not statistically significant, or merely noting that prior treatments did not influence treatment effects.

5.9.4 The use of meta-regression

Meta-regression analysis appears to be the most promising method of developing summary effect estimates that account for previous treatments, or sequence positioning. However, meta-regression was used predominantly for investigating whether the number of previous treatments was an important effect modifier, or to identify the source of statistical heterogeneity, rather than investigate the clinical effectiveness of treatment sequences as such. Meta-regression is generally used to evaluate the interaction between one or more study level covariates and treatment effect. Meta-regressions can be used to develop adjusted effect estimates according to previous treatments. None of the included studies used meta-regression to develop coefficients representing previous treatments. However, meta-regression was used to evaluate the impact of disease duration, which could be considered as a surrogate estimate for previous treatment, with extended disease indicating an increasing number of treatments tried. Disease duration was shown to be an important effect modifier for rheumatoid arthritis, and appeared to be better reported in primary studies than previous treatments. Some studies that pooled data across treatments used for both early and late stage disease used meta-regression to incorporate an interaction term for disease duration. The resultant regression coefficient provides a summary estimate of the amount by which the treatment effect changed on average per additional year of disease. Meta-regression was also used to characterise

the interaction coefficient for combination therapy, representing adding of a biological agent to methotrexate in rheumatoid arthritis.

The review showed that the implementation of meta-regression as a method of developing sequence-specific effect estimates, requires adequately reported data on the covariates from primary studies. However, the studies using meta-regression themselves also failed to report their results in full, with no data provided on covariates that were not statistically significant. For meta-regression to be useful in informing decision making or cost-effectiveness analyses of treatment sequencing, findings need to be reported in full, irrespective of statistical significance, especially when non-significance is likely to be due to lack of power.

5.9.5 The limitations and challenges of using meta-regression to inform treatment sequences

The relationship described in both meta-regression and subgroup analyses are observational in nature,²⁹⁶ as discussed in the Appendix, Volume I (Section A). Even when the reviews are restricted to RCTs, the study of effect-modifiers are inherently observational as it is not possible to randomise patients to one covariate value or another.³¹⁶ As such, they inherit all the difficulties of interpretation and inference that are attached to non-randomised studies, including confounding, correlation between covariates, and the inability to infer causality from association.^{296 316} Both meta-regression and subgroup analyses also suffer from aggregation or ecological bias, where inferences about the nature of individuals are deduced from inference for the group to which those individuals belong.^{296 317}

An important challenge of using these methods is the need to represent previous treatment as a single covariate. Included studies relied on a simplified, either dichotomised or categorised, summary of previous treatments, for example the number of previous treatments failed, or a pre-specified treatment graded as yes/no.

A practical limitation of using meta-regression is the availability of data from the primary studies, and having a sufficient number of studies and statistical power to detect a difference.²⁹⁶ Meta-regression requires the estimated treatment effect, its variance, and covariate values for each included study. Any analysis that is based on a subset of relevant studies, because these data are not available for all studies, will be flawed.²⁹⁶ If the number of relevant studies is also low then the observed impact of patient characteristics will be questionable.^{296 309}

The number of covariates which can feasibly be considered within a meta-regression is limited, especially if only a small number of studies are included in the review.^{208 296} Only those considered to be important covariates are therefore included, frequently chosen based on statistical significance in prior univariate analysis.^{20 205} The potential effect of previous treatment was often not included in the main analysis as it was not found to be a statistically significant covariate in preliminary analyses (Section 5.5.3).^{20 211} However, this assumes that there was sufficient power to detect statistical significance. Previous treatments were frequently poorly reported in primary studies, which probably

contributed to this non-selection for inclusion in the main analyses. For some clinical scenarios, including rheumatoid arthritis, there may be justification for including prior treatments as a clinically important covariate, irrespective of statistical significance.¹⁹⁸ The number of previous treatments is closely related to disease duration, which was generally included in meta-regression studies of rheumatoid arthritis, and may be acting here as a confounding factor. The inclusion of both covariates in the analysis means that it is possible to obtain the effect of each one controlling for the other.

The availability of individual level patient data would overcome a number of these limitations.^{175 296 316} The use of individual patient data is discussed further in Section 5.9.9.

5.9.6 The use of absolute treatment effects to identify optimal sequence

Another approach used was to develop absolute treatment effects for each individual treatment, and then use these to rank the treatments in order to identify the optimum sequence. This represents a naïve method that ignores the possible sequencing effects, the potential carry-over effect from patients who are resistant to any treatment, and the likely influence of changes over time or disease trajectory. This approach is therefore not recommended unless there is good evidence to show that clinical effectiveness is not affected by position in the treatment sequence.

5.9.7 The use of N-of-1 RCT

The N-of-1 trial is first described in Chapter 1 (Section 1.7) as a methodologically rigorous design (i.e. use of randomisation, blinding, formal outcome assessment) to determine the effect of the treatment in an individual patient.³¹⁸ It is also discussed in the Appendix Volume I (Section A4) as a trial design that reflects clinical practice.

No meta-analyses of N-of-1 RCTs for investigating treatment sequences were identified. This type of trial design is increasingly being recognised as a potentially useful data source for informing the use of complex interventions³¹⁹ and for personalised medicine.³²⁰ The last decade has seen a great increase in the analysis and meta-analysis of single-case-designs. This motivated a special issue of the *Journal of School Psychology* that included five articles providing an overview of current work on the topic, including standardised mean difference statistics, multilevel models, Bayesian statistics, and generalised additive models.³²¹ A group at the University of Alberta has also partnered with the *Journal of Clinical Epidemiology* to published a series of dedicated papers on the methodological issues related to the trial design and reporting of N-of-1 studies, as well as their synthesis.³²² This included an example of a systematic review and meta-analysis of N-of-1 trials, and a meta-analysis that combined data from both RCTs and N-of-1 trials.

The usefulness of N-of-1 trials for informing treatment sequencing, however, is yet to be recognised. Important limitations of this trial design for this purpose include their limited usefulness to disease conditions that are amenable to this type of study, and the use of a wash-out period between treatments. They are generally conducted in chronic and mostly stable conditions, where the optimal

treatment would provide a rapid effect that disappears quickly with treatment withdrawal,¹¹² for example the treatment of neuropathic pain. It is important that the disease or process persists long enough for the investigator to expose the patients to each experimental treatment and measure responses, which are generally based on short term outcomes.³²³ The limitations of this type of study, because they are designed to minimise any potential carry-over effects as much as possible, and attempts are also made to limit the period effect, are discussed in Chapter 1, Section 1.7. Three important limitations that have been identified with N-of-1 trials include incomplete reporting, marked variability in quality, and unacceptably high rates of prospective protocol registration.³¹⁸

5.9.8 The use of individual-patient data

The availability of RCTs reporting individual patient data, including full treatment histories, would greatly enhance the usefulness of meta-regression as a method of developing sequence-specific effect estimates. The use of individual patient data on the covariates provides a number of advantages, in that it provides an increase in power, improves the ability to investigate interaction and subgroup effects, overcomes issues regarding correlation, and reduces bias.¹⁷⁵ Meta-regression with individual patient data is capable of estimating effect modification with far greater precision, because of the much greater spread of covariate values.³¹⁶ It would also potentially enable the implementation of a separate interaction term for each treatment.¹⁷⁵ The use of individual patient characteristics, rather than summary characteristics of patients reported in the publications, will also prevent misleading inferences being made due to ecological fallacy.^{324 325} However limitations of using individual patient data include the challenge of obtaining access to the data, the fact that they require substantially more effort and statistical expertise to undertake, and are still susceptible to missing data on important covariates such as previous treatment.¹⁷⁵

The included meta-regression study by Anderson *et al.*, which aimed to identify which patient or disease activity factors predict response to second-line treatment for rheumatoid arthritis, limited inclusion to RCTs for which individual patient data were available on baseline and outcome variables.²¹ Despite having data on individual characteristics, the authors used a crude estimate of whether previous conventional DMARDs were used, classified as either yes or no, and the overall findings were based on the comparison of any 'active treatment' versus 'placebo'. Treatment was controlled for in the analyses using an indicator variable for each non-placebo treatment, whilst 'study' was not included as an effect because some active treatments were studied in only one RCT. The regression analysis showed that prior DMARD use was associated with a lower rate of treatment response, independently of disease duration and other factors. One further study (Rendas-Baum, 2011), which aimed to examine the relationship between clinical response to biological agents and the number of previous TNF-inhibitors received, used subgroup analysis based on individual patient data. The study included both RCTs and observational studies, but only data from the biological treatment arm of the RCTs were used in the analysis.

The use of meta-regression based on individual patient data for evaluating treatment sequences is dependent on the availability of sufficient individual level data on previous treatments. The likely scenario however, as discussed in Chapter 1, Section 1.4.1, is that even if IPD is available for some studies, it is unlikely to be so for all relevant studies.³²⁶ Individual patient data meta-analysis is prone to bias when individual patient data are only sought for a specific subset of studies.^{175 326} The evidence synthesis therefore needs to be based on both individual and aggregate level data. Any future work on developing methods for evaluating treatment sequences must address some of the challenges in pooling data based on individual patient and aggregate level data.^{67 175 309} Better reporting of previous treatment is also still needed in RCTs of single treatments for which IPD is available.

A further issue relating to IPD meta-analysis is whether or not to combine data from both RCTs and non-randomised studies. There is an ongoing debate in the literature regarding the validity of this approach, and under what conditions is it likely to be advocated.¹⁷⁵ To improve the validity of meta-analyses of IPD incorporating evidence from non-randomised studies and RCTs, the following suggestions have been made:

- i. Critically appraise the risk of bias in the available studies, and carefully decide whether inclusion is appropriate
- ii. Routinely account for confounding factors within the studies
- iii. Explore potential sources of between-study heterogeneity and compare the results between randomised and non-randomised studies¹⁷⁵

All three suggestions are also relevant for the inclusion of both randomised and non-randomised studies in meta-analyses of aggregate data. However, as shown in the review of sciatica treatments presented in Chapter 2 (and discussed in the Appendix Volume I, Section A), RCTs can also be poor quality, at which point data from a well conducted observational study may be more informative. Furthermore, obtaining and using IPD from observational studies or directly from patient registries is not straightforward and very time consuming, as identified by the case study from the GetReal project presented in Section 5.8.6.

5.9.9 The use of data from non-randomised studies

The review of methods showed that the evaluation of treatment sequences is likely to require the use of non-randomised studies, which include both real-world observational studies and non-randomised trials (also discussed in Chapter 1, Section 1.3.3). Non-randomised studies may be required as a complement to the available RCTs, or as the single source of evidence. However, as discussed in the Appendix Volume I (Section A), treatment effect estimates derived from non-randomised studies are at a greater risk of bias than those taken from RCTs. Non-randomised studies should therefore be used in conjunction with appropriate methods to account for this within the evidence synthesis.

5.9.9.1 Potential bias in non-randomised studies

A number of included studies relied on the use of real-world observational studies, and as such, some specific limitations and biases inherent within this type of study design for evaluating treatment sequences were identified. These include the following:

- i. *Selection (allocation) bias* resulting in systematic differences in prognostic factors between individuals in the treatment (e.g. first-line treatment) and control group (e.g. second or subsequent treatment). For example, Lloyd *et al.* included observational studies that compared the efficacy of TNF-inhibitors among cohort of patients receiving their first TNF-inhibitor with a cohort receiving a second or subsequent TNF-inhibitor, in order to examine the presence of sequencing class effect. However, patients who fail on biological agents have worse prognoses than those receiving biologic treatment for the first time, and are likely to show limited responses to all treatments, not just a second TNF-inhibitor.²⁹⁰ Herman *et al.* recommend that when the treatment strategies under investigation are sustained over time, adjustment for both baseline and post-baseline prognostic factors is necessary, to ensure the comparability (exchangeability) of treatment groups.³²⁷
- ii. *Channelling bias* favouring patients with more severe disease. New treatments create expectations of improved effectiveness and tolerability, which means that early, post-marketing users may not be representative of the eventual user population. For example, the first patients with rheumatoid arthritis to use a new immunomodulation drug are likely to be those who experienced little or no benefit from existing drugs and may therefore respond to the new drug in a way that the average patient would not.³²⁸
- iii. *Regression to the mean*. This phenomenon occurs because patients tend to be treated with a second or subsequent treatment, for example a biological agent after the failure of conventional DMARDs, at the height of their disease activity, where there is a greater than 50-50 likelihood that the disease activity will start improving after the intervention purely by chance.
- iv. *Confounding by disease duration*. In some conditions, such as sciatica and rheumatoid arthritis, the longer the disease duration, the less likely that patients will respond to treatments, irrespective of the treatments used.
- v. *Enrichment of successive treatment use with refractory patients*. A small proportion of patients have refractory disease that will not respond to any treatment. For example, epilepsy is resistant to drug therapy in a third of patients,³²⁹ and about 30% of patients with rheumatoid arthritis do not respond to TNF-inhibitors.³⁰³ The population receiving second-line or subsequent treatments are therefore more likely to be enriched with patients who are refractory to any, or a specific type of treatment. This is also related to a class effect. Patients who fail initial treatment due to a tolerability or safety issues are likely to also have the same problem with an alternative drug from the same class, which means the risk for developing an adverse event with, for example a second TNF inhibitor increases twofold in patients who switched due to an adverse event.³⁰³

- vi. *Immortal time bias*, which was found to be particularly relevant for cancer treatments. Heng *et al.*, who compared the effectiveness of second-line treatments for metastatic renal cell carcinoma using real-world observational studies, chose to exclude retrospective cohort studies that required patients to receive a third-line therapy after the initiation of second-line treatment, as the study design would result in the exclusion of a large proportion of second-line patients who would not reach third-line treatment due to loss to follow-up, continuation of second treatment at the time of the chart review, death during second-line treatment, or other reasons, resulting in immortal time bias for the effects of second-line treatment.²⁰³ Studies that limit inclusion to patients who have completed a sequence will also overlook patients who are continuing the initial treatment, or lost to follow-up after first-line treatment due to lack of efficacy, clinical deterioration, or drug acceptability issues.³³⁰
- vii. *Class effect bias*, which is the possibility that the comparison between drug classes may be confounded by differences in the type of patients treated with each class.²⁰³
- viii. *Aggregate data collection*. Many real-world observational studies do not report individual treatment or drug-level data, which means that any subsequent evidence synthesis has to be based on pooled data across treatments at class level, even when there is evidence that individual drug effects can vary within a class.²⁰³
- ix. The potential for *missing or inaccurate data* obtained from real-world practice. Patient registers and administrative databases are rarely set up for evaluating treatment sequencing, and do not generally involve the same level of rigour in recording events as in research studies.^{203 331}

5.9.9.2 Methods for adjusting for potential bias in non-randomised studies within evidence synthesis

The use of data from non-randomised studies means that methods to adjust for selection bias and time-varying confounding factors that were identified as part of the review of methods are required. An overview of different approaches used for bias adjustment for both internal and external biases is presented in the NICE Decision Support Unit (DSU) Technical Support Document (TSD3), entitled 'Heterogeneity: subgroups, meta-regression, bias and bias-adjustment'.³¹⁶ These include the following methods for using within pair-wise or network meta-analyses:

- i. Meta-regression
- ii. External priors to adjust for bias associated with markers of lower study quality
- iii. Network synthesis to estimate and adjust for quality-related bias internally
- iv. Expert elicitation for priors for bias

A more recent review by Efthimiou *et al.*, of the methodological developments and empirical studies in network meta-analysis also includes a summary of the latest methods for adjusting for study limitations and possible sources of bias.⁵⁹ This review, which was undertaken as part of the wider Get Real project (Section 5.8) did not identify any new approaches.

One of the included studies (Schmitz, 2012), which used meta-regression to examine the influence of previous treatments (Section 5.5.3), went on to explore the feasibility of extending their Bayesian network meta-analysis to include observational studies, due to the lack of RCTs directly comparing TNF-inhibitors.²¹¹ The methods are summarised here rather than in the results section as they do not relate to the evaluation of treatment sequences. They represent methods that are likely to be of value for the future evaluation of treatment sequencing and the need to incorporate a broader evidence base, and therefore relevant to the Framework presented in Chapter 8. Schmitz *et al.* differentiated between the effect of non-systematic and systematic bias, with the former requiring adjustment for over precision and the latter adjustment for overestimating or under estimating treatment effects. They compared three different approaches previously used for incorporating both RCTs and observational studies in the same network meta-analyses:

- i. Naive pooling across all study types, disregarding differences in study design.
- ii. Summarising the observational evidence to inform prior distribution in the network meta-analytic model. Here the observational information was down-weighted by inflating the variance parameter, and over or underestimation of the treatment effect adjusted by shifting the mean of the prior information.
- iii. Using a three level hierarchical modelling approach. Here overestimation was adjusted using an additive factor to the mean, and over precision by using a multiplicative factor to the variance.

In order to implement these approaches the authors assumed a 30% over precision, based on previous research. However, the research was inconsistent about overestimation of treatment effects in observational studies, and the actual size of the bias was difficult to estimate. Sensitivity analyses were therefore used to vary this estimate, with the results based on values ranging from 0.7 to 0.1

Naive pooling was the simplest method, but made the strong assumption of no difference between study designs and did not allow for bias adjustment or any additional uncertainty to be taken into account.²¹¹ Using the observational evidence to inform prior distributions allowed the adjustment for potential bias because of over precision or overestimation, but between study design heterogeneity was not taken into account, and it was not possible to extend the model to include more than two different study designs. The hierarchical modelling approach was able to account for the uncertainty arising from combining information from different study designs by random effects. The hierarchy levels also allowed the authors to quantify the impact evidence from different designs had on the results whilst adjusting for potential bias.

Schmitz *et al.* also included an uncontrolled trial in their network meta-analysis.^{211 332} Methods for synthesising data on single arm trials are still under-developed.³³³ The indirect treatment comparison should always be implemented using adjusted methods that preserve the randomisation or within-study comparison, as unadjusted methods do not provide reliable estimates.¹⁵⁰ (This was first referred to in Chapter 2, Section 2.8.2; see also Appendix Volume I, Section A) As such, Ades *et al.* strongly

recommend that single arm studies, whether based on RCTs or observational studies, should not be included in network meta-analyses.³³⁴ To overcome the challenge of including an uncontrolled open-label extension study in their network meta-analyses, Schmitz *et al.* used a matching approach, where baseline characteristics were compared across all the available studies to identify a suitable matching control.³³² However, this method does not control for unobserved variables, and the study was therefore excluded in sensitivity analyses.

5.9.10 Recent methodological developments in meta-analysis

Methodological development in meta-analytic approaches for informed decision-making have been quite prolific in the last two decades,^{58 59 164} and may be relevant for developing novel methods for evaluating treatment sequences. Section 2.8.2 reports on how ongoing methodological development in network meta-analysis (including hierarchical modelling approach), multivariate meta-analysis, and meta-regression could potentially provide the bases for developing novel methods for evaluating treatment sequences.

5.9.10.1 Multivariate network meta-analysis

The recent developments in network meta-analysis include incorporating multiple correlated outcome measures.⁵⁹ The case study from the GetReal project (work package 1) presented in Section 5.8 demonstrates the use of this approach to develop treatment effect estimates in the first and second-line by modelling first-line treatment as outcome one and treatment effects in second-line as the second outcome.

5.9.10.2 Model based meta-analysis

Another method to add to this list of meta-analytic approaches that could potentially provide the bases for developing novel methods for evaluating treatment sequences is longitudinal model based meta-analysis presented under meta-regression in Section 5.5.2 (and used by Mandema, 2011).²⁰⁶ Longitudinal model based meta-analysis capture two important components: magnitude of the treatment effect, which may be related to the dose in a linear or non-linear way, and it's time course.³³⁵

5.9.10.3 Meta-analysis of complex interventions and the use of component-based meta-analysis

Treatment sequencing, as noted in Chapter 1 (Section 1.5.1) can be conceived as a multicomponent complex intervention.⁹⁹ Meta-analytical methods developed for overcoming the challenges of evaluating complex interventions may also be useful for assessing treatment sequences.

The use of component-based network meta-analysis is becoming increasingly common for synthesising complex interventions in the presence of inevitable heterogeneity, where the interventions are disaggregated and individual components become the intervention nodes in the network.³³⁶ An example of this approach is provided by Welton *et al.*,⁶¹ which included the use of a

meta-regression-based extension to network meta-analysis in a Bayesian Framework.³³⁶ The complexity of the analysis was increased gradually over three models:

- i. An additive main effects model which assumed that the effect of each component adds (i.e., no synergistic or antagonistic effects)
- ii. A two way interaction model (allowing pairs of components to have either a bigger or smaller effect than would be expected from the sum of their effects alone)
- iii. A full interaction model for interventions described as having >2 components (e.g., cognitive + behavioural + support)

However, these methods do not account for the time-course of a treatment sequence, or take into account that the choice and application of each treatment or 'component' is dependent on the effect/impact of the previous one. Component network meta-analysis has also been used for evaluating the diagnostic accuracy of a sequence of two tests, the Ddimer test and the Wells Score, to diagnose deep vein thrombosis.³³⁷ The meta-analytic framework was able to allow for the fact that multiple diagnostic tests, when used in combination, may not be independent of one another.³³⁷ I return to this in Chapter 9 (Section 9.4.2).

A recent review of the challenges of conducting systematic reviews of complex multicomponent healthcare interventions, and approaches to address them, was conducted by the US Agency for Healthcare Research and Quality (AHRQ).³³⁸ The findings of this review, along with another related project, which assessed the theoretical foundations of complexity in systematic reviews of interventions,³³⁹ was subsequently used as part of the process for developing the latest guidance and tools for conducting reviews of complex interventions.⁹⁹ The guidance, which is based on a Delphi process and an expert consensus workshop, is presented as a series of papers published in 2017, one of which provides an overview of advanced analytic methods.³¹⁹ This paper presented methods to address a range of research questions; those assessing clinical effectiveness of the interventions included: network meta-analysis and single-case designs for overall effectiveness; multivariate meta-analysis and weighted least-squares estimates for average effect sizes for examining multiple outcomes; and meta-regression and finite mixture modelling (also known as latent class modelling) for the assessment of heterogeneity.³¹⁹

Another recent project funded by the European Commission, called 'Integrated assessments of complex health technologies' (INTEGRATE-HTA) also resulted in the production of a series of guidance for conducting health technology assessment of complex interventions. I discuss this project in more detail, along the usefulness of one of the guidance for evaluation treatment sequencing in Chapter 7, Section 7.4.6.1.

5.10 CHAPTER SUMMARY AND NEXT STEP

5.10.1 Chapter summary

The current chapter summarises the findings of the review of meta-analytic techniques. The main finding was the dearth of methods identified. The findings also showed that reviewing the evidence on treatment sequencing is not straightforward and is severely hampered by the limitation of the evidence base. No novel methods were developed, and the review outlines all the 'standard methods' that have been applied to evaluate treatment sequencing.

The evidence used to inform treatment-sequencing effects was broadly considered in two ways:

- i. In a piecemeal fashion usually based on stratified meta-analysis
- ii. Using research evaluating whole sequences

The former was generally based on RCTs of single treatments. The latter was mainly based on RCTs comparing fixed sequences of up to two treatment lines or observational studies of predefined sequences. Where the evidence is considered in a piecemeal fashion, it is important to develop summary effect estimates that allow for the previous treatments used.

5.10.2 The next step

The review of quantitative evidence synthesis methods identified a number of studies that used decision-analytic modelling to evaluate treatment sequences, most of which were undertaken as part of an economic evaluation. Treatment sequences within decision-analytic models are frequently characterised as a series of individual treatments, each requiring a summary treatment effect estimate that is conditional on positioning in the treatment pathway. However, as shown in the current chapter, the evidence base to inform such estimates is limited. This seriously hampers the ability to implement meta-analytic techniques to develop such estimates. Consequently, modellers had to resort to making simplifying assumptions as substitutes. The next step is to review the range and type of simplifying assumptions used for representing treatment-sequencing effects. The next chapter also explores the type of data selected to inform treatment effects. Key issues identified in the current chapter are used to develop the draft framework.

CHAPTER 6: REVIEW OF SIMPLIFYING ASSUMPTIONS APPLIED TO EFFECT-ESTIMATE IN MODELLING STUDIES FOR EVALUATING TREATMENT SEQUENCES

6.1 CHAPTER OVERVIEW

In this chapter the simplifying assumptions applied to effect-estimates in modelling studies for evaluating treatment sequences are reviewed. Ideally, in decision modelling, simplifying assumptions are used only when they are required to complement the limited data. The type and level of evidence available, as well as the decision problem being investigated, will in turn impact the extent and type of assumptions required.

The chapter starts with a brief introduction to decision analytical modelling and the potential impact of analytic judgements or assumptions on decision uncertainty. An overview of the type of decision problems relating to treatment sequences that decision models and economic evaluations may investigate is also provided. The specific aims and objectives of the review are provided in Section 6.3. A description of categorisation schemes developed for grouping the decision problems investigated, and the simplifying assumptions made, is also provided. The findings of the review of simplifying assumptions are presented in Section 6.5. This covers any clinical condition. The variation in methods and type of data sources selected to inform the treatment effects across modelling studies evaluating a similar decision problem are explored in Section 6.6. This is based on studies that investigated the sequential use of TNF-inhibitors for rheumatoid arthritis.

6.2 INTRODUCTION

6.2.1 Decision-analytic modelling and decision uncertainty

Decision-analytic modelling provides a means of making a decision about which treatment sequence to use despite a limited evidence base. Modelling can provide powerful tools to assess and compare the overall clinical and economic value of different treatment strategies when experimental studies are too complex, long, or expensive to carry out.³⁴⁰ Meta-analyses are conducted in order to make inference about the true treatment effect in a particular population. They may lead to the conclusion that there is insufficient evidence of an effect.³⁴¹ A decision analysis, on the other hand, starts with the premise that a decision must be made. A decision-analytic model makes full use of the best available data and, where the evidence is absent, analytic judgements and assumptions are used.⁹⁰ The impact of the resulting uncertainty is tested using sensitivity analysis.^{342 343} Uncertainty around the true values of model input parameters, which exists even when based on the best available evidence, can be accounted for using probabilistic sensitivity analysis. This is discussed in more detail in the Appendix Volume I (Section E3.6). Ultimately though, the findings of the model are only reliable if its design or structure is consistent with medical practice, its underlying assumptions are valid, and the data sources are robust and used appropriately. If incorrect assumptions or structures are used then the model results will be flawed and the decision may be incorrect.³⁴⁴ Any data sources or assumptions

used to inform sequencing effects therefore need to be valid and robust. Appropriate summary estimates of the decision uncertainty are also needed.³⁴⁵

6.2.2 Methods for handling structural uncertainty

The use of simplifying assumptions to represent sequencing effects may result in structural uncertainty, which relates to the conceptual and mathematical representation of a decision problem within a model.³⁴⁶ Structural uncertainty is less well described than parameter uncertainty,³⁴⁷ which is summarised in the Appendix Volume I (Section E3.6). It is commonly recommended that structural uncertainties are explored informally using scenario analyses, presenting the results under different model structures.³⁴⁷ Structural uncertainty clearly matters when different credible scenarios produce different results that suggest different decisions.³⁴⁸ However, this does not indicate which is the most credible scenario, and there are no established methods to formally assess the plausibility of alternative models.³⁴⁷ Proposed alternative approaches for characterising structural uncertainty include: *model averaging*, i.e. calculating the sum of the outcomes of a set of plausible models, weighted by some measure of their adequacy or credibility;^{348 349} and *discrepancy modelling*, which involves making judgements about the discrepancy between the model output and the 'true' target value.³⁴⁹ However, these methods require data to inform the model parameters for alternative assumptions. Where explicit assumptions cannot be assessed against data, Jackson and others, in their framework for addressing structural uncertainty,³⁴⁷ propose the use of expert elicitation.³⁴⁸ However, Jackson *et al.* go on to state that where sufficiently robust expert beliefs cannot be obtained, for example when extrapolating short-term evidence into the future, uncertainty must be addressed by presenting the results under different scenarios. This will show the main assumptions driving the results, and how the results change under different assumptions. Quantitative measures of model adequacy, such as the model weights can be used to help decision makers interpret a table of scenarios.³⁴⁷

6.2.3 The use of simplifying assumptions in decision-analytic models of treatment sequences

Decision-analytic models are generally implemented by first applying a summary treatment effect to the baseline patient population. This initial treatment effect, which needs to be based on the best available evidence, is generally informed by RCTs with short-term follow-up.³⁵⁰ This is then extrapolated using the model, supplemented by further evidence sources, in order to estimate the continued treatment effect and long-term impact.

The use of decision-analytic modelling to compare predefined sequences is usually based on the application of a summary treatment effect to each individual treatment in the sequence. The review explores the methods or assumptions used to inform the initial summary effect of each individual treatment used in the sequence being modelled. This includes individual treatments used early on in the treatment pathway or later stages. Where meta-analytic methods were used to produce these summary effect-estimate or sequencing effects, these are presented in Chapter 5. The current chapter focuses on the assumptions made regarding these effect-estimates conditional on their

positioning or the previous treatments used. The different modelling techniques used, are presented in Chapter 7.

6.2.4 Decision problem relating to treatment sequencing

The first stage of developing any decision-analytic model is to specify the question or decision problem.³⁵¹ The decision which a model aims to inform will then determine the appropriate model structure and complexity.³⁵²

A number of different types of decision problems relating to treatment sequencing may be investigated using a decision-analytic model. The decision-analytic model may be used to evaluate different sequencing approaches used over the course of the disease. For example, the decision problem may be about whether to use a 'step-up' versus 'step-down' drug regimen, starting with either the least toxic or the most effective drug first, or it could be about investigating whether a different class of drug is required upon treatment failure. Alternatively, it may relate to the best overall sequencing approach. A common problem faced by policy makers is the identification of the optimum positioning of a new treatment within the existing treatment pathway. Reimbursement and health technology assessment bodies, such as NICE, may be interested in the actual added value of a new treatment, and whether its addition to the current pathway is cost-effective, whilst taking into account the downstream costs of subsequent treatments. The decision problem could relate to a simple comparison of multiple new treatments, and their use within the current treatment pathway. I re-visit the range of decision problems relating to treatment sequences in Section 6.4.2, which describes a coding scheme developed to summarise the decision problem investigated by included studies.

Models that consider decision problems relating to the best overall sequencing encompass yet another layer of complexity. Examples include: which specific drug to use first, the duration of time over which to trial an intervention, whether to increase the dose before trying another treatment, and issues relating to individualising treatment. Albert *et al.* argue that, in these circumstances it's the philosophy of the approach, rather than the specific algorithms, that is really being debated. This is because the complexity and ambiguity of the situation is such that specific algorithms are not possible; there are too many options and variables to construct specific algorithms²¹⁶ (Chapter 1, Section 1.5.1). Furthermore, decisions on treatment sequencing made in clinical practice are adaptive in nature (as outlined in Chapter 1, Section 1.2.1). These are based on both the accruing data on the current individual circumstance over time and the embedded knowledge from previous cases. The practicality of conducting decision analysis means that decisions have to be made from the outset on which drug should be included in the algorithm and at which point, whether to treat all treatments in the sequences as a 'class effect', and how many treatments to allow for within a given sequence.

6.2.5 Accounting for treatments administered before and after the decision point of interest

All treatments for chronic conditions are in fact part of a treatment sequence. It could, therefore, be argued that all decision-analytic models are involved with 'treatment sequences'. For some decision

problems the previous and subsequent treatments may be ignored or greatly simplified in the model, but why and how this is done should be clearly justified.

A decision-analytic model represents a simplified reality, and usually considers only what is pertinent or relevant to the decision problem. This means that any variation between the treatment strategies being compared that does not impact the decision problem, or its cost-effectiveness, is generally ignored. The evaluation of treatment sequences acknowledges that treatment effects are likely to be influenced by the choice of both prior and subsequent treatments used.

Decision-analytic models that start at the point of diagnoses are more likely to reflect the complete treatment sequences used in chronic conditions (Chapter 1, Section 1.4.2.2). However, where treatments are modelled from an earlier point in the disease process, the likelihood that there is no matching evidence increases, and more assumptions are required. Modelling studies that only consider the impact of subsequent treatments are generally based on the assumption that the sequences being compared are starting from a level playing field. The potential impact of this is not generally considered within the sensitivity analysis, as it is not part of the cost-effectiveness estimates. The included modelling studies of biological agents for rheumatoid arthritis, the most common disease condition investigated, used a range of starting points in the treatment pathway, even when considering the same decision problem. The type of data selected to inform the effects of treatments administered before the decision point of interest is explored further in the Appendix Volume I (Section D).

The review focuses on decision-analytic modelling studies that aim to evaluate treatment sequences. However, economic models often focus on the comparison of single treatments, and tend incorporate treatment sequences to account for the impact of the downstream costs of subsequent treatments beyond this decision point. The downstream effects of subsequent treatments are also modelled in order to reflect clinical practice for chronic conditions. The effects of subsequent treatments, administered after the decision point of interest, are often handled in different ways. For example in the economic model developed as part of the NICE appraisal comparing the effectiveness of TNF-inhibitors for rheumatoid arthritis (NICE TA130; Chen, 2006, which was introduced in Chapter 4, Section 4.4),²²⁷ the initial treatment response to each conventional disease-modifying anti-rheumatic drug (DMARD) used after the TNF-inhibitor was explicitly modelled. In another technology appraisal comparing effectiveness of TNF-inhibitors for psoriatic arthritis (NICE TA199; Rodgers, 2011, Section 4.4),¹⁷ the economic model assumed that patients experienced a steady long-term deterioration after the failure of the TNF-inhibitor, and the fluctuations caused by response to subsequent conventional DMARDs, which were considered to be administered as part of palliative care, were ignored. This is explored further in the Appendix Volume I (Section D).

In oncology, there is a need to account for the clinical effect of downstream chemotherapy treatments for advanced cancer, as they have an impact on overall survival. (This issue was introduced in

Chapter 4, Section 4.4, and is discussed in more detail in the Appendix Volume I, Section C4). This is also handled in different ways. For example, in the NICE technology appraisal evaluating targeted therapies for advanced colorectal cancer (NICE TA93), Hind *et al.* limited inclusion to sequencing trials of up to two lines of treatment (discussed in Chapter 5, Section 5.3.2).²⁰⁴ In the NICE clinical guideline evaluating chemotherapy sequences for advanced breast cancer (NICE CG81), the overall survival was estimated as part of the modelling process, based on the sum of two separate pooled estimates of progression free survival for first and second line treatment, and a fixed arbitrary estimate for survival in progressive disease.²⁵⁷

6.2.6 Potential evidence gap and displacement effect

The available evidence to inform treatment-sequencing effects impacts the type of assumptions required. A frequent problem when evaluating the introduction of a new treatment to an established sequence is that specific evidence gaps sometimes appear regarding current treatments. These develop over time as a result of the process by which new drugs are introduced into the market, and the indication for which they are licenced. This was first introduced in Chapter 1, Section 1.3.2, and illustrated in the clinical case studies discussed in the Appendix Volume I (Section C). These evidence gaps are likely to remain, as future research concentrates on the evaluation of the subsequent new treatments. This in turn leads to a common dilemma in terms of the limited evidence base to inform the 'displacement effect'. When adding a new treatment to an established sequence, the treatment being displaced lower down the sequence will not only represent the subsequent treatment in the sequence with the new treatment, but also the comparator, in the same sequence without the new drug. The variation in the type of data used to address such an evidence gap is explored in more detail in Section 6.6, using sequential TNF-inhibitors in rheumatoid arthritis as an example.

6.3 AIM AND OBJECTIVES OF THE REVIEW OF SIMPLIFYING ASSUMPTIONS USED IN DECISION-ANALYTIC MODELLING STUDIES

The overall aim of this part of the review of decision-analytic modelling studies was to investigate the simplifying assumptions made regarding the treatment-sequencing effects, or the response to each treatment used in the sequence.

The specific objectives of this review were to identify and review:

- i. what analytic judgements or assumptions were used in the presence of limited, or absent data regarding treatment response conditional on positioning in the treatment pathway and
- ii. the extent of the variation in the choice and source of the accompanying efficacy data in the presence of limited, or the absence of data regarding treatment-sequencing effects for a similar decision problem.

The second objective, evaluating the consistency in the methods used to select data to inform treatment-sequencing effects, was addressed in more detail using modelling studies that investigated the use of treatment sequences for rheumatoid arthritis. However, the findings are also considered relevant to other disease conditions. The purpose here was to guide future methods, and not to identify and review the latest evidence on the effectiveness of treatments for rheumatoid arthritis.

Most modelling studies included the costs of palliative or best supportive care. The actual treatments used in palliative care were sometimes named in order to estimate these costs. However, only the sequencing effects of active treatments were considered here. The review focuses on treatment sequences, and the term 'treatment' is used for monotherapy, combination therapy or a category of treatments.

6.4 METHODS FOR THE REVIEW OF SIMPLIFYING ASSUMPTIONS USED IN DECISION-ANALYTIC MODELLING STUDIES

6.4.1 Identifying relevant studies

The literature searches and the methods used to identify and review the relevant modelling studies are described in Chapter 3.

6.4.2 Coding the decision problem and simplifying assumptions

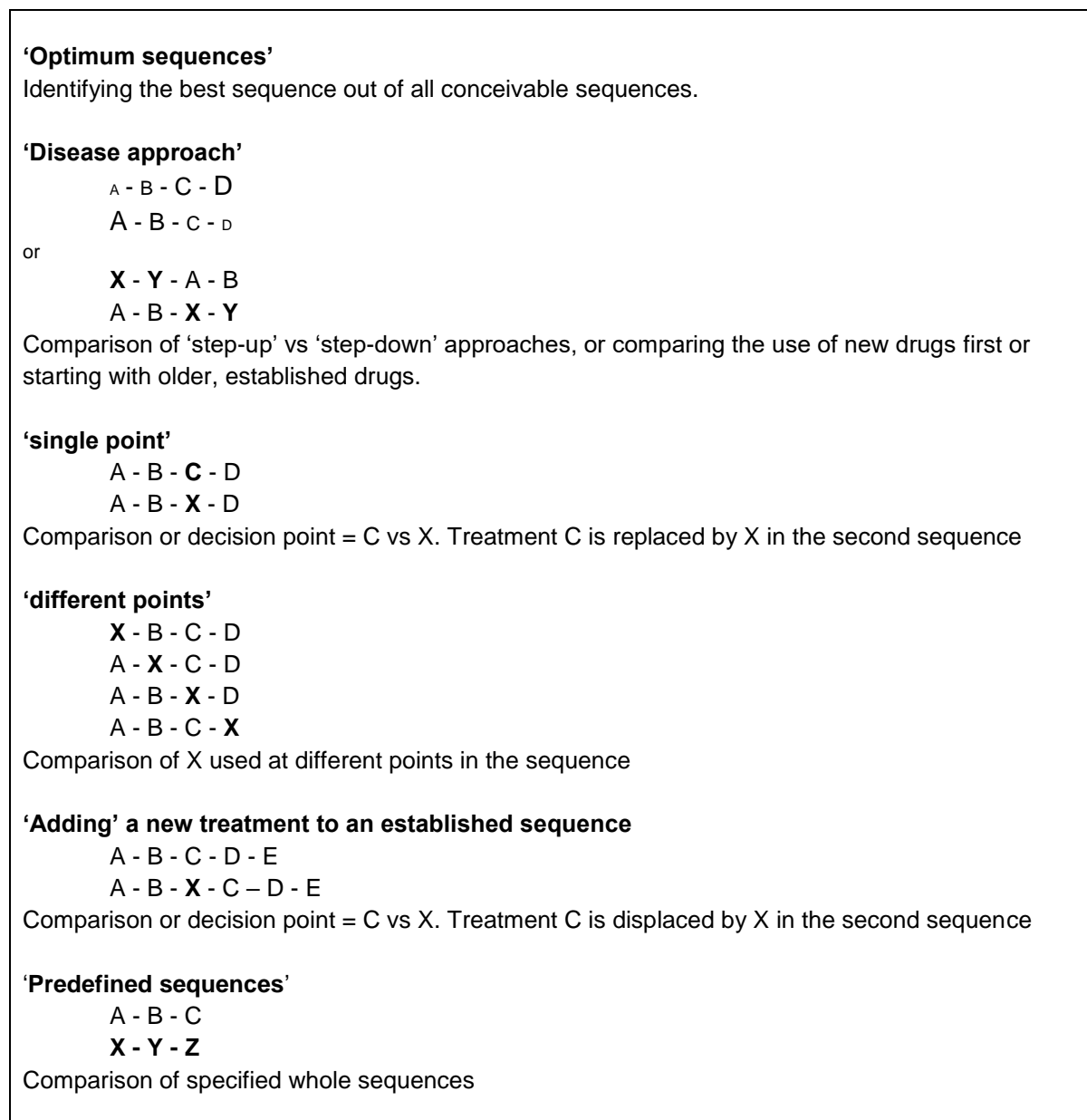
Included decision-analytic modelling studies were coded according to their underlying purpose, or decision problem, in relation to treatment sequencing. The decision problem categories are illustrated in Figure 6.1.

Studies were first assessed to determine whether they aimed to identify the 'optimum sequence' out of all conceivable treatment sequences, or to compare predefined sequences, thus selecting a manageable number of sequences for comparison in advance. Studies investigating predefined sequences were then coded according to the type of decision problem they related to, using the following categories:

- i. Disease approach, e.g. the comparison of step up vs step down treatment regimens
- ii. Single point, e.g. comparison of two treatments used as the same line of therapy
- iii. Different points, e.g. the same treatment used at different points in a treatment pathway
- iv. Adding, e.g. adding a new treatment to the start or the end of an established or current treatment pathway
- v. Predefined sequences, for studies that aimed to compare specified whole sequences

Decision problems were coded as adding if they included the comparison of the same sequence with and without the addition of a specified treatment. The different points coding was reserved for the comparison of a specific treatment used at different locations, but did not include the comparison of a sequence without this treatment. Some studies included multiple decision problems, and therefore had more than one decision code.

Figure 6.1: Illustration of treatment sequences according to decision problem



The simplifying assumptions used to represent sequencing effects were coded using the scheme developed as part of the review process and presented within the results in Section 6.5.3 and Figure 6.2. The codes were developed in an iterative process as part of the data extraction presented in the Appendix Volume II. (Appendix D). For each study a summary of the assumptions relating to the sequencing effect-estimate were initially extracted as a narrative, and then given a descriptor. The same descriptor was then used for the next study that used a similar assumption. The assumptions that were coded relate to those used to inform the initial treatment effect of individual treatments conditional to their positioning in the modelled sequences. Included studies that investigated a decision problem relating to adding a new drug were also assessed according to whether they ignored the displacement effect, by using the same treatment effect for the comparator, irrespective of

which sequence it related to. This is included here as it is an approach that assumes there are no sequencing effects at play.

6.5 RESULTS OF THE REVIEW OF SIMPLIFYING ASSUMPTIONS

6.5.1 Simplifying assumptions coding scheme

The new coding scheme developed as part of the review process is presented in Figure 6.2. Individual codes are described in more detail using the findings of the review of modelling studies presented in section 6.4.3. The coding scheme includes 12 codes grouped under six headers that relate the following:

- i. Treatment sequencing considered to have no or limited impact
- ii. The effect of some treatments is the same as an alternative, substitute treatment
- iii. The effect of treatment is thought to decrease with subsequent use
- iv. A treatment used for relapse is considered to have the same effect as its initial use
- v. Displacement effect is not accounted for
- vi. The use of alternative data sources to inform treatment sequence effects

The application of the coding system to the wider group of included studies highlighted some challenges. These were primarily due to the differential application of multiple simplifying assumptions (within the same study) across different treatments, different points in the sequences, and when conducting scenario analysis. The information provided on the simplifying assumptions used was frequently limited, and those applied to the treatments used subsequent to the decision-point of interest were especially poorly reported. In some instances, the simplifying assumptions reported by the authors did not account for all the compromises made due to the limited evidence base, especially for treatments used later in the sequences being modelled. One code (RDD) relating to the potential reduction of treatment effect due to disease duration, turned out to be particularly difficult to apply. This was not surprising because, as discussed in relation to the use of meta-regression (Section 5.5), it is difficult to disentangle whether long standing disease *per se* is associated with poor response to treatment, or whether prior failure of previous treatments predicts response to subsequent treatments.³⁰² Due to the iterative process of developing the coding system, this code was initially conceived from one study (Fitzsimmons, 2014), which used sensitivity analysis to account for the fact that treatment effect can diminish with time alone, as well as the number of previous treatments.¹¹⁵ However, studies that were subsequently data-extracted did not distinguish between these two issues, making it difficult to apply this particular code.

The coding system also represents an important contribution to the framework presented in Chapter 8. It can potentially be used as an aid to considering treatment sequences as part of the decision problem; to inform the choice of approach to use within a health technology assessment by clarifying what has actually been done previously; and to highlight whether modellers have used the same or different approaches within technology appraisals of a similar decision problem.

Figure 6.2: Coding scheme used for coding simplifying assumptions relating to treatment-sequencing effects used by included studies

Each study may include more than one code.

CODE SIMPLIFYING ASSUMPTION

Treatment independence

IP Treatment effect is *independent of positioning* in treatment sequence.

NPT Treatment effect is dependent on the *number of previous treatments* used, but independent of the type of treatments used.

Substitution with another treatment effect

GE Treatment effect is the same as an alternative treatment from the same class, or a generic class effect - irrespective of positioning in the sequence (*generic effect*).

PGE Treatment effect is the same as an alternative treatment from the same class, or a generic class effect - matching the same position in the sequence (*positional generic effect*).

ST Treatment effect is the same as an alternative (substitute) treatment from a different class of treatments, used at the same point in the sequence (*substitute treatment*).

Reduction of treatment effect

RF Treatment effect is reduced, in line with a multiplier or *reduction factor*, when used at a later point in the sequence. (Here the specific reduction or multiplication factor used to develop the diminishing effect, is informed by the available evidence that is also relevant to the treatment of interest.)

TD Treatment effect decrements by the same pre-set amount with each successive treatment (*decrementing effect*). (Here the same generic proportional reduction, used to represent the diminishing effects, is applied at each point in the sequence irrespective of the treatment used. The proportion is not necessarily based on a specific evidence base.)

RDD Treatment effect is *reduced with disease duration*, and treatments are not as effective when they are used in late disease.

Impact of time since previous treatment

LR Treatment effect is not affected by previous treatments if patients have been in *long term remission*, and thus can re-use the same treatment(s)/class of treatment(s) as that which achieved the prior remission.

Displacement effect ignored

DI A single treatment effect does not differ when it is *displaced* (i.e. its position in the sequence is changed) by the addition of a new prior treatment (*displacement ignored*)

The use of uncontrolled/observational studies without bias adjustment

UOBS *Uncontrolled* trials or *observational studies* provide an un-biased estimate of treatment (sequencing) effects.

EXC *Expert consensus* provides an un-biased estimate of treatment-sequencing effects

6.5.2 Description of included studies

Sixty six studies were included in the review assessing the range of simplifying assumptions used in decision-analytic models to account for sequencing effects.^{17 18 115 197 199 216-242 244-248 250-254 256-279} These are listed in Tables 6.1 and 6.2, which outline the modelling approaches used, the decision problems of interest, the length of treatment sequences, and the type of simplifying assumptions used. A large proportion of the modelling studies investigated treatment sequences for rheumatologic conditions (Table 6.1). These represent a single body of research evaluating sequential DMARDs for a group of arthritic conditions with common features (described in more detail in the appendix Volume I, Section C3), and are therefore listed separately from those of non-rheumatologic conditions (Table 6.2). A more detailed summary of each included study, including the patient population of interest, the treatment sequences evaluated, the available evidence base, and the simplifying assumptions made, is presented in the Appendix Volume II. (Appendix D).

Six studies, all of which evaluated treatments for non-progressive chronic conditions, aimed to identify the optimum sequence of treatments from all conceivable sequences. Four studies (Anis, 2011; Sawyer, 2013; Sizto, 2009; Woolacott, 2006) evaluated psoriasis,^{217 263 267 278} one included onychomycosis (fungal infection of the toenail; Frankum, 2005),²³⁹ and one study (Smith, 2007) was of postherpetic neuralgia (a nerve pain that persists after a shingles rash has cleared).²⁶⁸ The modelling approaches used, which are described in more detail in Chapter 7, were only feasible by assuming that treatment effects were independent of position in the sequence (which I come back to in Section 6.5.3.1). Treatments were ranked according to the outcome for individual treatments, in order to identify the optimum sequence.

The remaining studies, which included all the rheumatology studies, compared a set number of pre-specified sequences in order to evaluate:*

- i. Adding a new drug to a sequence (n=28)
- ii. Comparing different treatment approaches by incrementally increasing or decreasing the number of treatments used prior to a fixed sequence (n=10)
- iii. Comparing different drugs used at the same point within a sequence, or replacing one of the drugs in the sequence with a new treatment (n=20)
- iv. Using the same drug at different points within the same sequence (n=10)
- v. Simply comparing different predefined sequences (n=16)

(*Some studies included multiple decision problems as described in Section 6.4.2)

The need for simplifying assumptions regarding sequencing effects is dependent on the extent or limitation of the available evidence base. This in turn is influenced by the length and variability of the sequences evaluated, and whether reasons for discontinuation were considered. If, for example, the decision problem related to the comparison of a sequence of only two lines of therapy, or if there are matching trials for each line of therapy with subsequent treatments predefined in the protocol, then suitable sequencing trials may exist. However, when considering multiple interventions over a lifetime

horizon for chronic conditions, there is a greater potential for imperfect evidence, and a need for simplifying assumptions. Most (63%) of the rheumatology studies included a lifetime horizon, and the number of treatment lines within included sequences ranged from two to 12 (Table 6.2).

Table 6.1: Summary and coding of rheumatology studies

(Studies ordered alphabetically by author)

Author, year Country	Condition	Modelling approach (Time horizon)	Sequencing decision problem	Lines of treatment	Simplifying assumption code
Albert, 2000 US	RA	Markov cohort (not stated)	Disease approach	3	IP EXC
Bansback 2005 Sweden	RA	Individual sampling (lifetime)	Adding Single point (1st-line) (Simple sequence)	1 or 2	NPT IP UOBS GE DI
Barton 2004 UK (NICE TA36)	RA	Individual sampling (lifetime)	Adding (3 rd or last drug) Different points	9 or 10	NPT PGE IP GE DI
Beresniak, 2011 Spain	RA	Decision Tree (2 years)	Single point (2 nd and 3 rd) Predefined sequences	3	NPT PGE UOBS
Beresniak, 2013 Germany	RA	Decision Tree (2 years)	Single point (2 nd and 3 rd) Predefined sequences	3	NPT PGE UOBS
Brennan, 2004 UK (a submission for NICE TA36)	RA	Individual sampling (lifetime)	Adding (start)	3 to 4	NPT DI IP GE UOBS
Brennan, 2007 UK (a submission for NICE TA36)	RA	Individual sampling (lifetime)	Single point (2 nd TNF) (Simple sequence)	2	NPT GE IP PGE UOBS
Chen, 2006 UK (NICE TA130)	RA	Individual sampling (lifetime)	Adding (1-3) Different points (1 st , 3 rd , last)	10 to 11 (or 13)	IP RDD DI GE
Cimmino 2011 Italy	RA	Decision Tree (2 years)	Single point (2 nd and 3 rd) Predefined sequences	3	NPT PGE UOBS
Clark, 2004 UK (NICE TA72)	RA	Individual sampling (lifetime)	Adding (1 st , 3 rd , or last; after 1-2 TNFs) Different points	9, 10, or 11	IP RF RDD DI
Coyle, 2006 Canada	RA	Markov cohort (5 years)	Adding (4 th and 5 th or 2 nd and 3 rd , depending on toxicity) Different points	4 to 5 or 2 to 3 (based on toxicity)	IP DI
Davies, 2009 US	RA	Individual sampling (lifetime)	Single point (1 st drug) Adding (1-2 TNFs) Disease approach	4, 5 or 6	NPT IP GE UOBS DI
Diamantpoulus, 2012 Italy	RA	Individual sampling (lifetime)	Adding (start) Single point (1 st line)	4 or 5	NPT RF GE IP DI
Diamantpoulus, 2014 UK	RA	Individual sampling (lifetime)	Adding (1 st or 2 nd -line) Single point (1 st line)	3 or 4, and 6 or 7	IP NPT GE DI
Finckh, 2009 US	RA	Individual sampling (lifetime)	Disease approach	4 or 7	GE IP

Author, year Country	Condition	Modelling approach (Time horizon)	Sequencing decision problem	Lines of treatment	Simplifying assumption code
Hallinen, 2010 Finland	RA	Individual sampling (lifetime)	Adding (1 st , 2 nd or 3 rd - line) Disease approach	4, 5, or 6	IP DI
Jobanputra, 2002 UK (NICE TA36)	RA	Individual sampling (lifetime)	Adding (3 rd or last drug) Different points	9 to 10	NPT PGE IP GE DI
Kielhorn, 2008 UK	RA	Individual sampling (lifetime)	Adding (1 st -line)	3, 4, 5, or 6	IP GE DI
Kobelt, 2011 Sweden	RA	Individual sampling (10 years)	Single point (start) Disease approach	1 to 2	PGE UOBS
Lindgren, 2009 Sweden	RA	Individual sampling (lifetime) ¹	Adding (1 st -line)	3	PGE IP DI UOBS
Maetzel, 2002 Canada	RA	Markov cohort (5 years)	Adding (2 nd or 4 th drug, depending on toxicity)	3 to 4, or 5 to 6 (based on toxicity)	IP DI
Malottki, 2011 UK (NICE TA195)	RA	Individual sampling (lifetime)	Adding (start)	4 to 5	UOBS PGE NPT RDD DI
Merkesdal, 2010 Germany	RA	Individual sampling (lifetime)	Adding (start)	4 to 5	IP DI
Puolakka, 2012 Finland	RA	Decision Tree (2 years)	Single point (2 nd -line) Predefined sequences	3	IP NPT PGE UOBS
Rodgers, 2011 UK (NICE TA199)	PsA	Markov cohort (lifetime)	Adding (2 nd TNF) (<i>Simple sequence</i>)	2	RF UOBS
Russell, 2009 Canada	RA	Decision Tree (2 years)	Single point (1 st and 2 nd) Predefined sequences	3	NPT ST PGE RF DI
Saroux, 2010 France	RA	Decision Tree (2 years)	Single point (2 nd and 3 rd) Predefined sequences	3	NPT PGE UOBS
Schadlich, 2005 Germany	RA	Partitioned survival (3 years)	Adding (1 st and 2 nd) Different points	1 or 2, 4 or 5, and 5 or 6	IP RF DI
Schipper, 2011 Netherlands	RA (early)	Markov cohort (5 years)	Disease approach Different points	5	IP PGE UOBS
Tanno, 2006 Japan	RA	Markov cohort (lifetime)	Adding (start)	3 or 4	NPT GE UOBS DI
Tran-Duy, 2011 Netherlands	AkS	Individual sampling (lifetime)	Adding (3 rd and 4 th line)	5 or 7	NPT RF PGE GE IP DI
Tran-Duy, 2014 Netherlands	RA	Individual sampling (lifetime)	Adding (3 rd and subsequent) Disease approach	8 or 12	NPT DI PGE
Wailoo, 2006 US	RA	Individual sampling (lifetime)	Single point (1 st -line) Predefined sequences (with 2 nd and 3 rd TNF- inhibitor)	2 or 3	IP
Welsing, 2005 Netherlands	RA	Markov cohort (5 years)	Adding (start: 1-2 drugs) Disease approach	1, 2 or 3	IP TD DI
Wu, 2012	RA	Markov cohort	Adding (start: 1-2 drugs)	4, 5, or 6	NPT

Author, year Country	Condition	Modelling approach (Time horizon)	Sequencing decision problem	Lines of treatment	Simplifying assumption code
China		(lifetime)			IP GE UOBS DI

Simplifying assumption code: **GE** treatments from the same class or category produce identical (generic) treatment effects, and that these are independent of positioning in the treatment sequence; **DI** there is no treatment displacement effect; **EXC** expert consensus provides unbiased estimates of treatment sequencing; **IP** treatment effect is independent of positioning in sequence; **LR** treatment effects were not affected by previous treatments when patients had been in long term remission; **NPT** treatment effect was dependent on the number of previous treatments used, but independent of the type of treatments used; **PGE** treatments from the same class or category produce identical (generic) treatment effects, when used at the same position in the treatment sequence; **RF** treatment effect is reduced in line with a reduction factor; **RDD** there is a reduction in efficacy with increased disease duration; **ST** treatment effect is the same as a substitute treatment from a different class but used at the same position in the sequence; **TD** Treatment effect decrements by the same pre-set amount with each successive treatment; **UOBS** observational studies or uncontrolled trials provided an un-biased estimate of treatment effects.

Table 6.2: Summary and coding of non-rheumatology studies

(studies are ordered alphabetically by condition and then author)

Author, year Country	Condition	Modelling approach (Time horizon)	Sequencing decision problem	Lines of treatment	Simplifying assumption code
Cameron, 2008 UK	Cancer (aBC)	Markov cohort (10 years; lifetime)	Adding (2 nd or 3 rd drug)	4 to 5	NPT IP DI
Dranitsaris, 2011 Malaysia	Cancer (mCRC)	Decision tree (lifetime)	Adding (2 nd drug) (<i>Simple sequence</i>)	2	NPT DI
Heeg, 2015 Netherlands	Cancer (multiple myeloma)	Markov cohort (lifetime)	Predefined sequences	4	NPT
Lee, 2013 South Korea	Cancer (Ovarian cancer)	Markov cohort (10 years; lifetime)	Single point (2 nd -line = model starting point)	5	NPT IP
Lux, 2009 Germany	Cancer (aBC)	Markov cohort (10 years; lifetime)	Adding (2 nd -line)	4 to 5	NPT DI EXC
NICE CG81 (2009)	Cancer (aBC)	Decision tree (lifetime)	Predefined sequences	Up to 3	NPT IP TD UOBS
Soini, 2012 Finland	Cancer (Follicular non-Hodgkin lymphoma FL)	Markov cohort (lifetime)	Predefined sequences	2 (+/- maintenance therapy)	PGE LR
Wong, 2009 US	Cancer (mCRC)	Markov cohort (lifetime)	Predefined sequences	Up to 3	NPT IP
Shepherd, 2006 UK (NICE TA96)	Chronic Hep B infection	Markov cohort (lifetime)	Predefined sequences	up to 3	IP
NICE CG152 (2012)	Crohn's disease	Decision tree / Markov Cohort (2 years)	Predefined sequences	Up to 4	IP TD NPT
Greenhalgh, 2005 UK (NICE TA59)	Major depressive disorder (MDD)	Decision tree (1 year)	Different points	3	IP NPT
Erhardt, 2012 Germany	Type 2 diabetes mellitus	Individual sampling (lifetime)	Single point (2 nd line) (<i>Simple sequence</i>)	3	IP
Connock, 2006 UK (NICE TA79)	Epilepsy (<i>in children</i>)	Individual sampling (15 years)	Adding Different points	Up to 4	RF DI
Knoester, 2007 Netherlands	Epilepsy	Decision Tree (1 year)	Predefined sequences (<i>Simple sequence</i>)	Up to 2	PGE
NICE CG137 (2012)	Epilepsy (<i>in children and adults</i>)	Markov cohort (15 years)	Single point (1 st and 3 rd -line)	Up to 3	RF IP
Wilby, 2005 UK (NICE TA76)	Epilepsy	Semi-Markov cohort (15 years)	Single point (1 st and 3 rd -line)	Up to 3	IP
Beard, 2011 US	Fibromyalgia	Markov cohort (2 years)	Adding Different points	5 to 6	IP DI

Author, year Country	Condition	Modelling approach (Time horizon)	Sequencing decision problem	Lines of treatment	Simplifying assumption code
Denis, 2008 France	Glaucoma	Individual sampling (5 years)	Predefined sequences (<i>Simple sequence, estimating probability of starting 3rd-line treatment</i>)	2	IP
Orme, 2012 UK	Glaucoma	Markov cohort (10 years)	Single point (1 st -line)	3	IP GE
Holmes, 2006 South Africa	HIV	Individual sampling (lifetime)	Predefined sequence	Up to 2	RF GE UOBS
Tebas, 2001 US	HIV	Markov cohort (10 years)	Disease approach	Up to 3	EXC
Frankum, 2005 US	Onychomycosis	Decision Tree (1-3 years)	Optimum sequence	3	IP UOBS LR
Smith, 2007 US	Postherpetic neuralgia (PHN)	Markov cohort (lifetime)	Optimum sequence	4 to 5	IP
Anis, 2011 US	Psoriasis	Markov cohort (10-16 weeks)	Optimum sequence	Up to 6	IP
Sawyer, 2013 UK (NICE CG153)	Psoriasis	Markov cohort (1 year)	Optimum sequence	Up to 3	IP
Sizto, 2009 Canada	Psoriasis	Markov cohort (Unknown)	Optimum sequence	6	IP
Woolacott 2006 UK (NICE TA103)	Psoriasis	Markov cohort (10 years)	Optimum sequence	3 and 7	IP
Davies, 2008 UK	Schizophrenia	Markov cohort (10 years)	Predefined sequences	3	IP NPT
Heeg, 2008 Netherlands	Schizophrenia	Individual sampling (5 years)	Single point (1 st -line)	Up to 4	IP
Fitzsimmons, 2014 UK	Sciatica	Decision Tree (1 year)	Disease approach	Up to 3	IP TD RDD
Bensmail, 2009 France	Spasticity	Decision tree (2 years)	Single point (1 st -line)	3 to 4	IP UOBS EXC

Simplifying assumption code: **GE** treatments from the same class or category produce identical (generic) treatment effects, and that these are independent of positioning in the treatment sequence; **DI** there is no treatment displacement effect; **EXC** expert consensus provides unbiased estimates of treatment sequencing; **IP** treatment effect is independent of positioning in sequence; **LR** treatment effects were not affected by previous treatments when patients had been in long term remission; **NPT** treatment effect was dependent on the number of previous treatments used, but independent of the type of treatments used; **PGE** treatments from the same class or category produce identical (generic) treatment effects, when used at the same position in the treatment sequence; **RF** treatment effect is reduced in line with a reduction factor; **RDD** there is a reduction in efficacy with increased disease duration; **ST** treatment effect is the same as a substitute treatment from a different class but used at the same position in the sequence; **TD** Treatment effect decrements by the same pre-set amount with each successive treatment; **UOBS** observational studies or uncontrolled trials provided an un-biased estimate of treatment effects.

6.5.3 Summary of the simplifying assumptions used

The overall frequency with which each simplifying assumption was used is summarised in Table 6.3. Most studies had multiple codes assigned, where different assumptions were used for different treatments, treatments used at varying positions in the sequence, and different assumptions used in sensitivity analyses. The challenge of applying the coding system to these studies is highlighted in the examples provided in the text below.

Table 6.3: Summary of the frequency of use of the simplifying assumptions

Simplifying assumption used	Total (n=66)	Rheumatology studies (n=35)	Non-rheumatology studies (n=31)
<i>Treatment independence</i>			

Simplifying assumption used	Total (n=66)	Rheumatology studies (n=35)	Non-rheumatology studies (n=31)
Independent of positioning (IP)	48 (72%)	25 (71%)	23 (74%)
Dependent on number of previous treatments used (NPT)	29 (44%)	19 (54%)	10 (32%)
<i>NPT used in conjunction with IP</i>		11/19	6/8
<i>Substitution with another treatment effect</i>			
Generic effect (GE)	16 (24%)	14 (40%)	2 (6%)
Positional generic effect (PGE)	17 (26%)	15 (43%)	2 (6%)
Substitute treatment (ST)	1 (2%)	1 (3%)	-
<i>Reduction of treatment effect</i>			
Reduction factor (RF)	9 (14%)	6 (17%)	3 (10%)
Decrementing effect (TD)	4 (6%)	1 (3%)	3 (10%)
Reduced with disease duration (RDD)	4 (6%)	3 (9%)	1 (3%)
<i>Impact of time since previous treatment</i>			
Long term remission (LR)	2 (3%)	-	2 (6%)
<i>Displacement effect ignored</i>			
Displacement ignored (DI)	28 (42%)	23 (66%)	5 (16%)
<i>The use of uncontrolled/observational studies without bias adjustment</i>			
Uncontrolled trials or observational studies (UOBS)	20 (30%)	16 (35%)	4 (13%)
Expert consensus (EXC)	4 (6%)	1 (3%)	3 (10%)

6.5.3.1 Treatment independence

The most common simplifying assumption used by included modelling studies was that treatment effect was independent of positioning in the treatment sequence (***coded as IP***), essentially allowing treatment sequencing issues to be ignored. Forty-eight (72%) studies assumed that the effects of either all or some of the treatments were independent of treatment sequence. This included all six studies that aimed to identify the optimum sequence of treatments from all conceivable treatments, as opposed to comparing pre-defined sequences. The included studies highlighted the variation in the application of this common assumption and the data selected to supplement its use in practice.

Pivotal clinical trials or observational studies were frequently available for single treatments used at the specific decision-point of interest within the treatment sequence, generally the first treatment being modelled. This enabled the modellers to use evidence matching the line of therapy and the number of previous treatments, but not the specific sequence or specific previous treatments used. All but one of the included oncology studies used this simplifying assumption (***coded as NPT***). This was generally applied in conjunction with an assumption of treatment independence (*using codes NPT and IP*) for treatments used later in the sequence. Rheumatology studies also tended to use both assumptions in conjunction.

An alternative approach to assuming treatment independence for treatments used after the decision point was to use data matching each line of therapy or patient population (*coded as NPT only*). However, this was still hampered by the limited evidence base. An example of this is provided by a study (Heeg, 2015) which aimed to develop an analytic framework for comparing overall survival of different treatment sequences for multiple myeloma. The model required data for three treatment lines (Chapter 7, Section 7.3.3.3). The study obtained treatment-specific response rates (complete, partial, or non-response) for first-line treatments from a network meta-analysis of newly diagnosed patients. The response rates for subsequent treatment lines were obtained from a network meta-analysis of trials recruiting patients with relapsed or refractory disease.¹⁹⁷ Only treatments for which RCTs were available were considered in the model, so a number of treatments used in clinical practice could not be included. Only three RCTs, investigating three treatments, were available for relapsed/refractory disease, which were considered as representing third-line treatment. There were no RCTs of second-line treatment, and therefore the response rates for these were based on the results of the network meta-analysis used for third-line (for relapsed/refractory disease) combined with subgroup analysis on the treatment effect of second-line patients versus later-line patients, reported in the three RCTs. Treatment-switching in the model was assumed to be dependent on type of response and line of treatment, but independent of the specific treatment used. Data on duration of response, assumed to be the same across the different treatments, was taken from the bortezomib treatment arm in three clinical trials selected to match the relevant patient population.

Another study (Greenhalgh, 2005) evaluating the optimum positioning of electroconvulsive therapy (ECT), as first-, second-, or third-line treatment for major depressive disorder assumed that first and second-line treatments would have the same success rate (*coded as IP*); patients receiving third-line treatment were assumed to be treatment-resistant, and evidence sources matching this patient population were used (*representing NPT*).²⁴⁰ However, third-line treatment here could be interpreted as comparable to 'best supportive care' in other modelling studies, which would not have been coded. This highlights the challenge of coding the included studies based on the assumptions used at different parts of a sequence. A further example of the differential use of simplifying assumptions is provided by a study that reported using separate meta-analyses by line of therapy, but in fact also used the assumption of treatment independence, the NICE clinical guidelines study (NICE CG152), which evaluated treatment sequences for Crohn's disease.²⁵⁸ They obtained data on treatment effects for individual treatment regimens from two separate network meta-analyses, one of monotherapies used as first-line treatment and another of combination therapies used as second-line treatment. The study compared nine pre-defined sequences of up to four treatment lines, some of which included the monotherapy glucocorticosteroid as both second- and third-line treatment, and two combination therapies as third- and fourth-line treatment. The treatment sequences also included a generic biologic treatment used as second-, third- and fourth-line treatment. The same effect estimate was used for each treatment irrespective of their positioning, or specific previous treatments used.²⁵⁸

An important part of using simplifying assumptions in decision-analytic modelling relates to assessing their impact in sensitivity analyses (Section 6.2.1). Five studies evaluated the impact of using the assumption of treatment independence using sensitivity scenario analysis. Alternative data, to inform the sequencing effects, were based on the assumption that treatment effect was reduced in line with a reduction factor (**coded as RF**) in two studies,^{229 264} and the assumption that treatment effects decreased by the same set amount with each successive treatment (**coded as TD**) in three studies.^{115 257 258} However, one study only applied the alternative assumption coded as TD to one treatment, glucocorticosteroid following budesonide failure, in the study of treatments for Crohn's disease (NICE CG152).²⁵⁸ One further study used expert consensus (**coded as EXC**) to provide an estimate of sequencing effects.²¹⁶ These alternative assumptions are described in more detail in the corresponding sections below.

6.5.3.2 Substitution with another treatment effect

Another frequently-used simplifying assumption was that treatment effects were the same as an alternative, substitute treatment, generally taken from the same class or category of treatments. This incorporated the assumption that the treatment effect was equivalent to an alternative treatment used at a relevant position in the sequence (**coded as coded as PGE**) or was independent of positioning in the sequence (**coded as GE**). Seventeen (26%) studies used a substitute treatment effect matching the relevant positioning in the sequence (*coded as PGE*) and sixteen studies (24%) used a substitute treatment effect that was independent of positioning in the sequence (*coded and GE*). Most studies used a substitute treatment effect from another single named treatment. However, three studies of rheumatoid arthritis (Schipper, 2011;²⁶⁵ Diamantopoulos, 2012;²³⁴ Diamantopoulos, 2014²³⁵) applied a generic effect for TNF-inhibitors as a substitute treatment effect for a specific TNF-inhibitor, for example etanercept, and one study (Brennan, 2004)²²⁴ used a generic effect for conventional DMARDs as a substitute effect for gold. The numbers also include six further rheumatology studies (Brennan, 2007;²²⁵ Finckh, 2009;²³⁸ Kobelt, 2011²⁴⁸; Lindgren, 2009;²⁵¹ Tran-Duy, 2011;²⁷² Tran-Duy, 2014²⁷³) that modelled treatments within a sequence as a class (non-TNF-inhibitor biological agents, TNF-inhibitors, conventional DMARDs, or NSAIDs) rather than individual treatments. The treatment sequences modelled by eight studies (Brennan, 2007; Diamantopoulos, 2012; Diamantopoulos, 2014; Finckh, 2009; Kobelt, 2011; Lindgren, 2009; Schipper, 2011; Tran-Duy, 2014) of rheumatoid arthritis are illustrated in Table 6.4 and discussed further in Section 6.6.

This is an assumption of exchangeability of treatment effects within the same group or class. It should be noted, however, that class effects were not always used consistently. One study (Russel, 2009) investigating sequencing of biological agents for the treatment of rheumatoid arthritis, used data from an RCT of a non TNF-inhibitor, abatacept, to represent the treatment effect a TNF-inhibitor used as the second or third-line biologic (**coded as ST**),²⁶¹ whereas in other studies, for example Schipper, 2011 (Table 6.4)²⁶⁵ and Lindgren, 2009 (Table 6.4)²⁵¹, TNF-inhibitors were considered as a class separate from other biologics. I come back to this in Section 6.6.2.

6.5.3.3 Reduction of treatment effect

The simplifying assumption of treatment independence ignores any potential sequencing effects. Some studies used alternative assumptions in order to acknowledge a potential decrease in effect of treatments used later in the sequence. Nine studies (14%) applied the assumption that treatment effect is reduced in line with a multiplier or reduction factor when used at a later point (**coded as RF**).^{17 18 199 229 235 244 261 264 272} The use of a multiplier or reduction factor and the methods used to develop these are described in more detail in Chapter 5, Section 5.6, since they are considered as part of the meta-analytic approach used to produce estimates of clinical effectiveness for developing the parameters of the decision-analytic model. However, only two studies (Connock, 2006; Rodgers, 2011)^{17 199} reported the methods used to develop the multiplier or reduction factor; in the remaining studies it was merely noted that the figure was 'informed by the available evidence'. The reduction in treatment effects was applied as part of a sensitivity analysis in two studies (Schadlich, 2005; Clark, 2004), where the treatment-sequencing effects were ignored in the base case analysis.^{229 264} In one study (NICE GC137) the figure was varied in one-way sensitivity analyses.¹⁸ No other studies reported the conduct of sensitivity analyses to assess the impact of the reduction factor. However, Diamantpoulus *et al.*, noted in their discussion that 'sensitivity analyses showed that treatment response was not a major driver of the results'.²³⁴

Four (6%) studies (Fitzsimmons, 2014, NICE CG81, NICE CG152, Welsing, 2005) applied the assumption that the treatment effect would diminish by an arbitrary amount at each point in the sequence (**coded as TD**).^{115 257 258 275} This differs to the use of a 'reduction factor' (**coded as RF**), where the development of a specific multiplication factor based on relevant evidence is reported, because here all treatment effects are considered to decrease by the same or a generic proportion at each point in the sequence. Three of the studies (Fitzsimmons, 2014; NICE CG81; NICE CG152) applied this assumption within their sensitivity analyses (discussed in Section 6.5.3.1).^{115 257 258} One study (Fitzsimmons, 2014) evaluated the cost-effectiveness of using a stepped-care approach to treating sciatica.¹¹⁵ In that study the average treatment effect for each treatment category was derived from the network meta-analysis presented in Chapter 2; the potential reduction in effectiveness of treatments used during the later stages of the stepped approach was explored in sensitivity analysis using a relative reduction of 10%.¹¹⁵ A second study (NICE CG81) evaluated 17 chemotherapy sequences of up to three-lines of treatment.²⁵⁷ Whilst data for first-line treatments were obtained from a network meta-analysis, treatment effects for the two second-line treatments modelled were obtained from single studies, with the same effect estimate used to represent these drugs used as 3rd-line treatment in the base-case analysis. These estimates were reduced by a third in the sensitivity analysis, to represent their use as third-line treatment. The population of interest entering the model in a third study (Welsing, 2005) included patients with rheumatoid arthritis who were eligible for treatment with a first TNF-inhibitor, having had an inadequate response to at least two conventional DMARDs including methotrexate.²⁷⁵ The treatment effect of leflunomide, the first drug in the 'control' sequence, representing current practice, was derived from an RCT of leflunomide versus methotrexate in patients who had not previously used methotrexate. Patients entering the model were

assumed to have already failed methotrexate, and the effect estimate for leflunomide was therefore arbitrarily reduced by 25% in order to match the patient population indication of the economic evaluation. However, leflunomide was also the second treatment in the 'intervention' sequence, after inadequate response to a TNF-inhibitor (coded as DI, reported under Section 6.5.3.5). The same treatment effect was used to represent both indications. In the fourth study (NICE CG152), which included an evaluation of nine pre-defined sequences of up to three treatment lines for Crohn's disease, the treatment effect of glycocorticosteroid following budesonide failure was adjusted in sensitivity analyses.²⁵⁸ This study is also described in section 6.5.3.1 under treatment independence. The data for first-line treatments were obtained from a network meta-analysis of monotherapies used as first line. However, glycocorticosteroid was also used as a second-line monotherapy in some of the included sequences, after the failure of budesonide, sulfasalazine, or mesalazine. The guideline development group reasoned that glycocorticosteroid would be less effective when used after budesonide failure, and that it would be appropriate to multiply the efficacy by an adjustment factor between 0-1. The probability of remission used in the base case analysis was 75%, and values used in the sensitivity analysis ranged from 50-100%. However, the probability of remission with glycocorticosteroid, obtained from the network meta-analysis of first-line treatment, was 66%.

Four (6%) studies assumed a specific reduction in treatment efficacy with disease duration (*coded as RDD*). This included one non-rheumatology study (Fitzsimmons, 2014) that investigated the cost-effectiveness of treatment sequences for sciatica.¹¹⁵ The assumption was used in their sensitivity analyses. In the base case analysis, it was assumed that there was no reduction in utility as a result of previous unsuccessful interventions. However, in clinical practice the likelihood of improvement is prone to diminish with time alone, as well as the failure of each successive treatment. The potential decreased effects with successive treatments were explored using the diminishing effects assumption described earlier. Sensitivity analyses were also used to explore the possibility that the effect of an individual treatment may depend more on how early it is employed after the onset of sciatica than on the specific treatment used. Sensitivity analysis therefore incorporated decreasing the utility achieved with resolution of symptoms following the failure of prior treatments by an arbitrary estimate of 25%. A separate arbitrary estimate of 5-10% was also used to represent the proportion of patients being non-responders at each stage of the pathway (*coded as TD*).

The remaining three studies coded as RDD, all of which used the Birmingham rheumatoid arthritis model (BRAM), show that the distinction between the use of simplifying assumptions regarding the decreasing effect of disease duration (or time alone) and that of the number (or type of) previous treatments failed is not straightforward,^{227 229 254} which was also discussed in Chapter 5, Section 5.5 (the use of meta-regression). The positioning of a treatment later in the sequence can represent both, and it could be argued that these studies should have been coded as TD, representing decreasing effect. One of these studies (Malottki, 2011)²⁵⁴ cited and used a similar approach to a prior study (Chen, 2006),²²⁷ whose authors reported being unable to find data to support the quantification of a reduction in effectiveness with disease duration.²²⁷ Treatment effects of conventional DMARDs,

obtained from trials of early disease, were therefore halved to represent their later use, and the assumption assessed using scenario analysis. In the third study (Clark, 2004) the treatment effects of three conventional DMARDs, taken from selected studies of early disease, were each increased by 0.125 points on the Health Assessment Questionnaire Index (HAQ) as part of the sensitivity analyses to account for the potential decrease in effect when conventional DMARDs are used later in the sequence.²²⁹ The model also assumed that an individual's HAQ score increased gradually in steps of 0.125 over time, representing progressive disease. This rate was reported to have been chosen to reflect the empirically observed increase reported by a separate study based on a database of leflunomide clinical trials.³⁵³

The fact that treatment effect is likely to diminish with time was also acknowledged in some rheumatology studies, but without using an explicit assumption. For example, one study (Finckh, 2009) which compared treatment sequences representing early versus late introduction of TNF-inhibitors (Table 6.4) used data from RCTs of early disease to inform initial treatment effects.²³⁸ It was acknowledged that the ability to induce remission or achieve a good or moderate response was likely to decline over time, and therefore, the probabilities for treatment response six months after therapy were estimated using a multivariate relationship based on a number of covariates including, among others, disease duration, baseline HAQ, and number of previous DMARDs. This was based on the regression analysis of a patient registry conducted by Wailoo *et al.*²⁷⁴ The study by Chen *et al.*, described earlier, provides another example of this.²²⁷ Here the treatment effects for TNF-inhibitors added to a sequence of conventional DMARDs, as either the first, second, or third drug, were derived from two RCT data sets, one based on a population with 'early disease' and the other 'late disease'. Data for the TNF-inhibitors modelled as the first drug were based on the dataset for early disease, representing patients who were naïve to methotrexate. Both datasets, for early and late disease, were used for TNF-inhibitors introduced as the third drug in the sequence, and the data set for late disease only were used for TNF-inhibitors added to the end of the sequence. The findings of scenario analyses showed that the results were sensitive to these assumptions; less favourable results were obtained for TNF-inhibitors using the 'late disease' data set for the third drug, and results obtained for TNF using the 'early disease' data set for the first drug were less favourable than those obtained when using the 'early disease' data set for third-line use. The study (Chen, 2006) also evaluated adding a further 2-3 subsequent TNF-inhibitors to the TNF used as the third drug, using the same treatment effect as the first TNF, which was not investigated in sensitivity analysis.²²⁷

6.5.3.4 Impact of time since previous treatment

Two studies assumed that treatment effects were not affected by previous treatments when patients had been in long term *remission* (**Coded as LR**). One (Soini, 2012) investigated treatment sequences for advanced cancer,²⁶⁹ and one (Frankum, 2005) evaluated treatments for onychomycosis.²³⁹ The treatments which achieved the remission could be re-used as subsequent treatments; in other words, treatment sequencing was not considered an issue. Neither study assessed the impact of this using sensitivity analysis.

6.5.3.5 Displacement effect ignored

All five non-rheumatology studies that investigated adding a new treatment to an established sequence, ignored the potential effect of treatment displacement (**Coded as DI**). It was also ignored in 23 rheumatology studies that investigated adding a new treatment to an established sequence. It was not applicable to one study that investigated adding a second TNF-inhibitor to the end of the treatment sequence for psoriatic arthritis,¹⁷ as this did not displace anything.

6.5.3.6 The use of uncontrolled trials or observational studies without bias adjustment

Due to the limited evidence base, a number of studies used estimates of treatment effects derived from observational studies or uncontrolled trials. None of these studies used any methods to adjust the treatment effects to account for potential bias in the data source. These studies can be interpreted as having used the simplifying assumption that the observational studies or uncontrolled trials provided an unbiased estimate of treatment effect (**Coded as UOBS**).

Four (6%) further studies used expert consensus to inform the treatment effects of interventions used primarily towards the end of treatment sequences (**Coded as EXC**). None accounted for any potential limitation within these estimates. However, one study (Albert, 2000) used the data from expert consensus on sequencing effects as part of a scenario analysis, whilst assuming that treatment effects were independent of other treatments in the base case analysis.²¹⁶ The study compared four pre-defined sequences of three conventional DMARDs. The authors noted that their literature review did not identify any data for the composite effectiveness of whole sequences, or estimates of the effectiveness of each drug used at each specific point along the sequence. The authors therefore compared the use of different estimates of effectiveness obtained from three different sources, including expert rheumatologists, in order to assess the assumption that treatment effects were independent of other treatments in the sequence (Appendix Volume II, Appendix E)

6.6 AN ASSESSMENT OF THE VARIATION IN HOW THE DATA ON TREATMENT SEQUENCES WAS SELECTED ACROSS MODELLING STUDIES EVALUATING A SIMILAR DECISION PROBLEM IN RHEUMATOID ARTHRITIS

The choice of data should always be based on the best available evidence. However, the type of data chosen within the included studies varied substantially, even for the comparison of similar treatment options within a specific point in a treatment pathway. The impact of a limited evidence base, and how the data on treatment-sequencing effects are obtained to develop the decision-analytic model parameters are explored here in more detail using 22 studies that considered the sequential use of TNF-inhibitors when investigating treatments sequences for rheumatoid arthritis. An overview of the treatments for rheumatoid arthritis and the available evidence base is provided in Appendix Volume I, Section C3. The sequential use of TNF-inhibitors was identified as representing an important RCT evidence gap (Chapter 5, Section 5.4).

6.6.1 Modelling studies that evaluated sequential TNF-inhibitors

The decision problem and treatment sequences evaluated by the included modelling studies which considered sequential TNF-inhibitors are presented in Table 6.4. Twenty-one studies modelled the sequential use of TNF-inhibitors, whilst one study (Malottki, 2011) modelled a second-line TNF-inhibitor as the first treatment. Studies that considered the sequential use of TNF-inhibitors evaluated the introduction or use of biological agents at different points in the treatment pathway. Five studies investigated the introduction of biological agents in early disease, representing a patient population who were either DMARD-naïve or only failed to respond to one previous conventional DMARD (Table 6.4, Section A).^{227 232 238 248 265} Five studies evaluated the introduction of biological agents in established disease, after an inadequate response to at least two conventional DMARDs, one of which was usually methotrexate (Table 6.4, Section B).^{234 235 261 273 274} Twelve studies investigated the use of biological agents in a patient population who have had an inadequate response to previous TNF-inhibitors (Table 6.4, Section C).^{222 223 225 228 229 241 246 251 254 256 260 262}

Table 6.4: The decision problem and treatment sequences evaluated by included modelling studies that considered sequential TNF-inhibitors

AUTHOR, YEAR COUNTRY	DECISION PROBLEM	TREATMENT SEQUENCES
A. Studies that evaluated the introduction of biological agents in early disease		
Chen, 2006 UK (NICE TA130)	TNF-inhibitor added to a sequence of DMARDs as the 1st, 3rd or last active drug (early vs late introduction of the TNF-inhibitor); also investigated adding a further 1-2 consecutive TNF inhibitors to the TNF used as the 3rd active drug.	MTX - SSZ - SSZ+MTX - [SSZ+HCQ+MTX -] LEF - Gold -AZA - CyC - CyC+MTX - PEN TNF - MTX - SSZ - SSZ+MTX - SSZ+HCQ+MTX - LEF - gold - AZA - CyC - CyC+MTX - PEN MTX - SSZ - SSZ+MTX - TNF I-III - LEF - gold - AZA - CyC - CyC+MTX - PEN MTX - SSZ - SSZ+MTX - SSZ+HCQ+MTX - LEF - gold - AZA - CyC - CyC+MTX - PEN- TNF <i>When TNF(s) added as the third drug it replaced SSZ+HCQ+MTX</i>
Davies, 2009 US	Comparing sequences starting with 1-2 consecutive TNF-inhibitors vs conventional DMARDs	MTX - MTX+HCQ - LEF - gold TNF I - MTX+HCQ - LEF - gold TNF I - TNF II - MTX+HCQ - LEF - gold
Finckh, 2009 US	Early vs late introduction of TNF-inhibitors	NSAIDs - 3 DMARDs - 3 TNFs 3 DMARDs - 3 TNFs 3 TNFs - 3 DMARDs <i>Each treatment represented three consecutive treatments</i>
Kobelt, 2011 Sweden	Comparing sequences starting with a TNF-inhibitor vs a conventional DMARD	MTX - standard DMARD therapy ETA - standard DMARD therapy MTX - 1st biologic - standard DMARD therapy ETA - 2nd biologic - standard DMARD therapy
Schipper, 2011 Netherlands	Early vs late introduction of biological agents including TNF-inhibitors.	MTX - MTX+LEF - TNF I - TNF II - RTX MTX+LEF - TNF I - TNF II - RTX - MTX TNF I - TNF II - RTX - MTX - MTX+LEF
B. Studies that evaluated the introduction of biological agents in established disease		
Diamantpoulus, 2012 Italy	Adding a new biologic (non-TNF-inhibitor) to a sequence of biologics, and comparing biologics (TNF-inhibitor vs non-TNF-inhibitor) as the first drug	TOC - ETA - ADA - RTX - ABA TOC - ADA - RTX - ABA ETA - ADA - RTX - ABA TOC - ETA - RTX - ABA ADA - ETA - RTX - ABA INF - ETA - RTX - ABA
Diamantpoulus, 2014 UK	Adding a new biologic (non-TNF-inhibitor), as first or second drug, to two different sequences of biological agents, depending on patient's tolerance to MTX	<i>For MTX contraindicated population:</i> CZP - ETA - ADA TOC - CZP - ETA - ADA CZP - TOC - ETA - ADA <i>For MTX tolerant population:</i> CZP - RTX - ETA - ADA - ABA - ADA - INF TOC - CZP - RTX - ETA - ADA - ABA - ADA - INF CZP - TOC - RTX - ETA - ADA - ABA - ADA - INF
Russell, 2009 Canada	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ETA - INF - ADA ABA - ETA - INF ETA - ABA - INF

Tran-Duy, 2014 Netherlands	Comparing sequences with and without biological agents. TNF-inhibitors randomly chosen from ETA, ADA, INF, GOL, CZP; and the 2 non-TNF-inhibitor biologics from RTX, ABA, TOC	MTX – SSZ/LEF AZA - CyC - CYC - HCQ - gold MTX – SSZ/LEF – TNF I - TNF II - non-TNF I - non-TNF II - AZA - CyC - CYC - HCQ – gold
Wailoo, 2006 US	Initial analysis compared 4 biologics (followed by cDMARDs). Subsequent analysis included treatment sequencing, which compared the use of 2nd or 3rd TNF-inhibitor with a single TNF (INF) followed by cDMARDs	[ETA, ADA, ANA, INF] - cDMARDs INF - ETA - cDMARDs INF - ADA - cDMARDs ETA - ADA - cDMARDs ADA - INF - cDMARDs INF - ETA - ADA - cDMARDs INF - ADA - ETA - cDMARDs ETA - ADA - INF - cDMARDs ETA - INF - ADA - cDMARDs
C. Studies that evaluated biological agents in a patient population who have had an inadequate response to a previous TNF-inhibitor		
Beresniak, 2011 Spain	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ETA - ABA - ADA ETA - RTX - ADA ETA - ADA - ABA ETA - ADA - INF
Beresniak, 2013 Germany	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ADA - ABT - ETA ADA - RTX - ETA ADA - ETA - ABT ADA - ETA - INF
Brennan, 2007 UK (a submission for NICE TA36)	The primary aim of the study was to compare the use of a first TNF-inhibitor with conventional DMARDs, but a secondary aim was to investigate sequential TNF-inhibitors, which is why this study is included here.	(≥2 cDMARDs) - cDMARDs (≥2 cDMARDs) - TNF I - cDMARDs (≥2 cDMARDs) - TNF I - TNF II - cDMARDs <i>Rather than specifying particular cDMARDs at different positions, a generalised DMARD was used. The number of treatments included in the sequences were not stated, but it was assumed that after the 6th treatment patients would no longer respond but still receive some maintenance therapy on cDMARDs.</i>
Cimmino 2011 Italy	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ETA - ABA - ADA ETA - RTX - ADA ETA - ADA - ABA ETA - ADA - INF
Clark, 2004 UK (NICE TA72)	adding a new biologic (a non-TNF-inhibitor) to two sequences containing 1-2 consecutive TNF-inhibitors. The new biologic was added after TNF-inhibitor(s) or as the last drug.	SSZ - MTX - LEF [- ETA] - INF [- ANA] - gold - AZA - CyC - CyC+MTX - PEN SSZ - MTX - LEF [- ETA] - INF - Gold - AZA - CyC - CyC+MTX - PEN [- ANA] SSZ - MTX - HCQ - Gold - LEF [- ETA] - INF [- ANA] - AZA - CyC - CyC/MTX - PEN SSZ - MTX - HCQ - Gold - LEF [- ETA] - INF - AZA - CyC - CyC+MTX - PEN [- ANA] <i>Both baseline sequences implemented with and without ETA, representing 1-2 consecutive TNFs.</i>
Hallinen, 2010 Finland	Adding biological agents to sequences, representing a gradual increase in the number of previous biological agents	(TNF) - gold - CyC - MTX (TNF) - [ADA, ETA, INF, RTX, or ABT] - gold - CyC - MTX (TNF) - RTX - [ADA, ETA, INF, or ABT] - gold - CyC - MTX (TNF) - RTX - INF - [ADA, ETA, or ABT] - gold - CyC - MTX <i>Assumed all patients entering model have had an IR to one TNF inhibitor. The sequence 'gold - CyC - MTX' was described as best supportive care..</i>
Kielhorn, 2008 UK	Adding a new biologic (a non-TNF-inhibitor) to two sequences, with and without 2 consecutive TNF-inhibitors.	(TNF) - LEF - gold - CyC - (MTX) (TNF) RTX - LEF - gold - CyC - (MTX) (TNF) - ADA - INF - LEF - Gold - CyC - (MTX) (TNF) RTX - ADA - INF - LEF - Gold - CyC - (MTX) <i>Assumed all patients entering model have had an IR to one TNF inhibitor. MTX was described as palliative treatment</i>
Lindgren, 2009 Sweden	Adding a new biologic (a non-TNF-inhibitor) to sequence of TNF-inhibitors.	(TNF I) - TNF II - TNF III - TNF IV (TNF I) - RTX - TNF II - TNF III
Malotki, 2011 UK (NICE TA195)	Adding a biological agent to sequence of cDMARDs, representing the comparison of second-line biological agents	(TNF) [ADA, ETA, INF, RTX, or ABT] - LEF - gold - CyA - AZA (TNF) - LEF - gold - CyA - AZA <i>Assumed all patients entering model have had an IR to one TNF inhibitor.</i>
Merkedal, 2010 Germany	Adding a new biologic, a non-TNF-inhibitor to a standard treatment sequence of 2 TNFs followed by 2 conventional DMARD.	(ETA) - ADA - INF - gold - CyC - BSC; (ETA) - RTX - ADA - INF - gold - CyC - BSC <i>Assumed all patients entering model have had an IR to ETA, a TNF inhibitor.</i>
Puolakka, 2012 Finland	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ADA - ABT - ETA ADA - RTX - ETA ETA - ABA - ADA ETA - RTX - ADA INF - ABT - ETA INF - RTX - ETA
Sarau, 2010 France	Comparing various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)	ETA - ABA - ADA ETA - RTX - ADA ETA - ADA - ABA ETA - ADA - INF

Abbreviations: Conventional DMARDs: AZA azathioprine; CyC cyclosporine / cyclosporin A; HCQ hydroxychloroquine; LEF leflunomide; MTX methotrexate; PEN D-penicillamine; SSZ sulfasalazine.

TNF-inhibitors (biological agents): ADA adalimumab; ETA etanercept; INF infliximab; GOL Golimumab, CZP Certolizumab pegol.

Biological agents targeting different proteins: ABA abatacept; ANA anakinra; RTX rituximab; TOC tocilizumab

Other abbreviations: DMARD disease-modifying anti-rheumatic drugs; cDMARD conventional disease-modifying anti-rheumatic drugs; TNF-inhibitor (or TNF) tumour necrosis factor α inhibitor; vs versus

6.6.2 The evidence used to inform sequencing effects of TNF-inhibitors

The main issue here is that there were no RCTs comparing the efficacy of successive first-generation TNF-inhibitors in patients with an inadequate response to a previous TNF-inhibitor (Appendix Volume I, Section C3). Four main data sources used are outlined below, the choice of which often impacted the data used for other treatments in the modelled sequences.

To avoid repetition, the phrase 'inadequate response to previous treatment' is used to represent both insufficient response and intolerance. For the same reason, no distinction is made between biological agents used as combination therapy or monotherapy. Biological agents are used as both add-on therapy to existing conventional DMARDs, usually methotrexate, and monotherapy, with the former generally used in established rheumatoid arthritis, and the latter in early disease (Appendix Volume I, Section C3).

6.6.2.1 Data obtained from patient registries

Five studies obtained the treatment effects of sequential TNF-inhibitors from observational studies of patient registries.^{225 248 251 265 273} All five used individual patient level data to inform their models. One study (Brennan, 2007) used patient registry data to obtain a generic effect estimate for TNF-inhibitors irrespective of their positioning, whilst three studies (Schipper, 2011, Tran-Duy, 2014, Kobelt, 2011, Lindgren, 2009) used the data to obtain generic estimates that reflected the line of treatment (discussed in Section 6.5.3.2). In the fifth study (Tran-Duy, 2014), because the number of observations for some drugs were too small and the effectiveness of those with more observations were found to be similar, specific pairs of TNF-inhibitors were grouped; the data were sampled distinguishing between first- and second-line use.

Studies using observational data to inform line specific effects for TNF-inhibitors:

Two of the four studies that used patient registry data to obtain line-specific effect-estimates for TNF-inhibitors (Schipper, 2011 and Tran-Duy, 2014) used data from the Dutch Rheumatoid Arthritis Monitoring (DREAM) biologic registry.^{265 273} Both studies assumed the effectiveness of second-line TNF-inhibitors was independent of specific previous TNF-inhibitors. The DREAM registry includes patients with established disease starting their first TNF-inhibitor. One study (Schipper, 2011) investigated the introduction of TNF-inhibitors in early rheumatoid arthritis (Table 6.4, Section A) and, in order to implement the registry based data, assumed that the treatment effects of TNF-inhibitors in DMARD-naïve patients were comparable to those in the registry who had failed at least two conventional DMARDs.²⁶⁵ The impact of this was assessed using scenario analysis, implemented using an estimated response rate of 30% taken from an RCT, instead of the 20% taken from the registry based on clinical practice data. Two studies (Kobelt, 2011 and Lindgren, 2009) used data

from the Southern Swedish Arthritis Treatment Group (SSATG) registry, which also included patients with established rheumatoid arthritis.^{248 251} One of these (Kobelt, 2011) investigated the introduction of TNF-inhibitors in early disease.²⁴⁸ Individual patient data for the first treatments modelled, methotrexate or etanercept, were therefore obtained from an RCT of early disease, which included patients who were methotrexate-naive. The treatment effects of subsequent treatments, either a first or second TNF-inhibitor, were obtained from the patient registry. The second study (Lindgren, 2009) compared two sequences consisting of sequential TNF-inhibitors, with and without the addition of a new biological agent rituximab (a non-TNF-inhibitor) at the start of the sequence.²⁵¹ Patients entering the model were assumed to have had an inadequate response to their first TNF-inhibitor (Table 6.4, Section C). The model differentiated between TNF-inhibitors used as different treatment lines, but not the specific agent. Patients who discontinued rituximab were assumed to go back to TNF-inhibitors immediately, using event data for second-line TNF-inhibitors, based on the rationale that rituximab should not influence the magnitude of the effect of the next TNF-inhibitor. The register provided data for up to three lines of TNF-inhibitor, and the treatment effect of the fourth TNF-inhibitor was assumed to be the same as the third. The effect of this assumption was considered minimal as the average number of treatments resulting from the model was 2.6 in the TNF-inhibitor arm and 2.3 in the rituximab arm.²⁵¹ The data were analysed using Cox regression with treatment line as one of the covariates.

Observational data used to obtain a generic effect for TNF-inhibitors:

One study (Brennan, 2007) used the alternative assumption that response to a second TNF-inhibitor was independent of the response to a first.²²⁵ The study compared the use of TNF-inhibitors as a class, with ongoing conventional DMARD therapy, in patients who had failed at least two conventional DMARDs (Table 6.4, Section C).²²⁵ The impact of a second TNF-inhibitor was only investigated as part of further analysis. A computer simulation model was used to synthesise evidence based on individual patient data from the British Society for Rheumatology Biologics Registry (BSRBR) database with other available sources from systematic reviews (*reference given for data source was Barton, 2004²¹⁹*). At the time of the analysis, the register followed 7,083 patients with active rheumatoid arthritis treated with a TNF-inhibitor, and 870 patients treated with conventional DMARDs over three years. There were too few patients in the control group to restrict the analysis to patients starting a new conventional DMARD so the analysis included all patients, adjusting for the number of previous DMARDs used as well as other key parameters. The results showed a higher likelihood of response with fewer previous DMARDs. The decision model was conducted over the patients' lifetime, with observational data from a Swedish cohort used to extrapolate estimates from BSRBR to a longer time horizon. A generalised weighted average effect estimate was used for the initial TNF-inhibitor and conventional DMARDs, irrespective of the specific drugs used. The assumption that response to a second TNF-inhibitor was independent of the response to a first (Section 6.5.3.1), was then used for the second TNF-inhibitor, based on the absence of evidence on any correlation. Subgroup analyses were also reported according to number of previous conventional DMARDs used (<2, <3, <4, <5, 5+).

6.6.2.2 Data obtained from RCTs of first-line TNF-inhibitors

Nine studies used data obtained from RCTs of TNF-inhibitors used as the first-line biological agent.²²⁷

229 232 234 235 238 241 246 256

One study (Diamantpoulous, 2012) applied a reduction to the treatment effects obtained from the RCTs in order to represent the TNF-inhibitors' use as second-line (Section 6.5.3.3).²³⁴

Diamantpoulous *et al.*, investigated incorporating a new biological agent, tocilizumab, to an established sequence of four biological agents starting with a sequence of two TNF-inhibitors in a patient population with inadequate responses to conventional DMARDs (Table 6.4, Section B). Tocilizumab was either added to the start of a sequence representing standard care, or replaced the initial TNF-inhibitor, which could be one of three agents. The baseline sequence included etanercept followed by adalimumab, and in the sequences where tocilizumab replaced the first drug, the second was either adalimumab or etanercept. Response rates for first-line etanercept or tocilizumab were obtained from a published network meta-analysis of placebo-controlled trials of six commonly used biological agents. Because of the lack of evidence about the efficacy of TNF-inhibitors after etanercept or tocilizumab, response rates for adalimumab were reduced by 30% to correspond to its second-line position in the sequence representing standard care (Chapter 5, Section 5.6.2). However, etanercept was also considered as the second TNF-inhibitor in alternative sequences and it was not stated whether the effect of etanercept was reduced here.

The other eight studies (Clark, 2004; Davies, 2009; Chen, 2006; Hallinen, 2010; Diamantopoulos, 2014; Kielhorn, 2008; Merkesdal, 2010; Finckh, 2009) all assumed that the treatment effect of a second TNF-inhibitor was the same as its use in the first-line setting (Section 6.5.3.1).^{227 229 232 235 238 241 246 256} Only one study (Clark, 2004) investigated the potential impact of using the assumption that the treatment effects were independent of positioning.²²⁹ Clark *et al.* compared predefined sequences made up of both conventional DMARDs and TNF-inhibitors, with and without the addition of anakinra, a non TNF-inhibitor, at different points in the sequences, middle, late, and last (Table 6.4, Section C). The sequences incorporated either one or two consecutive TNF-inhibitors, which included etanercept followed by infliximab. Anakinra was added after the TNF-inhibitors as either the second or third-line biologic, or alternatively as the last active drug in the sequence. The authors noted that the effect of removing etanercept was that patients reached the divergent point earlier, and therefore tended to be younger. Therefore, although the base case analysis ignored sequencing effects, sensitivity analyses included replacing the treatment effects used for the two TNF-inhibitors with a higher score for the first biologic, etanercept, and a lower score for the second biologic, infliximab (discussed in Chapter 5, Section 5.6.2.3). Both had the same HAQ score in the base-case analysis. (This study was based on the Birmingham Rheumatoid Arthritis Model (BRAM) discussed in Chapter 7, Section 7.3.6.4.)

One of the studies (Davies, 2009) using data obtained from RCTs of first-line TNF-inhibitors for second-line TNF-inhibitors, interestingly, obtained the data to inform the effects of subsequent conventional DMARDs from a follow-up study based on SSATG registry, as it included patients with

established rheumatoid arthritis.²³² The study compared three TNF-inhibitors versus methotrexate as the first drug in a predefined sequence for early rheumatoid arthritis (Table 6.4, Section A). The treatment effects of each initial TNF-inhibitor were based on an RCT that included a methotrexate-naïve patient population. A second TNF-inhibitor (etanercept) was then added to the most cost-effective initial TNF-inhibitor (adalimumab) before switching to subsequent conventional DMARDs. It was assumed that etanercept efficacy in adalimumab-treated patients would be equivalent to etanercept efficacy in TNF-inhibitor-naïve patients.

6.6.2.3 Data obtained from uncontrolled open-label studies

An alternative data source used for the treatment effects of sequential TNF-inhibitors was the ReACT (Research in Active Rheumatoid Arthritis) study (Appendix Volume I, Section C3), a large, uncontrolled, open-label study that aimed to evaluate the effectiveness and safety of adalimumab in patients who had previously discontinued TNF-inhibitors in clinical practice.³⁵⁴ The study enrolled participants with active rheumatoid arthritis who had previously been treated with conventional DMARDs or TNF-inhibitors. Of the 6610 included patients, 899 (14%) had a history of etanercept and or infliximab therapy.

The ReACT study was used by six included modelling studies (Cimmino, 2011; Saraux, 2010; Puolakka, 2012; Beresniak, 2011; Beresniak, 2013; Malotki, 2011) that evaluated the use of biological agents in a patient population with inadequate response to a previous TNF-inhibitor (Table 6.4, Section C) (Section 6.5.3.1).^{222 223 228 254 260 262} One study (Malotki, 2011) compared three different TNF-inhibitors used as second line.²⁵⁴ It was assumed that all patients entering the model had an inadequate response to their first TNF-inhibitor. Data on sequencing effects for two TNF-inhibitors were obtained from uncontrolled open-label studies. Treatment effects of second line adalimumab were obtained from the ReACT study,³⁵⁴ whilst the data for etanercept were obtained from a study (Bringham, 2009)³⁵⁵ of patients treated with etanercept after failing infliximab (Appendix Volume I, Section C3). The treatment effect of infliximab, for which there were no studies, was assumed to be the same as etanercept (Section 6.5.3.2). This study was based on the Birmingham Rheumatoid Arthritis Model (BRAM) discussed in Chapter 7, Section 7.3.6.4.

The remaining five modelling studies investigated different fixed sequences of three biological agents using the same evidence base and modelling approach, known as the advanced simulation model (Chapter 7, Section 7.3.2.3).^{222 223 228 260 262} The fixed sequences were used to compare the cost-effectiveness of a TNF-inhibitor versus an alternative biologic as either the second- or third-line biological agent. The treatment sequences included specific drugs, presented in Table 6.4 (Section C) under the following authors: Beresniak, 2011,²²² Beresniak, 2013,²²³ Cimmino 2011,²²⁸ Puolakka, 2012,²⁶⁰ and Saraux, 2010.²⁶² A 100% failure rate was assumed for the first TNF-inhibitor, generally etanercept or adalimumab, in all sequences. Adalimumab was used as a second TNF-inhibitor in three studies,^{222 228 262} and etanercept in one.²²³ However, the same treatment effect, taken from the ReAct study, was used for both TNF-inhibitors as second-line therapy. Although all the studies

modelled a third TNF-inhibitor as part of the fixed sequences, which included adalimumab,^{222 228 260 262} infliximab,^{222 223 228 260 262} or etanercept,^{223 260} only infliximab was used as the comparator for the alternative biologic in sequences investigating a non-TNF-inhibitor as the third biologic (Table 6.4, Section C). A 100% inadequate response rate was assumed for both first and second TNF-inhibitors, etanercept and adalimumab. In the absence of published effects data for infliximab used after an insufficient response to two TNF-inhibitors, the results of the ReACT study were used as surrogate evidence (Section 6.5.3.2).

6.6.2.4 Data obtained from RCTs of ‘new’ biological agents that include a patient population with inadequate response to TNF-inhibitors

The advanced simulation modelling approach was also used by Russell *et al.*, who assessed the use of abatacept, a non-TNF-inhibitor, as an alternative first-line biologic in patients with an inadequate response to conventional DMARDs, and a second-line biologic in patients with an inadequate response to a first TNF-inhibitor (Table 6.4, Section B) (discussed in Section 6.5.3.2).²⁶¹ Modelling was based on fixed sequences of three biological agents. Data on treatment effects were taken from three pivotal trials. Two RCTs were of abatacept, one of which included a patient population with inadequate response to conventional DMARDs who had not previously received a TNF-inhibitor, and the other included participants with an inadequate response to TNF-inhibitors. The third RCT was of etanercept in a patient population with an inadequate response to conventional DMARDs. In the absence of relevant controlled trials, the treatment effect for etanercept used as both second- and third-line biologic was also taken from the RCT of abatacept in patients with an inadequate response to previous TNF-inhibitors using the assumption of a 10% reduction in effectiveness after each switch (Section 6.5.3.3). The assumption was based on observational studies reporting that switching TNF-inhibitors is often associated with lower efficacy.

6.7 DISCUSSION

6.7.1 Summary of findings

Decision-analytic models can be used to evaluate treatment sequences, where no sequencing clinical trials exist. Treatment sequences can be modelled as a series of individual treatments, but each treatment requires a specific treatment effect estimate that is conditional on its positioning in the sequence, or the previous treatments used. The scarcity of data to inform such estimates means that simplifying assumptions are used in conjunction with the available data on discrete treatments. The review of modelling studies identified a range of simplifying assumptions used to represent treatment-sequencing effect-estimate.

A coding scheme was developed as part of the data extraction process in order to group the studies according to the simplifying assumptions used. This coding system provides a useful addition to the framework presented in Chapter 8, in that it could help clarify to policy makers, modellers, and reviewers what actually has been done in previous or completed modelling studies.

Four issues were identified, which made it challenging to apply the coding scheme to the included studies:

- i. The differential application, within the same study, of multiple simplifying assumptions across different treatments, different points in the sequences, and in the use of scenario analysis.
- ii. The limited information provided on the simplifying assumptions used, especially for those applied to treatments used subsequent to the decision point of interest. In some instances, the simplifying assumptions reported did not account for all the compromises made due to the limited evidence base.
- iii. Studies generally did not distinguish between the potential decrease in treatment effect due to time alone (or disease duration) and that of the number and type of previous treatments used.
- iv. Variability in the way that treatments representing palliative or best supportive care were modelled and reported.

The findings of the review demonstrated that priority was often given to matching the evidence for the decision point, rather than considering treatment sequences as a whole. Sequencing trials were rarely available, and the uncertainty in the quality of the alternative evidence to inform the sequencing effects was not investigated in depth. Only five (10%) studies using the most commonly applied assumption of treatment independence evaluated its impact in sensitivity analyses, by reducing the effect of treatments used later in the sequence using a factor based on evidence, an arbitrary amount, or expert consensus. The assumption that treatment effect is dependent on line of therapy was frequently used in conjunction with the assumption of treatment independence, applied to treatments adopted later in the sequence.

The type of simplifying assumptions made is dependent on both the decision problem and the limitations or type of available evidence. The type of data selected for the discrete treatments used to inform treatment-sequencing effects varied considerably, even when considering similar decision problems. For example, data sources used for informing sequential TNF-inhibitors included observational studies of patient registries, RCTs of TNF-inhibitors used as first-line treatment, uncontrolled open-label studies, and RCTs of novel biological agents used after the failure of TNF-inhibitors (used as a substitute effect). None of the data sources were ideal, and they necessitated simplifying assumptions to be made in order to apply them. Patient registry data was frequently used, but even this did not provide estimates of treatment effects conditional on the failure of a specific previous treatment. The review findings also showed that the modelling technique chosen can impact the extent of the treatment sequences that need to be modelled, the type of data required to inform the sequencing effects, and the assumptions required. The coding scheme may provide a useful tool for appraising and comparing different approaches used to address similar treatment sequencing decision problems. A review of the modelling technique used is provided in Chapter 7.

6.7.2 Previous reviews

Two existing systematic reviews of economic evaluations investigating treatment sequences of biologic agents for rheumatoid arthritis (Tosh, 2014)²⁸ or psoriasis (Mauskopf, 2014)¹⁸⁵ were identified that also investigated the simplifying assumptions used. Mauskopf *et al.* aimed to analyse the assumptions about treatment sequencing after failure on a first-line biological agent in cost-effectiveness models of biological therapy for moderate to severe plaque psoriasis, and to compare them with the most recent treatment guidelines.¹⁸⁵ They identified five modelling studies, one of which investigated sequences of active treatments after the failure of a first-line biological agent. The authors concluded that cost-effectiveness models of first-line biological agents either do not include subsequent treatments or include only some of the regimes recommended in the current guidelines. They also concluded that the cost-effectiveness results may be sensitive to the assumptions about treatment sequencing and the choice and efficacy of subsequent treatment sequencing regimens. The second review, by Tosh *et al.*, aimed to assess and critique how sequential DMARDs have been modelled for economic evaluations of their use in rheumatoid arthritis.²⁸ They identified 25 studies modelling a sequence of treatments, but none that identified the optimum sequence of multiple treatments given a set of treatment options. They noted that the reporting of the methods and evidence used to assess the effect of downstream treatments in the sequences was generally poor; when lifelong models and downstream treatment sequences were considered, evidence gaps were identified. They concluded that methods were not consistently applied, leading to varied estimates of cost-effectiveness, and that treatment sequences were not fully considered and modelled, potentially resulting in inaccurate estimates of cost-effectiveness.

A more recent review, identified after completing the review of methods, by Zheng *et al.* aimed to provide practical guidance on conceptualising whether and how to model sequences in health economic models, by undertaking a review of the approaches used to model treatment sequences in published NICE technology appraisals.¹⁰⁴ The authors concluded that the biggest challenge to modelling treatment sequences is the scarcity of clinical data that capture long-term impacts of sequences on efficacy and safety. They identified three commonly used assumptions to bridge the evidence gap, but noted that each had its own limitations. These included the assumption that the efficacy of a treatment stayed unchanged regardless of line of therapy, the use of data from trials in different lines of therapy to directly model a treatment sequence, and the use of retrospective studies of clinical registries or databases. This review did not identify any new assumptions that were not covered by the coding scheme that I developed. In fact, my review provides a much more in-depth evolution of the simplifying assumptions used, and it also separates out the issue of the type of data selected to inform the efficacy of subsequent treatments.

6.7.3 Assessing the validity of assumptions relating to sequencing effects and their impact of structural uncertainty

The validity of the simplifying assumptions made regarding treatment sequencing in terms of representing reality was not generally investigated. However, the limited evidence base means that

there is generally no gold standard to use as a reference point. One potential approach might be to compare the summary treatment effects developed by making simplifying assumptions with the findings of a sequencing RCT but, as discussed in Chapter 4 and 5, these are limited in both their availability and extent of the treatment sequences considered. An alternative approach might be to emulate the sequencing trial using big data (real world data).^{327 356} I come back to this in Chapter 9, Section 9.4.6. The issue of assessing the validity of a simplifying assumption is also closely related to the assessment of the external validation of a treatment sequencing model, which I discuss in Chapter 7, Section 7.4.4. I also return to this issue in the main discussion for the thesis (Chapter 9).

The use of simplifying assumptions increases the uncertainty in estimating the effectiveness and cost-effectiveness of treatments, the extent of which should be explored as part of the analysis of structural uncertainty (discussed in Section 6.2.2). Structural uncertainty can have a greater impact upon the model results than parameter uncertainties, which are more frequently reported in health economics.³⁴⁶ Only a small proportion of the studies that made the simplifying assumption of treatment independence assessed its impact on the overall results using scenario analysis. However, although scenario analyses can provide an indication of the impact of the uncertainty when making the structural assumptions on the model results, these methods cannot capture the overall model uncertainty,³⁴⁶ and do not provide an indication of the most credible scenario.³⁴⁷ No study used alternative methods to explore the impact of making simplifying assumptions on the model results, and more research is needed to develop these. However, current approaches for handling structural uncertainty are underdeveloped and research in this area is ongoing.³⁴⁶

6.7.4 Implications for practice

The implications of the findings of the review of simplifying assumptions for practice, which also inform the framework provided in Chapter 8, are presented here. The recommendations for future research are presented in Chapter 9.

The evidence and rationale for making simplifying assumptions, such as treatment independence, needs to be clearly reported. Their impact on the model output or estimates of cost effectiveness also needs to be explored using a range of plausible scenarios, representing different assumptions. The selection of data to inform these scenarios must be justified, and where they have to be based on clinical opinion due to the limited evidence base, should be extracted using appropriate methods of expert elicitation.^{357 358}

The review considered any modelling study that investigated treatment sequences in any disease condition, and as such identifies all the assumptions used to account for treatment-sequencing effects. This comprehensive list of simplifying assumptions representing treatment-sequencing effects can be used as a checklist to inform future practice. The new coding scheme can help commissioners, policy makers, reviewers, and modellers to appraise and understand a model better, and consider whether treatment sequencing should be considered. My research also revealed that

treatment sequencing may need to be considered, whatever the decision problem. The list of simplifying assumptions can also be used by modellers to help them consider this, and what implicit assumptions they may be making; for example, are they assuming all previous treatments are equal? However, in order to apply the coding system, better reporting of the simplifying assumptions made is required.

Most economic evaluations aim to compare single treatments and focus on the downstream costs of subsequent treatments, for example when considering a lifetime horizon.²⁸⁴ They frequently employ the assumption that subsequent treatments, generally used to reflect usual care, are the same for all comparators, with any potential treatment-sequencing effects likely to be ignored or simplified. However, the proportion of patients going on to receive subsequent treatments, or the duration over which they are applied is likely to differ between the 'single' treatments of interest. The current review focused primarily on economic models that aimed to evaluate the effectiveness of treatment sequences. However, even when evaluating the cost-effectiveness of single treatments, the decision model should adequately account for the potential effects of treatment sequencing. The list of simplifying assumptions identified during the review will be useful for clarifying what is currently being used in all modelling studies.

Modelling treatment sequences, as an alternative to comparing individual treatments used at a single point in the pathway, is important as it is more representative of real-life management of chronic conditions. This in turn reflects the importance of using the best available evidence to inform whole treatment sequences, or the treatments used beyond the decision point of interest. However, there are currently no standards or guidance available for selecting appropriate assumptions or data sources, which would be useful for informing and standardising practice. Treatment sequencing trials are likely to be limited, and the uncertainty in the quality of the evidence base used as alternatives to inform the sequencing effects needs to be investigated and its potential impact on the decision making assessed.

The need to consider treatment sequences in the decision analysis, and the validity of using the assumption of treatment independence can be explored as part of the clinical evaluation, based on the assessment of clinical and statistical heterogeneity within the meta-analysis as outlined in Chapters 2 and 4. Ideally the clinical evaluation would also incorporate the development of the reduction or multiplication factors for adjusting treatment effects to account for treatment sequencing, or develop summary treatment effect-estimates that are conditional on positioning in the sequence. However, this would require treatment sequencing to be part of the original health technology assessment brief and not merely part to the economic model to account for the cost of downstream treatments.

A potential reason for variation in the data sources used is the available time for identifying and selecting the evidence base. The time frame for producing an economic model and identifying the

best available evidence to inform it is generally limited. There is an increasing demand by policy makers for rapid reviews of the evidence base, especially for the introduction of single new technologies, which will lead to an even more condensed timeframe to work with. The clinical evaluation undertaken within a health technology assessment frequently fell short of the needs of the economic evaluation that included a model of treatment sequences, as they rarely considered treatment-sequencing effects, or concluded that insufficient evidence was available to evaluate treatment sequences. The lack of integration or direct use of the systematic review to inform the economic evaluation was also identified in a previous review of NICE technology appraisals.³⁵⁹ This review also identified the need for those undertaking a health technology assessment to consider the data requirement of the economic model at an early stage.³⁵⁹

Economic evaluations undertaken by, or on behalf of, industry tended to focus on a specific decision point, which generally reflected the treatments used in available RCTs. The need to consider treatment sequencing should be identified during the scoping stage of the health technology assessment, and incorporate both the clinical and economic evaluation. The development and use of a conceptual framework to inform the whole technology appraisal may help this. I return to the use of a conceptual framework in Chapter 7, Section 7.4.6.

6.8 THE NEXT STAGE

The next chapter reviews the range and type of modelling approaches used for evaluating treatment sequences and how they were characterised within the model. The advantages and disadvantages of different modelling approaches are also reviewed.

CHAPTER 7: METHODOLOGICAL REVIEW OF DECISION ANALYTIC MODELLING APPROACHES USED FOR EVALUATING TREATMENT SEQUENCES

7.1 CHAPTER OVERVIEW

This chapter summarises the actual modelling approaches used for evaluating treatment sequences in practice, as identified in the literature review. It includes a description and evaluation of the different modelling approaches identified using a series of examples illustrating their application. These are ordered according to increasing complexity or sophistication of the modelling structure. It also includes a summary of the advantages and disadvantages of each approach for modelling treatment sequences.

The term 'modelling approach' is used to describe the overarching modelling method, which includes the modelling technique, structure and assumptions used. Modelling technique is used to refer to the actual procedure used, and modelling structure is adopted here to describe the conceptual configuration of the decision problem.

7.2 DEVELOPING CRITERIA FOR ASSESSING MODELLING APPROACHES USED FOR EVALUATING TREATMENT SEQUENCES

Several decision analytic modelling techniques can be used for evaluating treatment sequences. Most techniques used in health economics in general have been borrowed from other fields and have different features that make them more or less applicable for different circumstances. The key features and underpinning assumptions of different modelling techniques and structures can potentially impede or assist the overall modelling and configuration of treatment sequences within the model.

Two published taxonomies, developed for categorising different modelling techniques according to their key features, were used to guide the review of modelling studies.^{344 360} These were used to categorise the included studies and inform the criteria used for data extraction (Section 7.2.2). A summary of the two taxonomies and an overview of the key features they describe is provided in the Appendix Volume I (Section E). This also includes a description of parameter uncertainty and its representation in the model (Section E3.6).

7.2.1 Existing guidance on selecting an appropriate modelling technique

In addition to the two modelling taxonomies,^{344 360} a number of guides and algorithms have been developed to aid the selection of an appropriate modelling technique (also referred to as structures in some guides) for economic evaluation in general.^{242 361-371}

The predominant modelling techniques used for health economic evaluation include the cohort-based models, the decision tree and the Markov cohort model. The guidance developed by the International Society for Pharmacoeconomics and Outcomes Research and the Society for Medical Decision Making (ISPOR-SMDM) Joint Modelling Good Research Practices Task Force recommends that a state transition model, rather than a simpler model with limited ability to reflect time, for example a decision tree, should be used if the model requires time dependent parameters, time to an event, or repeated events.³⁷² However, they also note that other modelling techniques, such as discrete event simulation, are suitable for these situations. A number of other guides recommend the use of patient-level simulation models as the preferred option when the assumptions required by a Markov model are not sustainable,^{361 364 373 374} and for modelling treatment sequencing.^{254 272 373} A review of NICE technology appraisals published up until January 2005 identified the following reasons for choosing an individual patient-level simulation approach: treatment switching, sampling patient characteristics, and dependence of patient histories including previous events and time in state.³⁴⁵ However, the authors went on to advocate further adapting the cohort framework for evaluating treatment sequences, instead of an individual patient level model, due to the increased computational expense, decreased transparency, and the potential requirement for specialist skills.³⁴⁵ This is also compounded by the need to conduct probabilistic sensitivity analyses. In terms of the computational expense, some argue that a discrete event simulation represents a more efficient approach^{367 373} than patient-level state-transition models, which require the current health state to be calculated for each patient at each model cycle making them computationally inefficient.³⁷³ Unlike state transition models that are focused around health states, discrete event simulation is conceptualised around events.³⁶⁷ The model moves forward in time to the point at which the next event is experienced, therefore requires fewer calculations per patient than a patient-level state-transition model.^{367 373} It also allows greater flexibility in modelling timing of health-related events.³⁴⁴ Discrete event simulation can represent sophisticated methods, requiring extra time and expertise to implement.^{345 367 374} However, Davies *et al.* argued that discrete event simulation is more transparent and intuitive to clinicians than complicated cohort models.³⁷³

The type of data required to implement and characterise the parameters of the chosen modelling technique can be an important factor. Modelling approaches based on individual patient simulation require more data to populate than a cohort-based model. However, for state-transition models, developing accurate transition probabilities is a potential problem whether they are based on cohort or individual-level simulation.^{363 372 375} Pooled estimates from meta-analyses, and associated uncertainty, need to be converted to transition probabilities before they can be applied to the model and, for individual-patient level simulation models, time-to-event data, for example survival data, need to be converted to time-dependent transition probabilities.^{83 84 373}

7.2.2 Criteria used for categorising and assessing included modelling approaches

The key features outlined in the two modelling taxonomies,^{344 360} and supported by the various guides, which are likely to be relevant for modelling treatment sequences were used for categorising included models. The modelling approaches used were grouped according to whether they:

- i. Were conceptualised around states or events
- ii. Simulated a closed cohort or a dynamic population
- iii. Simulated all individuals simultaneously or one at a time
- iv. Based on the Markovian assumption or not
- v. Allowed transitions to occur only at specified time intervals or not
- vi. Used expected or stochastic variables

Within these groupings included models were also assessed in terms of how treatment sequencing was conceptualised, how the potential limitations of the Markovian property were overcome, and how time was interpreted in the model. They were also assessed in terms of their ability to account for:

- i. Heterogeneity in the target population
- ii. The outcomes of different subgroups
- iii. The time dependency of certain parameters
- iv. Repeated events
- v. Competing risks
- vi. Patient history or previous treatments
- vii. The need for differential treatment selection based on reason for discontinuing previous treatments
- viii. Parameter uncertainty
- ix. Dynamic decision making

Other factors that are likely to influence the choice of modelling approach include the extent of the available evidence base, time and resources, and the need to include probabilistic sensitivity analysis to inform decision making. These factors were also taken account when assessing the advantages and disadvantages of included modelling approaches.

The choice of an appropriate modelling approach depends on the complexity of the underlying decision problem, the extent of the treatment sequences being investigated, and the disease condition. The methods used to code the modelling studies according to the type of decision problem relating to treatment sequencing evaluated, is described in Chapter 6, Section 6.2.3 and 6.4.2. The type of sequencing decision problems evaluated by included studies are listed in Tables 6.1 and 6.2. The time horizon and number of treatment lines modelled are presented in Tables 7.1 and 7.2. Modelling studies relating to rheumatologic conditions (Table 7.1) are listed separately from those of non-rheumatologic conditions (Table 7.2) in the same way as presented in Chapter 6 (Section 6.5.2).

7.3 MODELLING APPROACHES USED FOR EVALUATING TREATMENT SEQUENCES

The literature searches and the methods used to identify and review the relevant modelling studies are described in Chapter 3.

7.3.1. Summary of modelling approaches used by included studies

The review of modelling approaches included 70 studies,^{17 18 115 199 204 207 216-279} 36 (51%) of which were rheumatology studies,^{17 216 218 219 222-225 227-230 232 234 235 238 241 245 246 248 249 251 253 254 256 260-262 264 265 270 272-275 279} and ten (14%) were oncology.^{204 207 226 236 243 250 252 257 269 277} The type of modelling approach used for investigating treatment sequences included deterministic decision tree, stochastic decision tree, Markov cohort model, partitioned survival cohort model, semi-Markov cohort model, individual-patient simulation (IPS) state transition models, discrete event simulation, non-terminating population based simulation, terminating population based simulation, and dynamic Markov cohort model. The modelling technique used by each study is listed in Tables 7.1 and 7.2, along with information on the type of economic analysis conducted. A number of studies used a previously published model and, where feasible, these have been collated and are referenced under the original model. The models are referred to here using the model name listed in the third column of Tables 7.1 and 7.2. These are based on the name of the model provided in the publication, where given, or the name of the lead author. One exception to this is the model published by Fitzsimmons *et al.*, which is named as the 'sciatica model' as this was developed using the data from the network meta-analysis presented in Chapter 2.¹¹⁵

Barton *et al.* proposed the use of 'individual sampling model' to describe all models in which the ability to track individuals is an essential part of the model structure, and where only one individual is modelled at a time.³⁶¹ This was due to inconsistency in the use of the descriptors, 'discrete event simulation' and 'state transition models', to describe what appeared to be identically structured models in some instances. However, I have tried to differentiate between these two modelling techniques, as a number of researchers advocate the use of discrete event simulation over patient-level state-transition model. For the purpose of this review the descriptor 'state-transition models' was used for those that only allow transitions to occur at specified time intervals. However, the modelling technique was often not adequately described, which means that the accuracy of this categorisation may not always be assured. Furthermore, it is also acknowledged that several modelling techniques fall under the broad category of 'state transition models', including models that use a continuous state process, and those that allow interactions between groups.³⁷² However, none of the included modelling studies used these model types.

Table 7.1: Modelling approaches used by rheumatology studies

(ordered by model type then alphabetically by author)

Author, year Country	Condition	Modelling technique	Model name*	Type of analysis (PSA conducted)	Time horizon	Lines of treatment in model
Beresniak, 2011 Spain	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3

Author, year Country	Condition	Modelling technique	Model name*	Type of analysis (PSA conducted)	Time horizon	Lines of treatment in model
Beresniak, 2013 Germany	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3
Cimmino 2011 Italy	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3
Puolakka, 2012 Finland	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3
Russell, 2009 Canada	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3
Sarau, 2010 France	RA	Stochastic decision tree	Advanced simulation model	CEA (YES)	2 years	3
Albert, 2000 US	RA	Markov cohort	Albert model	CE (NO)	Not stated	3
Coyle, 2006 Canada	RA	Markov cohort	Maetzel model	CUA/CEA (YES)	5 years	4 to 5 or 2 to 3 (based on toxicity)
Schipper, 2011 Netherlands	RA	Markov cohort	Welsing model	CUA (YES)	5 years	5
Tanno, 2006 Japan	RA	Markov cohort	Tanno model	CUA (YES)	Lifetime	3 or 4
Welsing, 2005 Netherlands	RA	Markov cohort	Welsing model	CUA/CEA (YES)	5 years	1, 2 or 3
Wu, 2012 China	RA	Markov cohort	Wu model	CUA/CEA (YES)	Lifetime	4, 5, or 6
Maetzel, 2002 Canada	RA	Markov cohort	Maetzel model	CUA/CEA (YES)	5 years	3 to 4, or 5 to 6 (based on toxicity)
Rodgers, 2011 UK (NICE TA 199)	PsA	Markov cohort	York psoriatic arthritis model	CUA (YES)	Lifetime	2
Launois, 2008 France	RA	Dynamic Markov cohort	Launois model	Budget impact	Lifetime	3 or 4
Schadlich, 2005 Germany	RA	Partitioned survival	Schadlich model	CUA/CEA (YES)	3 years	1 or 2, 4 or 5, and 5 or 6
Brennan, 2004 UK (NICE TA36 - IS)	RA	Individual sampling (STM)	Sheffield Etanercept model	CUA (YES)	Lifetime	3 to 4
Bansback 2005 Sweden	RA	Individual sampling (STM)	Bansback model (<i>Sheffield Etanercept and BPM</i>)	CUA (YES)	Lifetime	1 or 2
Brennan, 2007 UK (NICE TA 130 - IS)	RA	Individual sampling (STM)	Sheffield BSRBR model	CUA (YES)	Lifetime	2
Davies, 2009 US	RA	Individual sampling (STM)	Bansback model	CUA (YES)	Lifetime	4, 5 or 6
Diamantpoulus, 2014 UK	RA	Individual sampling (STM)	Diamantpoulus model	CUA (YES)	Lifetime	3 or 4, and 6 or 7
Diamantpoulus, 2012 Italy	RA	Individual sampling (STM)	Diamantpoulus model	CUA (YES)	Lifetime	4 or 5
Finckh, 2009 US	RA	Individual sampling (STM)	Sheffield AHRQ model	CUA (YES)	Lifetime	6 or 7
Hallinen, 2010 Finland	RA	Individual sampling (STM)	Kielhorn model	CUA (YES)	Lifetime	4, 5, or 6
Kielhorn, 2008 UK	RA	Individual sampling (STM)	Kielhorn model	CUA (YES)	Lifetime	3, 4, 5, or 6
Kobelt, 2011 Sweden	RA	Individual sampling (STM)	Kobelt model	CUA (YES)	10 years	1 to 2
Merkesdal, 2010 Germany	RA	Individual sampling (STM)	Kielhorn model	CUA (YES)	Lifetime	4 to 5

Author, year Country	Condition	Modelling technique	Model name*	Type of analysis (PSA conducted)	Time horizon	Lines of treatment in model
Wailoo, 2006	RA	Individual sampling (STM)	Sheffield AHRQ model	CUA (YES)	Lifetime	1 to 3
Barton 2004 UK (NICE TA36)	RA	Individual sampling (DES)	BRAM	CUA (NO)	Lifetime	9 or 10
Chen, 2006 UK (NICE TA130)	RA	Individual sampling (DES)	BRAM	CUA (NO)	Lifetime	10 to 11 (or 13)
Clark, 2004 UK (NICE TA72)	RA	Individual sampling (DES)	BRAM	CUA (NO)	Lifetime	9, 10, or 11
Jobanputra, 2002 UK (NICE TA36)	RA	Individual sampling (DES)	BPM	CUA (NO)	Lifetime	9 to 10
Malottki, 2011 UK (NICE TA195)	RA	Individual sampling (DES)	BRAM	CUA (YES)	Lifetime	4 to 5
Tran-Duy, 2014 Netherlands	RA	Individual sampling (DES)	Tran-Duy model	CUA (YES)	Lifetime	8 or 12
Tran-Duy, 2011 Netherlands	AkS	Individual sampling (DES)	Tran-Duy model	CUA (YES)	Lifetime	5 or 7
Lindgren, 2009 Sweden	RA	Individual sampling (DES)	Lindgren model	CUA (YES)	Lifetime	3

* A number of studies used a model developed by a previous study. This is the name of the representative model under which the model is described within the text.

Abbreviations: AkS Ankylosing spondylitis; BPM Birmingham Preliminary Model; BRAM Birmingham Rheumatoid Arthritis Model; CEA cost effectiveness analysis; CE clinical effectiveness; CUA cost utility analysis; DES discrete event simulation; IS independent submission (to NICE, including industry or charity); PSA probabilistic sensitivity analysis; PsA psoriatic arthritis; RA rheumatoid arthritis; STM state transition model

Table 7.2: Modelling approaches used by non-rheumatology

(ordered by model type then alphabetically by author)

Author, year Country	Condition	Modelling technique (For base-case analysis)	Modelling name*	Type of analysis (PSA conducted)	Time horizon	Lines of treatment in model
NICE CG131 (2011)	Cancer (advanced CRC)	Stochastic decision tree and partitioned survival	NICE CG131 (Added marginal value method)	CUA (YES)	Lifetime	Up to 2
Dranitsaris, 2011 Malaysia	Cancer (metastatic CRC)	Decision tree	Dranitsaris model	CUA (NO)	Lifetime	2
NICE CG81 (2009)	Cancer (advanced BC)	Decision tree	NICE CG81	CUA (YES)	Lifetime	Up to 3
NICE CG152 (2012)**	Crohn's disease (induction / maintenance)	Decision tree / Markov Cohort	NICE CG152	CUA (YES)	30 weeks / 2 years	Up to 4
Fitzsimmons, 2014 UK	Sciatica	Decision Tree	Sciatica model	CUA (NO)	1 year	Up to 3
Frankum, 2005 US	Onychomycosis	Decision Tree	Frankum model	CUA (YES)	1-3 years	3
Knoester, 2007 Netherlands	Epilepsy (>12 yrs of age)	Decision Tree	Knoester model	CUA (YES)	1 year	Up to 2
Bensmail, 2009 France	Spasticity	Stochastic decision tree	Advanced simulation model	CEA (YES)	2-years	3 to 4
Greenhalgh, 2005 UK (NICE TA59)	Major depressive disorder (MDD)	Stochastic decision tree	Greenhalgh model	CEA (YES)	1 year	3
Anis, 2011 US	Psoriasis	Markov cohort	York psoriasis model	CUA (YES)	10-16 weeks	Up to 6
Beard, 2011	Fibromyalgia	Markov cohort	Beard model	CUA/CEA	2 years	5 to 6

Author, year Country	Condition	Modelling technique (For base-case analysis)	Modelling name*	Type of analysis (PSA conducted)	Time horizon	Lines of treatment in model
US				(YES)		
Cameron, 2008 UK	Cancer (advanced BC)	Markov cohort	Cameron model	CUA (YES)	10-years; lifetime	4 to 5
Davies, 2008 UK	Schizophrenia	Markov cohort	Davies model	CUA (YES)	10 years	3
Heeg, 2015 The Netherlands	Cancer (Multiple Myeloma)	Markov cohort	Heeg cancer model	CE (YES)	Lifetime	4
Lee, 2013 South Korea	Cancer (Ovarian cancer)	Markov cohort	Lee model	CUA (YES)	10-years; lifetime	5
Lux, 2009 Germany	Cancer (advanced BC)	Markov cohort	Cameron model	CUA (YES)	10-years; lifetime	4 to 5
Orme, 2012 UK	Glaucoma	Markov cohort	Orme model	CUA (NO)	10 years	3
Sawyer, 2013 UK (NICE CG 153)	Psoriasis	Markov cohort	Sawyer model	CUA (YES)	1 year	Up to 3
NICE CG137 (2012)	Epilepsy (in children and adults)	Markov cohort	NICE CG137 (structure based on York epilepsy model)	CUA (YES)	15 years	Up to 3
Shepherd, 2006 UK (NICE TA96)	Chronic Hep B infection	Markov cohort	Shepherd model	CUA (YES)	Lifetime	up to 3
Sizto, 2009 Canada	Psoriasis	Markov cohort	York psoriasis model	CUA (YES)	Unknown	6
Smith, 2007 US	Postherpetic neuralgia (PHN)	Markov cohort	Smith model	CUA (YES)	Lifetime	4 to 5
Soini, 2012 Finland	Cancer (Follicular non- Hodgkin lymphoma FL)	Markov cohort	Soini model	CEA/CUA (YES)	25 years; lifetime	2
Tebas, 2001 US	HIV	Markov cohort	Tebas model	Virology (NO)	10 years	Up to 3
Wong, 2009 US	Cancer (metastatic CRC)	Markov cohort	Wong model	CUA (NO)	Lifetime	Up to 3
Woolacott 2006 UK (NICE TA103)	Psoriasis	Markov cohort	York psoriasis model	CUA (YES)	10 years	3 and 7
Hind, 2008 UK (NICE TA93)	Cancer (advanced CRC)	Partitioned survival	Hind model	CEA/CUA (YES)	Lifetime	Up to 2
Wilby, 2005 UK (NICE TA76)	Epilepsy (in adults)	Semi-Markov cohort	York epilepsy model	CUA (YES)	15 years	Up to 3
Holmes, 2006 South Africa	HIV	Individual sampling (STM)	Holmes model	CEA (YES)	Lifetime	Up to 2
Connock, 2006 UK (NICE TA79)	Epilepsy (in children)	Individual sampling (DES)	Birmingham epilepsy model	CUA (NO)	15 years	Up to 4
Denis, 2008 France	Glaucoma	Individual sampling (DES)	Denis model	CE (NO)	5 years	2
Heeg, 2008 The Netherlands	Schizophrenia	Individual sampling (DES)	Heeg schizophrenia model	CUA (YES)	5 years	Up to 4
McEwan, 2010 UK	Type 2 diabetes mellitus	Non-terminating population based simulation	Cardiff T2DM model	CUA/CEA (NO)	Lifetime	3
Erhardt, 2012 Germany	Type 2 diabetes mellitus	Terminating population based simulation	Cardiff T2DM model	CUA (NO)	Lifetime	3

* A number of studies used a model developed by a previous study. This is the name of the representative model under which the model is described within the text.

** Different modelling approach used for two clinical situations: induction of remission and maintenance of remission.

Abbreviations: CRC colorectal cancer; BC breast cancer; CE cost-effectiveness analysis; CG clinical guideline; CUA cost utility analysis; DES discrete event simulation; T2DM Type 2 diabetes mellitus; PSA probabilistic sensitivity analysis; STM state transition model

The next section includes a summary of each modelling approach and provides an overview of the range of relevant features that were accounted within the models using each approach. These are illustrated using a series of examples. A more detailed description of each included model is provided in Appendix Volume II (Appendix E), where different modelling approaches are presented in separate tables, as are studies investigating rheumatological and non-rheumatological conditions. A summary of the treatment sequences evaluated, the source of clinical effectiveness estimates, and assumptions used to inform treatment sequencing effects for each included study is presented in the Appendix Volume II (Appendix D), which is linked to the review of assumptions in Chapter 6. The review of the advantages and disadvantages of each approach, summarised at the end of each section is based on all included models.

7.3.2 Decision tree

7.3.2.1 Description of studies using decision tree modelling

Fourteen studies used a decision tree model to evaluate treatment sequences, including three NICE clinical guidelines (CG81, CG131, CG152).^{115 207 221-223 228 236 239 240 247 257 258 260-262} Two clinical guideline studies used the decision tree in conjunction with another modelling technique. One clinical guideline study (NICE CG131) used a simple stochastic decision tree alongside a partitioned survival analysis, also known as the added marginal value method (Table 7.2).²⁰⁷ A second clinical guideline study (NICE CG152) used separate modelling techniques for treatment induction and subsequent maintenance of remission in Crohn's disease.²⁵⁸ Here, a decision tree model was used for comparing predefined treatment sequences for induction of remission in patients with an acute exacerbation, whilst a separate Markov cohort model was used for comparing single maintenance treatments, with patients who relapsed receiving the most cost-effective acute induction treatment sequence identified using the decision tree model.

Seven studies based on the same model called the advanced simulation model, were cost-effectiveness analyses.^{221-223 228 260-262} The remaining studies were cost-utility analyses. Three studies (NICE CG 81; NICE CG131; Dranitsaris, 2011) used the decision tree framework for evaluating treatment sequences for advanced cancer over a lifetime horizon.^{207 236 257} The timeframe for remaining models ranged from one to three years. Most studies evaluated fairly simple treatment sequences consisting of two to three lines of treatment.^{222 223 228 236 239 240 247 260-262} Details of each included model is presented in the Appendix Volume II (Appendix E). A schematic diagram of a decision tree, using the Dranitsaris model as an example, is provided in Figure 7.1.

Two studies only conducted a deterministic analysis of the decision tree model,^{115 236} whilst five used a deterministic analysis of the base case model along with subsequent probabilistic sensitivity analysis.^{207 239 247 257 258} These have been categorised here as deterministic decision trees. The decision tree modelling technique called advanced simulation model is described as a type of decision tree which allow computing variable distributions.²²³ This is essentially the same as probabilistic sensitivity analysis, and has been categorised here as a stochastic decision tree,³⁶⁰ and

summarised separately.^{221-223 228 260-262} One additional study (Greenhalgh, 2005), which implemented the model using Monte Carlo simulation, is also categorised here as a stochastic decision tree.²⁴⁰

7.3.2.2 Deterministic decision tree

A deterministic tree was used for modelling treatment sequences for a wide range of different conditions including advanced cancer (NICE CG81, Dranitsaris model, NICE CG131),^{207 236 257} neurological pain (Sciatic model),¹¹⁵ fungal infection (Frankum model),²³⁹ epilepsy (Knoester model),²⁴⁷ and Crohn's disease (NICE CG152).²⁵⁸ The deterministic decision tree modelling approach was chosen in some studies due to the limitation of the evidence base. Some examples of alternative modelling approaches used to model treatment sequences for similar clinical conditions, for example epilepsy, are presented in the next sections.

The different deterministic tree modelling approaches used are described in this next section using, primarily, the studies modelling overall survival for advanced cancer as examples. However, it starts with a description of the sciatica model, which provides an example of a fairly simple deterministic decision tree model used for implementing a very large number of treatment sequences. The same decision tree model can be used for modelling multiple pre-defined treatment sequences, or alternatively each sequence can be modelled separately. The NICE CG157 model for Crohn's disease provides an example of a simple decision tree used to model each sequence separately. A detailed summary of this model is provided in the Appendix Volume II (Appendix E). The Knoester model, which is also described in more detail in the Appendix, provides an example of a decision problem that incorporated the differential use of a second treatment depending on the reason for discontinuing current treatment. The Frankum model is described briefly at the end of this section, as it provides an example of a decision problem where a successful treatment of onychomycosis could result in either permanent resolution or a recurrent infection that requires re-treatment.

The sciatica model

The sciatica model compared more than 100 treatment strategies, used over a 12-month period, within a single deterministic decision tree.¹¹⁵ This included the evaluation of three different treatment pathways, which represented: i) *primary care management*, which only included an initial treatment (n=5); ii) *a stepped approach*, based on one of the initial treatments followed by an intermediate treatment (n=6) and then invasive treatments, which included epidural, or epidural followed by disk surgery; and iii) *immediate referral to surgery* following an initial treatment (n=5) in primary care. The probability of success for each individual treatment was based on comparisons with placebo taken from a single network meta-analysis presented in Chapter 2. Most comparisons demonstrated wide confidence intervals around the summary estimate of effect, which was why a deterministic rather than a probabilistic approach was chosen. A counter-argument to this, is that this uncertainty could have been propagated through the model, thus reflecting the resulting large uncertainty in the probability of each intervention being cost-effective. The limited time horizon was chosen because of the heterogeneity in duration of effect, and lack of data on relapse and recurrence made it difficult to

extend the analysis beyond twelve months. Where multiple treatments were used in a sequence, their summary effects were combined. In other words, successive treatments were assumed to provide an additive effect, rather than the probability of success at each chance node being attributed only to the treatment used immediately prior to this point. It was also assumed that there was no reduction in utility with previous unsuccessful treatments, with the probability of a successful outcome multiplied by the same utility estimate, regardless of how many interventions were required to achieve this. However, potential reductions in effectiveness of intermediate therapies and surgery in the stepped approach were evaluated in sensitivity analyses. A reduction in utility and the potential subsequent effects of non-responders at each stage of the pathway were also explored in sensitivity analysis. These were described in more detail in Chapter 5, Section 5.6.2.3, summarising the assumptions of diminishing effects with consecutive treatments, and reduced efficacy with disease duration.

NICE CG81 model

The NICE CG81 model provides an example of a more complex decision tree model used to evaluate treatment sequences for advanced cancer, which required a lifetime perspective, and all outcomes result in eventual death from progressive disease.²⁵⁷ The economic evaluation investigated 17 fixed chemotherapy sequences for patients with metastatic breast cancer who had received prior anthracycline therapy. Treatment sequences included up to three treatment lines, each administered for a fixed period.

Time is not made explicit in a decision tree, rather the aim of the model was to measure how long patients spent in the progression free 'health' state for each sequence. The model also needed to account for multiple levels of treatment response, for which time to progression will differ; different reasons for treatment switching; and the possibility of patients experiencing toxic death, which will have different probabilities and timing to that of all-cause mortality. The initial chance node represented a choice between four first-line treatments, each leading to a decision tree with 28 branches. Time was implicitly incorporated based on the assumption made about the number of cycles that had elapsed prior to the chance node or event occurring. For example, it was assumed that after a patient received one cycle of treatment they would reach a point at which they might die of toxicity. Those who survived received two more cycles, after which they may experience major toxicity. Major toxicity prompted treatment discontinuation with a one-month time-lag included before starting the next treatment. Those who continued treatment faced the probability of having a complete or partial response, stable disease, or not responding. Responders and stable patients went on to receive additional cycles of treatment, whilst non-responders switched to the next treatment if feasible. Overall survival for each treatment sequence was estimated based on the assumption that chemotherapy impacts on time to progression and, through that, overall survival. Overall survival was calculated as the sum of the time to progression from first-line treatment, time to progression from second-line, time to progression from third-line and the period from progression to death, which was assumed to be a fixed period of five months, regardless of chemotherapy treatment (outlined in equation 7.1).

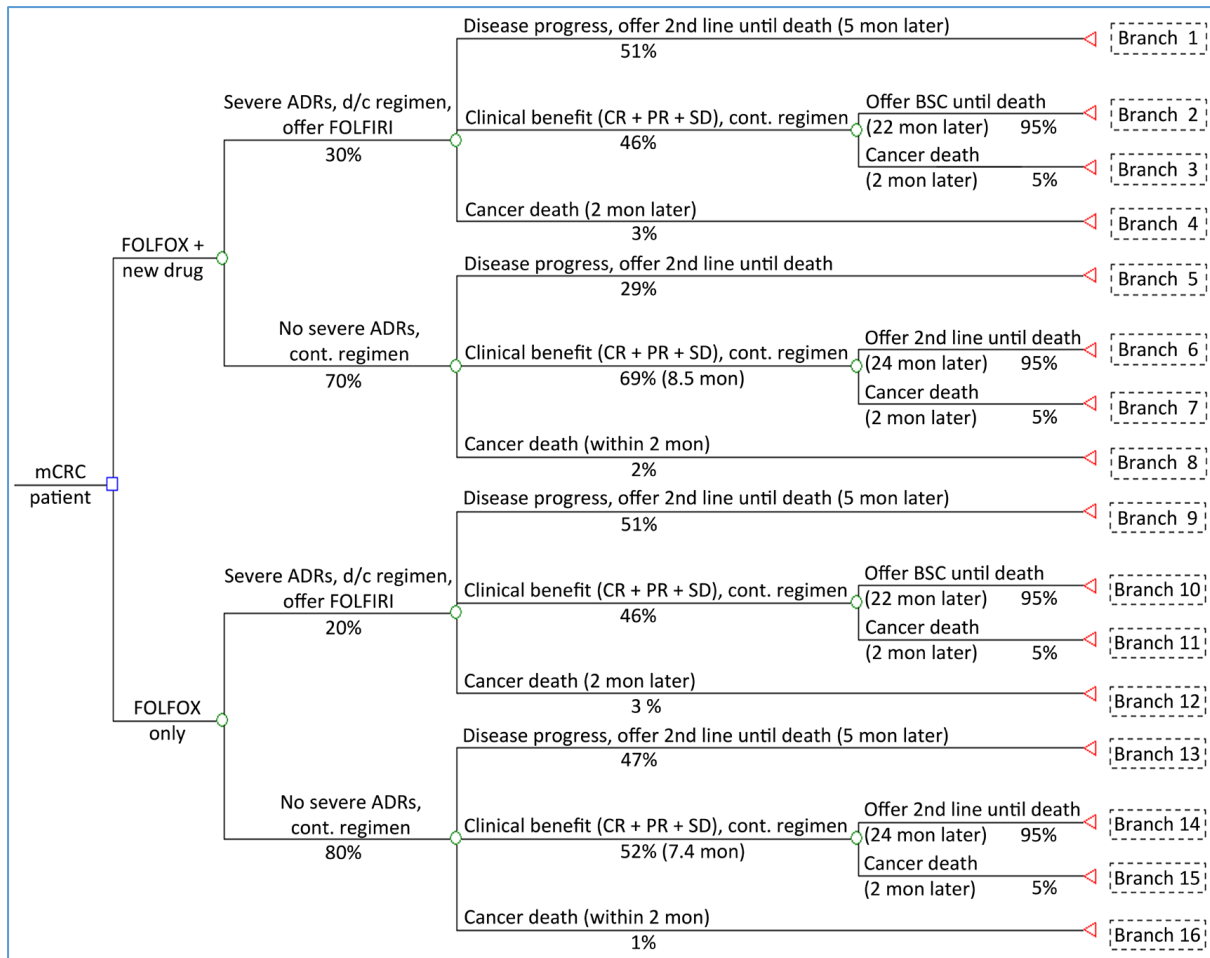
$$OS = (PFS1) + (PFS2) + (PD) \quad [7.1]$$

The use of the endpoint progression free survival for evaluating treatment sequences in advanced cancer is discussed in Chapter 5, Section 5.3.2 using the review by Stenner *et al.* (2012) as example. This includes a discussion on the different interpretation of the outcome progression free survival 2, which is referred to here as the PFS associated with second-line of treatment. These outcomes are also defined and explained further in the Appendix Volume I (Section C4.2).

Dranitsaris model

A similar modelling approach to NICE CG81 was used by Dranitsaris *et al.*, which compared a chemotherapy sequence of only two treatment-lines, with and without the targeted therapy bevacizumab, added to the first line treatment (Figure 7.1).²³⁶ Here the branches, after ascertaining whether the patients experienced severe adverse effects on initial treatments or not, were associated with fixed time periods representing time until progression (treatment duration) or death for the different regimens. This provided a fixed time until death for each terminal branch. No distinction was made between the type of response, with clinical benefit incorporating complete response, partial response, or stable disease. The authors chose to use a decision tree approach as they did not have data on disease progression and toxicity for each cycle of chemotherapy, with only the median number of cycles being presented in the published clinical trials. The authors noted that a Markov model would have been preferable, given its ability to incorporate the element time. Another included study (Wong, 2009), which also evaluated treatment sequences with bevacizumab, used a Markov model (described in Section 7.3.3.3).²⁷⁷

Figure 7.1: A schematic diagram of the Dranitsaris model: a decision tree model for the treatment of metastatic colorectal cancer



Taken from: Dranitsaris, G., Truter, I., Lubbe, M. S., Sriramanakoppa, N. N., Mendonca, V. M., & Mahagaonkar, S. B. (2011). Using Pharmacoeconomic Modelling to Determine Value-Based Pricing for New Pharmaceuticals in Malaysia. *The Malaysian Journal of Medical Sciences*, MJMS, 18(4), 32–43.

Abbreviations: mCRC metastatic colorectal cancer; FOLFOX oxaliplatin in combination with infusional 5-fluorouracil; FOLFIRI irinotecan in combination with infusional 5-fluorouracil; ADR adverse drug reaction; CR complete response; PR partial response; SD stable disease; BSC best supportive care; mon month; d/c discontinued; cont. continue.

NICE CG131 model

The NICE CG131 model, which was used to evaluate treatment sequences for advanced colorectal cancer, was based on a combination of a simple decision tree model and partitioned survival analysis.²⁰⁷ A very simple generic decision tree model was used to account for the proportion of patients discontinuing first-line treatment who do not then go on to receive further active chemotherapy treatments. The study evaluated the cost-effectiveness of 10 fixed chemotherapy sequences containing only two lines of treatment. Survival time was partitioned in the model using the progression free survival and overall survival. Clinical effectiveness was based on quality adjusted life years (QALYs) gained. The model did not explore survival conditional on best response to treatment, as there was insufficient detail reported in the literature to facilitate survival analysis dependent on tumour response. The decision tree started with all patients receiving first-line treatment, which was

represented as a single branch. Disease progression is an inevitable outcome for patients with advanced cancer. Following which the model allowed for a proportion of patients to discontinue treatment and receive no further treatment (and would have disease progression until death), whilst the remaining proportion went on to receive second-line treatment. Patients would continue second-line treatment until disease progression, which they would then receive until death.²⁰⁷ Patients were assumed to be in a stable disease state whilst receiving chemotherapy, and prior to the onset of progressive disease. Following the point of disease progression, patients were assumed to be in a progressive disease state with a lower overall quality of life. It was considered that 60% of patients on first-line treatment go on to receive second-line treatment, based on published studies. However, it was not possible to obtain separate overall survival curves for the two subgroups, of patients who only received one line of treatment or two. The QALY calculations were therefore based on a weighted average of quality-adjusted survival across the combined patient populations and not as a separate absolute estimate for each subgroup. The QALY for patients who only received one line of treatment was estimated using equation 7.2, and 7.3 for patients who received two lines of treatment.

$$(PFS1 \times \text{utility in stable}) + ((OS - PFS1) \times \text{utility in progression}); \quad [7.2]$$

$$(PFS1 \times \text{utility in stable}) + (PFS2 \times \text{utility in stable}) + ((OS - PFS1 - PFS2) \times \text{utility in progression}) \quad [7.3]$$

Progression free survival for first-line treatment (PFS1) was taken from a network meta-analysis of clinical trials of first-line treatment, whilst progression free survival for second-line treatment (PFS2), and overall survival (OS) were taken from network meta-analyses of prospective sequencing studies.

The modelling technique partitioned survival analysis is described in more detail in Section 7.3.4, as it was also proposed as an alternative to Markov modelling, for incorporating time dependency within a cohort framework.

Frankum model

The Frankum model was developed to evaluate the budgetary effect of up to three lines of treatment for toe nail onychomycosis.²³⁹ Unlike treatment failure due to adverse effects, recurrent infection can be treated using the same treatment that was used previously. A sequential treatment pathway framework was used to represent the series of recurring health states, or possible outcomes, as well as the corresponding treatment switching decisions, which patients could follow if they were treated with a predefined sequence of treatments. An initial treatment failure, which could include an adverse effect requiring discontinuation of treatment during the first prescription or a lack of response after a full course of treatment, resulted in switching to a new treatment, whilst a subsequent relapse or recurrence was treated with a second course of the successful initial treatment. The time horizon of the model was assumed to span between one and three years, based on the duration of follow-up and treatment pathway implemented.²³⁹

7.3.2.3 Stochastic decision tree

There were two distinct stochastic decision tree models, the advanced simulation model and Greenhalgh model.

Advanced simulation model

The advanced simulation modelling approach was used primarily for comparing fixed treatment sequences of three biological agents for rheumatoid arthritis.^{222 223 228 260-262} It was also used in one study for evaluating treatment sequences with or without intrathecal baclofen therapy, used as first-line treatment, for the management of severe spasticity.²²¹ The model is discussed in more detail here using the rheumatoid arthritis example (discussed previously in Chapter 6, Section 6.6.2.3 and in the Appendix Volume I, Section D2.1).

The decision trees were designed as simulation models and programmed to take into account the entire distribution of each pre-defined parameter according to specific distribution curves. These were performed using 5000 Monte-Carlo simulations. The results were expressed for one single hypothetical patient.²²³

The rheumatoid arthritis model was mainly used for comparing second- or third-line treatments, whilst still accounting for the treatment switching used in general medical practice (Table 6.4). Separate simulations were run for each fixed treatment sequence, and for two clinical endpoints representing treatment success (achieve remission or low disease activity). The model was run over two years using four six-month treatment periods. A 100% inadequate response rate was assumed for the first TNF-inhibitor at six months in all models, and a 100% inadequate response rate for the second biological agent, at 12 months, for the comparison of third-line agents (Appendix Volume I, Section D2). This approach assumed that treatment success was sustained over the entire six-month treatment period and that each biological agent was maintained as long as clinical response was deemed adequate. The model assumed that if a patient failed to improve after three successive biologic agents, that they would receive conventional DMARDs with a residual level of effectiveness. The overall clinical effectiveness of each treatment sequence was based on the expected number of days in therapeutic success, which was calculated over the two-year time horizon using the following formula [7.4]:

$$N = \sum [S_{r_i} * 180] \quad [7.4]$$

N= expected number of days in therapeutic success

Sr= Success rate over 6 months

I = 6 months treatment period

The advanced simulation modelling approach was used to inform cost-effectiveness analysis, and not used for cost utility analysis. The overall results were based on the cost per day in therapeutic success, and clinical effectiveness was therefore not transformed into utilities. The two-year time

horizon was chosen in order to reflect the available data and avoid having to speculate or make assumptions about long-term effectiveness.^{222 223 228 260-262} The authors noted that the simulation models consumed large amounts of computer processing time, and were implemented using powerful workstations with parallel-processors, and adapted programming language.²⁶¹

Greenhalgh model

The Greenhalgh model was used to assess the optimum positioning of electroconvulsive therapy (ECT) in major depressive disorder.²⁴⁰ It was evaluated as first, second, or third-line treatment, compared with pharmacological treatment. The three treatment lines represented acute treatment, which was followed by psychotherapy in non-responders. Each sequence was modelled separately. The model was based on a one-year time horizon and evaluated using one-week intervals. During each treatment line there was a probability that the patient would end treatment early, due to lack of efficacy or adverse event, and so move to the next treatment. At the end of each treatment line there was a probability that the treatment was successful, and the patient was discharged, or unsuccessful and the patient moved to the next treatment. The probabilities for successful treatment and leaving the treatment early were related to both treatment type and line of therapy (Chapter 6, Section 6.5.3.1). The duration of each treatment, however, was based on a generic estimate of 6 weeks for pharmacological treatments, with dropouts averaging 2 weeks of treatment; and 4 weeks for electroconvulsive therapy, with dropouts averaging 1 week of treatment. Following successful treatment a patient may be given maintenance therapy to help prevent relapse. For each week the model determined whether the patient was in one of the following four states, relating to both the treatment and associated level of depression:

- i. Severely depressed and receiving acute treatment
- ii. Responder: successfully completed acute treatment, no longer severely depressed, and receiving maintenance /continued therapy
- iii. Non responder: receiving long-term psychotherapy, and on completing psychotherapy assumed to improve to mild depression
- iv. Relapsed state following successful treatment. Patients who relapsed from maintenance therapy assumed to require treatment to maintain a quality of life equivalent to moderate depression

The model attributed a specific quality of life utility score to each state (representing severe, moderate, mild, and depression in remission) and determined the movement through these states. The authors noted that a 12-month time horizon was chosen as valid data for longer time periods were not available.²⁴⁰

7.3.2.4 Advantages and disadvantages of the decision tree approach

The decision tree approach is the easiest modelling method to implement, and has the advantage of simplicity and transparency.³⁷⁶ It provides a logical structure for the treatment sequencing decision problem, depicting the related actions and consequences, as they unfold over a specified time

horizon. It has also been successfully used as a simulation model, in order to account for the uncertainty in parameter estimates, but, this may have been at the expense of the simplicity of implementation. Separate models can be used for each treatment sequence or, alternatively, different treatment pathways can be modelled within the same decision tree. However, when a large number of different treatment sequences were implemented in the same model, simplistic assumptions were required regarding treatment switching and the impact of previous treatments. More complexity in the decision problem was successfully accounted for in decision trees used for the comparison of a limited number of fairly simple fixed sequences, generally over the short term, such as a one- or two-year time period. This included accounting for: different reasons for treatment discontinuation and potential impact on subsequent treatment selection; different types or levels of treatment response; the option of relapse being treated with the same previously successful treatment; different reasons for mortality; and the fact that not all patients will be eligible for every subsequent treatment. However, as the number and length of potential sequences increases, or the model needs to account for varying treatment options as a consequence of different reasons for quitting treatment, etc. then the number of branches required will become more extensive. Decision trees may also be inefficient at modelling recurring events over time. Another potential limitation of this approach for modelling treatment sequences is the fact that decision trees are governed by fixed timing of the treatment outcomes, which may be problematic for considering a lifetime horizon. However, duration of time spent in a health state, or on each treatment, can be incorporated using assumptions regarding the timing of events or chance nodes. Overall survival for each treatment sequence can then be estimated based on the assumption that treatment impacts on time to progression and, through that, overall survival, by summing the time to progression for each treatment line plus a period representing progression to death. Alternatively overall survival can be estimated using partitioned survival analysis.²⁰⁷ Modelling treatment sequences for a chronic condition over a lifetime may also be better achieved using a Markov model.

7.3.3 State transition (Markov) cohort model

7.3.3.1 Description of studies using Markov cohort modelling

Eight rheumatology studies (Table 7.1)^{17 216 230 253 265 270 275 279} and sixteen non-rheumatology (Table 7.2)^{217 220 226 231 243 250 252 259 263 266-269 271 277 278} used a Markov cohort modelling approach to evaluate treatment sequences. Most of the studies were cost utility analyses, or included both cost utility and cost-effectiveness outcomes (Table 7.1 and 7.4). However, three studies did not investigate cost-effectiveness. One aimed at evaluating clinical effectiveness only (Albert, 2000),²¹⁶ one overall survival (Heeg, 2015),²⁴³ and the other was based on virologic, rather than clinical, outcomes (Tebas, 2001).²⁷¹ Most studies included probabilistic sensitivity analysis.

The Markov cohort modelling approach was used as a method for both identifying the optimal sequence from all conceivable strategies, and for comparing predefined sequences. The decision models were generally implemented using TreeAge Software or Microsoft Excel.

7.3.3.2 Markov cohort model used for identifying optimal sequence from all conceivable sequences

York psoriasis model: net benefit per unit time

The York psoriasis model, which was developed as part of a NICE appraisal (TA103; Woolacott, 2006), provides an example of a simple Markov cohort model used to identify the optimal ordering of treatments.²⁷⁸ The same model was used by two other studies (Table 7.2).^{217 267} The model was used to investigate the cost-effectiveness of biological agents for the treatment of moderate to severe plaque psoriasis, which is a chronic, relapsing, but non-progressive disease.²⁷⁸ The modelling approach was developed based on the premise that:²⁷⁸

- i. It is the cost-effectiveness of treatment strategies, rather than individual treatments, which needs to be considered for decision-making, but it is not tractable to compare all possible treatment sequences
- ii. The earlier in the sequence a treatment is tried, the greater the proportion of patients who receive and respond to it will be

The authors proposed that in order to maximise the expected total net-benefit per unit time for the overall treatment strategy, individual treatments should be tried in order of decreasing expected treatment period net-benefit per unit time.

The *treatment period net benefit* for individual treatments represents the net-benefit during the entire period a patient receives that treatment, and equates to the weighted average of the expected net benefit incurred over the treatment *lifetime* for those patients who respond to the treatment, plus the expected net benefit over the treatment's initial *trial period* (the interval during which a new treatment is used to see whether it works or not) for those who do not respond to treatment. This means that the analysis is able to account for the proportion of patients who do not respond to treatment but experience some gain in quality of life while still on the treatment before switching to the next in the pathway. Importantly, this approach also accounts for the net benefit of the future treatments needed by those who do not respond to the initial treatment.

Each treatment option was modelled separately using a very simple two-state Markov chain model, with patients either being in a responding or non-responding state. Data on the proportion of patients who responded to each treatment were obtained from a network meta-analysis of placebo controlled trials, which did not consider treatment sequences. The duration of the 'trial' period was based on the design of efficacy trials and expert opinion, whilst duration of the 'treatment' period for responders was based on observational data. Both parameters were entered into the model as fixed parameters. The mean treatment response period was then estimated from a 10-year Markov model using a cycle length of one year.

Interventions that offered a lower expected net-benefit than supportive care were disregarded as they were not cost-effective. The optimum sequence was then identified by ordering the remaining treatments according to their net benefits per unit of time, using those with higher net benefits earlier

in the treatment sequence. The expected net-benefit per unit time for the overall strategy was estimated using the equation presented below [7.5].²⁷⁸

$$NB_A + (1 - P_A)NB_B + (1 - P_A)(1 - P_B)NB_C \quad [7.5]$$

where NB_A , NB_B , and NB_C are the expected net benefits per unit time estimated for treatment A, B, and C, respectively; P_A and P_B is the probability of responding to treatment A and B, respectively.

The same approach was also implemented using cost-effectiveness ratio per unit time (Anis, 2011 and Sizto, 2009).^{217 267}

7.3.3.3 Markov cohort model used for comparing predefined sequences

Most studies used the Markov cohort modelling approach for comparing predefined treatment sequences.^{17 216 220 226 230 231 243 250 253 259 263 265 266 268-271 275 277 279} Treatment sequences were characterised within these models in three different ways, as:

- i. A series of health states, representing successive treatment lines that patients progressed through in a forward motion
- ii. A series of health states representing each treatment-line, but with additional health states to account for other relevant factors or attributes
- iii. A Markov cycle tree to implement treatment switching with the health states used to represent the patients' transition through different levels of disease activity, or model the natural history of the condition

These are illustrated in Figures 7.2 to 7.6, which provide schematic diagrams of Markov cohort model examples. The first approach is illustrated in Figures 7.2 and 7.3, using the Cameron model and Heeg model, respectively. The second approach is illustrated in Figure 7.4 using the Lee model. The third approach is illustrated in Figures 7.5 and 7.6, using the Maetzel model and Welsing model, respectively.

Models using Markov states to represent successive treatment lines

Six studies (one rheumatology²⁷⁰ and five non-rheumatology studies^{226 243 252 269 271}) characterised treatment sequences as a series of health states. This approach is illustrated here using four studies (Cameron, 2008; Heeg, 2015; Lux, 2009; Soini, 2012)^{226 243 252 269} that investigated treatment sequencing for cancer. In essence, this approach can be viewed as building on the three-state structure (described in Appendix Volume I, Section C4.5, Figure C3) typically used for modelling single cancer treatments consisting of progression free survival (PFS), survival with progressive disease, and death. The model presented by Soini *et al.* represents the simplest adaptation of this typical structure, and is therefore discussed first. Two cancer studies (Cameron, 2008 and Lux, 2009) used the same modelling approach, referred to here as the Cameron model (Figure 7.2), which accounted for the fact that not all patients receive all treatments in the sequence The Heeg cancer model (Figure 7.3) accounted for differences in survival according to level of response. As noted previously in Section 7.3.2.2, under the description of the NICE CG81 decision tree model, the

outcomes used to evaluate treatment sequences for cancer are defined and explained in further in the Appendix Volume I (Section C4.2). The models used in the two non-cancer studies (Tanno, 2006 and Tebas, 2001) are described in the Appendices.

Soini model

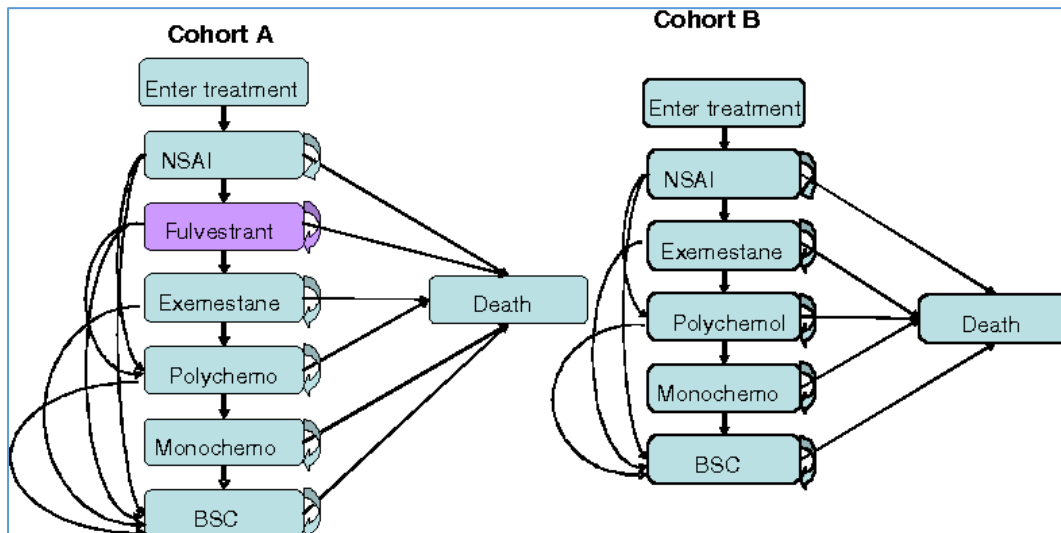
The Soini model was used for evaluating treatment sequences for Follicular lymphoma, which is the most common and incurable form of non-Hodgkin Lymphoma.²⁶⁹ The model structure was aligned with the clinical objective of placing patients into a progression free state for the longest period possible, and was used to estimate both mean progression free survival and overall survival for each sequence. The model included four health states: progression free first-line treatment (PF1), progression free second-line treatment (PF2), progressive disease, and death. Essentially, the 'progressive disease state' here is equivalent to the 'best supportive care state' used in the Cameron model described below. However, despite the Soini model structure being simple, the four treatment sequences that were investigated were fairly complex, and not entirely clear. They represented the use of first-line maintenance treatment compared with observation, followed by the comparison of two second-line induction treatments. First-line maintenance was only used for patients who responded to first-line induction treatment, and second-line induction treatments only differed between sequences for patients who relapsed within one year of first-line maintenance treatment. In the case of early treatment failure a non-cross resistant scheme was preferred, whilst patients with long remission were assumed not to be affected by previous treatments (Chapter 6, Section 6.5.3.4).

Cameron Model

The Cameron model was used to compare treatment sequences with and without fulvestrant in a hypothetical population of hormone receptor-positive postmenopausal women with advanced breast cancer (Figure 7.2).^{226 252} Virtual patients were drawn and randomised to one of two cohorts differing only in whether fulvestrant was administered or not, as either the second,^{226 252} or third treatment.²²⁶ This enabled the implementation of treatment-specific states. The treatment concept for the cohort without fulvestrant encompassed seven health states, which included an initial health state, four active treatment states, best supportive care, and a terminal state, death. Patients progressed forward through the model and could not switch back to any earlier treatments. A cycle length of 28 days was used, which corresponded to one interval of fulvestrant administration. At the end of each cycle, patients could either remain on the same treatment, experience disease progression and make a transition to another state, or die. The probability of any of these events occurring were dependent on treatment-specific median time to progression data and the probability of dying on each treatment-line. Once patients had experienced a progressive event they could move to any later health state, including best supportive care. This was to reflect clinical practice, where not all patients receive all the treatments in the sequence. Estimates for the probability of skipping one or more lines of treatment or dying whilst on each treatment-line, were derived from a clinician survey. The patients' progression sequences were generated by drawing from estimated probability distributions describing

progression rates, proportion of patients skipping treatment lines, and proportions deceased during each treatment line.²⁵²

Figure 7.2: A schematic diagram of the Cameron model: a Markov cohort model for the treatment of advanced breast cancer



Taken from: Cameron, D.A., Camidge, D.R., Oyee, J. & Hirsch, M. (2008) Economic evaluation of fulvestrant as an extra step in the treatment sequence for ER-positive advanced breast cancer. *British Journal of Cancer*, 99(12), 1984-90.

Note: each compartment represent individuals mutually exclusive health states. Arrows represent allowed transitions between these states

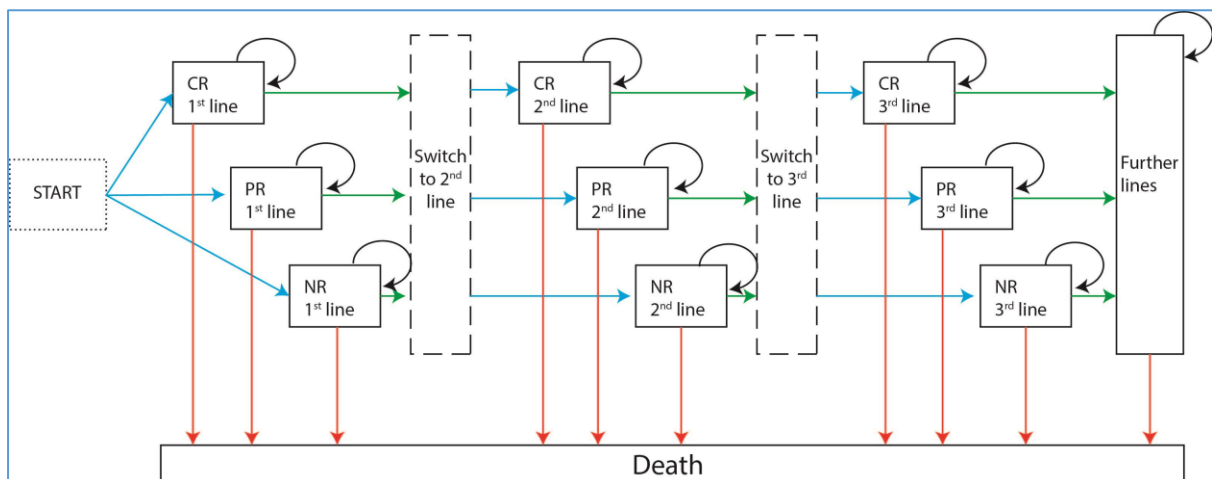
Abbreviations: BSC best supportive care; NSAI non-steroidal aromatase inhibitors (anastrozole or letrozole)

Heeg cancer model

The Heeg model was used for comparing fixed chemotherapy sequences of four treatment lines, for multiple myeloma ineligible for stem cell transplantation (Figure 7.3).²⁴³ The model was developed for comparing overall survival (OS) of 17 different treatment sequences. In clinical practise, complete response is used as a short-term marker for treatment success, and has been shown to be a predictor of overall survival in multiple myeloma.²⁴³ The first three treatment lines were therefore divided into three different response states, representing complete, partial, or non-response on each treatment line. The model also included a health state representing 'subsequent treatment lines', and the terminal state, death. A cycle length of one month was chosen as this is the shortest interval at which the patient's response to therapy is typically measured in clinical practice. At the start of the model patients were distributed over the three response categories to first-line treatment, based on the specific treatment they receive in first line. During each following monthly cycle, members of the cohort could remain in their current response state, switch treatment, or die. In the model, response rates were combined with the probability of switching treatment and the mortality probability. The probability of switching treatment was based on time to next treatment. Patients who switched treatment were then redistributed over the three response health states in second line, where they could again remain on treatment, switch treatment or die. Treatment-specific probabilities of response (complete, partial, and non-response) on each first-line treatment were taken from a network meta-

analysis of RCTs of newly diagnosed patients (Chapter 6, Section 6.5.3.1). The response specific probability of transition to the next treatment or death were obtained from a single trial (VISTA study)³⁷⁷ of newly diagnosed patients (*using a Weibull model*). The treatment-specific probabilities of response on both second and third-line treatments were obtained from a second network meta-analysis of RCTs of patients with relapsed or refractory cancer. The response and line specific probability of transiting to the next treatment or death were obtained from another trial (APEX trial)³⁷⁸ of relapsed patients (*using exponential survival curve, and assuming constant treatment switch and mortality transition over time*). The probability of dying whilst in the ‘further-treatment lines’ health state were obtained from a third trial (SUMMIT trial),³⁷⁹ which included patients who had already received many treatments before entering the trial. For consistency, the probabilities for treatment switching and mortality were derived from the results of the bortezomib arm in each of the three selected clinical trials.

Figure 7.3: A schematic diagram of the Heeg model: a response-based Markov cohort model for the treatment of multiple myeloma



Taken from poster presentation: Heeg, B., Van Agthoven, M., van Beurden-Tan, C., Liwing, J., Mellqvist, U. H., Plesner, T., Logman, F., Aschan, J., Einsele, H., Treur, M., Barendse, M., Richardson, P. G., Palumbo, A., Nahi, H., Sonneveld, P. (2014). PCN24 Treatment Sequencing Survival Model for Patients with Multiple Myeloma Ineligible for Stem Cell Transplantation (SCT). *Value Health*, 17(7), A617-8.

Abbreviations: CR complete response; NR non-response PR partial response

In all three models (Soini, Cameron, and Heeg cancer) the evidence used to inform time to treatment failure or progression free survival was line specific, but not generally treatment-specific.^{226 243 269} This is based on the assumption that response to a specific treatment is independent of response to the previous treatments used, and ignores any potential cross-resistance. The probability of switching treatment in the Heeg cancer model was both response type and line specific.²⁴³ However, the model assumed the same duration of response across all treatments, and although the probability of switching treatment for first-line treatment was based on a Weibull distribution that can allow the probability to increase or decrease over time, the probability of switching for second and third line treatments were based on an exponential distribution, and therefore assumed to be constant over time. The Heeg cancer model provides an example of an adaptation of the basic three state cancer

model, which accounts for both the different levels of treatment response and treatment sequences. Other Markov models used a similar approach to also account for other attributes relating to the decision problem.

Models using Markov states to represent response to each treatment line and other attributes

Four studies used Markov health states to represent a combination of individual treatment lines and other factors to better represent the decision problem.^{231 250 263 277} All the models were based on the basic premise that each treatment line (or specific treatment) was associated with two health state, treatment response or non-response. Non-response generally led to treatment switching, whilst a treatment response state was generally recursive, thus representing continued treatment. The models also included other states in order to monitor additional attributes, such as serious adverse effects, or make state transitions conditional on certain attributes, such as relapse. This approach is initially illustrated here using a study of psoriasis (Sawyer, 2013) and then the two cancer studies (Wong, 2009 and Lee, 2013). The Sawyer model adds to the narrative of modelling treatment sequences for psoriasis, and provides an example of modelling relapse.²⁶³ The study by Wong *et al.*, represents the only cancer study to account for the possibility of toxic death and all-cause mortality separately within the Markov model, by incorporating toxicity as a separate health state.²⁷⁷ It also accounted for the fact that patients who develop toxicity can continue on the same treatment at a lower dose. However, similar to Soini *et al.* it evaluated fairly limited sequences of up to three lines of active treatment. The Lee model was used to evaluate more extensive treatment sequences, and also accounted for the fact that treatments were only administered for a fixed period (Figure 7.4).²⁵⁰ However, the chemotherapy sequences being compared only differed in terms of the first-line treatment used. The model used in the fourth study (Davies, 2008), of treatment sequences for schizophrenia does not add anything new, and is summarised in the Appendices.²³¹

Sawyer model

The Sawyer model was used to investigate early treatments for psoriasis, which was described as a relapsing and remitting condition for which there was no evidence that response to treatment has an effect on its natural history.²⁶³ The model was used to inform the NICE clinical guideline 153. The model structure represented the movement of patients through a fixed sequence of three topical treatments used in primary care, over a one year time-horizon. The model included an additional health state, 'relapse', in order to account for the fact that a patient who initially responds to treatment and then relapses, may be eligible for the same treatment that was initially successful. Patients who failed to respond to all three treatments were referred to secondary care for more intensive treatments. Patients were assumed to undergo a maximum of eight weeks continuous therapy with a given topical agent, except corticosteroids, which were assumed to be trialled for four weeks. A cycle length of four weeks was used, and patients could respond within the first four weeks (early responder) or the second (late responder). The model appeared to include nine health states, depicting either a response or non-response to each treatment line, plus an additional health state representing relapse, which could occur whilst on any line of treatment. Only the responder states

were recursive, whilst relapse appeared to have been implemented as a temporary state through which patients could pass from each responder state back to the prior treatment. After receiving the first treatment, hypothetical patients entered the first-line responder state or the first-line non-responder state. Those who responded stopped treatment and either maintained response (*re-entering the first-line responder state*) or relapsed (*entering temporary relapse state*). Patients who relapsed resumed the initial treatment and, again, either responded or did not respond to treatment (*entering either the first-line responder or non-responder state, respectively*). Patients who did not respond (*in first-line non-responder state*) moved to the second treatment, to which they could respond (*second-line responder*) or not respond (*second-line non-responder*) etc. Probabilities of early and late response were treatment related but assumed to be independent of positioning in the sequence. All treatments were assumed to have the same relapse rate, which could occur at any point following response. The Markovian assumption also meant that patients in each treatment response state were treated as a homogenous group, and the patients' history ignored. For example, those in the first-line responder state would have had equal probability of responding or relapsing to treatment, irrespective of whether they have had a previous relapse or not.

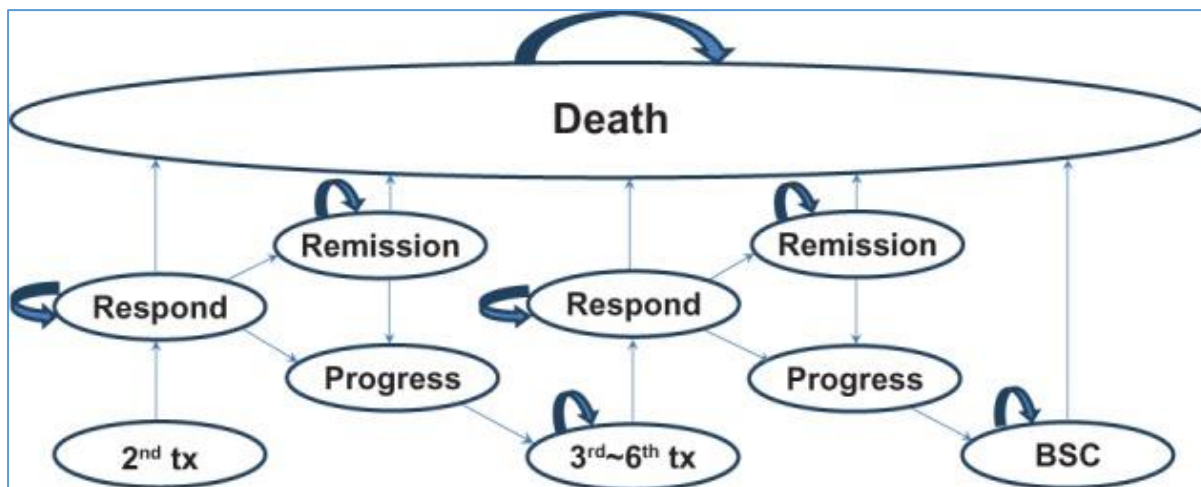
Wong model

The model by Wong *et al.* was developed for comparing fixed treatment sequences for newly diagnosed metastatic colorectal cancer.²⁷⁷ The model included toxicity and progression as separate health states. These can be conceived as competing risks, which were implemented in the Markov model by using a short one week cycle. At the end of each Markov cycle, patients could either remain on the same treatment, develop toxicity, experience progression, or die from all-cause mortality. Patients who developed toxicity could die, continue therapy at a reduced dose, or change treatment. Patients could have up to two toxic events. It was assumed that toxicity and progression were independent and mutually exclusive events within the course of a one week cycle. The main clinical effectiveness estimates were the rates of progression and toxicity. Progression rates, taken from clinical trials, were converted to weekly probabilities, and the data on the rates of Grade 3 or 4 toxicities were converted to the probability of fatal or non-fatal outcomes.

Lee model

The Lee model was used for comparing two chemotherapy sequences containing four treatment lines for platinum-sensitive ovarian cancer (Figure 7.4).²⁵⁰ The model included four health states: responsive, progressive, clinical remission, and death. 'Progress' was referred to as a tunnel state, but actually appears to have been a temporary state, linking the 'respond' and 'remission' states to the next active treatment or best supportive care. A cycle length of nine weeks was used, which reflected the timing of treatment response assessment in clinical practice. Patients remained on the same treatment for 18 weeks if they responded. Patients who did not progress or experienced adverse effects within this period entered the clinical remission state, withdrawing from the drug. Patients could progress whilst on treatment or from remission, and would then enter the next line of treatment.

Figure 7.4: A schematic diagram of the Lee model: a Markov cohort model for the treatment of recurrent ovarian cancer



Taken from: Lee, H-Y., Yang, B-M., Hong, J-M., Lee., T-J., Kim, B-G., Kim, J-W., et al. (2013). Cost-utility analysis for platinum-sensitive recurrent ovarian cancer therapy in South Korea: results of the polyethylene glycolated liposomal doxorubicin/carboplatin sequencing model. *ClinicoEconomics and Outcomes Research: CEOR*, 5, 297–307.

Notes: The different compartments are mutually exclusive health states. Arrows represent allowed transitions between states. “Progress/stable” state is “tunnel” state.

Abbreviations: BSC best supportive care; tx treatment

Models using Markov states to represent disease condition or treatment response

The use of Markov states to represent the patients’ disease condition or treatment response, rather than treatment lines, means that the model can account for different levels of disease activity or response. This approach was used by seven rheumatology studies (Albert, 2000; Coyle, 2006; Maetzel, 2002; Rodgers, 2011; Schipper, 2011; Welsing, 2005; Wu, 2012) representing five unique models,^{17 216 230 253 265 275 279} and four non-rheumatology studies (Beard, 2011; Orme, 2012; Shepherd, 2006; Smith, 2007).^{220 259 266 268} Treatment sequencing was generally implemented using a Markov cycle tree, where the terminal branches represent the Markov states, and provides the distribution of the cohort among these states at the end of a model cycle.

This approach is described here using mainly the studies of rheumatoid arthritis, which is a chronic condition with a varying disease course over time, characterised by periods of high disease activity alternating with low disease activity, or remission.²⁷⁵ The Albert model first provides an example of using a decision tree structure to implement treatment sequencing within a simple three state Markov model representing the response categories: improved, active disease, and toxicity. This initial model was subsequently modified by changing the improved state into a tunnel state in order to incorporate time dependency, which can vary between treatments. Maetzel *et al.*, used the Markov cycle tree in a slightly different way, in order to account for what happens to patients who may continue treatment despite not having a clinical response (Figure 7.5).²⁵³ The Welsing model,²⁷⁵ which was further adapted by Schipper, *et al.*,²⁶⁵ provides an example of where four Markov health states were used to account for the fluctuating levels of disease severity that can occur over the course of a patient’s

lifetime (Figure 7.6). Finally, the Shepherd model, which is the most complex Markov cohort model in the review is also described here as it provides a good example of how tunnel states can be used to overcome the Markovian assumption of no memory.²⁶⁶ The model was used to evaluate sequences of up to two lines of antiviral drugs for chronic hepatitis B, which is a complex chronic condition that includes two distinct patient subgroups, and individual treatments with varying profiles and durations of administration.

The remaining four models are described in more detail in the Appendices as they do not add any new information on modelling treatment sequences as such. However, they do provide examples of accounting for various additional factors in the model. The York psoriatic arthritis model accounted for the differential effect of the reason for withdrawal on the cost-effectiveness of a second TNF-inhibitor, by using separate subgroups for those who discontinued the first TNF-inhibitor due to adverse effects or intolerance.¹⁷ The Beard model, which was developed for evaluating the optimal positioning of duloxetine within a standard treatment sequence for fibromyalgia, accounted for three discrete-pain response levels as well as toxicity. It included the following five health states: full response, partial response, full response and intolerable adverse effects, partial response and intolerable adverse effects, and inadequate pain response. The Beard model also accounted for the fact that some patients decide not to have any further treatments. The model allowed for a proportion of patients to drop out of current treatment and be lost to subsequent treatments.²²⁰ The same percentage (25%) was used for all active treatments, and explored across a specified range (20-30%) in sensitivity analyses. The Smith model provides an example of a model based on the natural history of the condition, postherpetic neuralgia, the duration of which was not affected by the treatment.²⁶⁸ Treatment was considered for pain relief only, and it was assumed that failure to respond to one treatment would have no effect on likelihood to respond to others. The Orme model provides an example of the implementation of a fairly complex treatment pathway for glaucoma or ocular hypotension.²⁵⁹ The model allowed different subsequent treatment selection based on reasons for quitting the prior treatment. It also considered patients with both high and low level of risk within the same model, and the differential follow-up required by patients with mild and moderate glaucoma. It used four levels of glaucoma severity as discrete health states. The model structure was based around three triggers for switching treatment: treatment intolerance; intraocular pressure not meeting the benchmark; and progression in visual field defect. The choice of next treatment was implemented using a decision tree structure.

Albert model

The Albert model was developed for assessing the clinical effectiveness of different management approaches for rheumatoid arthritis.²¹⁶ This included the comparison of four fixed sequences of three drugs, representing strategies starting with either the least toxic or most effective drug first. The model allowed patients who did not respond to switch treatments, whilst those who responded or improved remained on their current treatment, which was initially modelled as a terminating state. In other words, it was assumed that once a successful treatment was found, its success would be maintained.

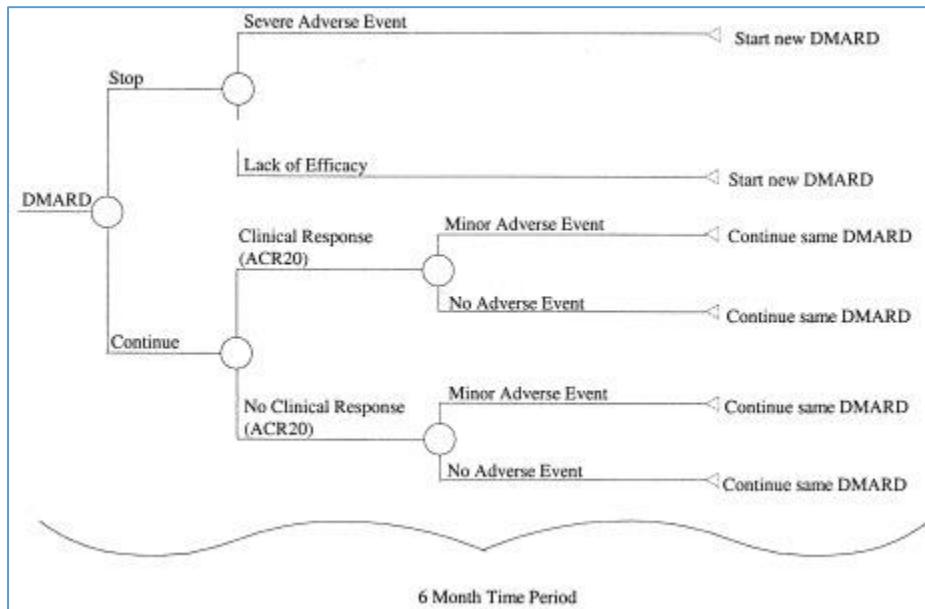
Every patient in the cohort was assumed to start on the first drug and then, after one cycle of treatment, moved to the 'improved', 'active' take second drug, or 'toxic' take second drug state etc. A cycle length of six months was used and at any given cycle the patients could be in any of the three health states. The model was used to estimate the mean time spent in the improved state for each sequence. Different sequences were compared by running the same model for each sequence, using specific transition probabilities for the different drugs but assuming treatment independence. The findings showed that by the third cycle, regardless of the sequence of drugs used, most individuals were improved. Although the rate of increase of patients in the improved category was greatest if the most effective agents were used first, the results converged after several cycles. The authors noted that this indicated that the model was not a good reflection of reality. They therefore develop an expanded Markov analysis, where the improved state was modelled as a separate tunnel for each drug instead of an absorbing state, and duration of therapy was modelled as a means to terminate the improved state. This meant that patients would have a period of time in the improved state and then be cycled back to the active or toxic states. The revised model was able to reflect the fact that probabilities for continuing the drug and develop toxicity vary with time, and for each drug. Tables with probability of remaining on one drug over time were used to calculate the transition probability of continuing a particular drug at each state.

Maetzel model

The Maetzel model was used to compare a sequence of conventional DMARDs, with or without leflunomide (Figure 7.5).²⁵³ The same model was used later by Coyle *et al.* for evaluating similar pre-defined sequences of conventional DMARDs with and without the addition of a TNF-inhibitor.²³⁰ Separate models were run for each sequence and compared in terms of the average time spent in the state of treatment response. Leflunomide was added after a sequence of up to three treatments containing methotrexate. Patients who were intolerant to methotrexate were allowed to cycle through a different treatment sequence, thus avoiding combination therapy with methotrexate. The model included two health states, 'continue same DMARD' and 'start new DMARD'. Treatment response was measured using the American College of Rheumatology (ACR) criteria. Different degrees of improvement are referred to as ACR20, ACR50, ACR70, which represent 20%, 50%, and 70% improvement, respectively, on a 28-level symptom scale. Patients discontinued treatment if they failed to achieve ACR criteria for 20% improvement (<ACR20) or experienced severe adverse effects. A sub-decision tree, for each six-month cycle, was then used to depict the patient pathway, to assess the reasons for discontinuation and subsequent events for patients who continued treatment. The initial decision node in the tree, representing 'DMARD' treatment, branched into 'continue' or 'stop treatment'. Stop treatment then branched, at a chance node, into severe adverse effect or lack of efficacy, with both branches ending in the health state 'start new treatment'. Whilst 'continue' branched into 'clinical response' (ACR50/70) or 'no clinical response' (ACR20), with both then branching into 'minor adverse effect' or 'no adverse effect'. All four branches stemming from 'continue' ended in the health state 'continue same DMARD'. This approach meant that the model could account for the fact that in clinical practice some patients would continue treatment despite having 'no

response', as per ACR20, with the decision not to modify treatment taken, for example, based on X-ray results. It also enabled the economic evaluation to include the cost of treating minor adverse effects in patients who continued treatment.

Figure 7.5: A schematic diagram of the Maetzel model: a Markov cohort model for the treatment of rheumatoid arthritis



Taken from: Maetzel, A., Strand, V., Tugwell, P., Wells, G. & Bombardier, C. (2002) Cost effectiveness of adding leflunomide to a 5-Year strategy of conventional disease-modifying antirheumatic drugs in patients with rheumatoid arthritis. *Arthritis & Rheumatism*, 47(6), 655–661.

Note: Decision analysis tree representing the conditions under which patients move from one branch (i.e., treatment) to another within a 6-month treatment cycle.

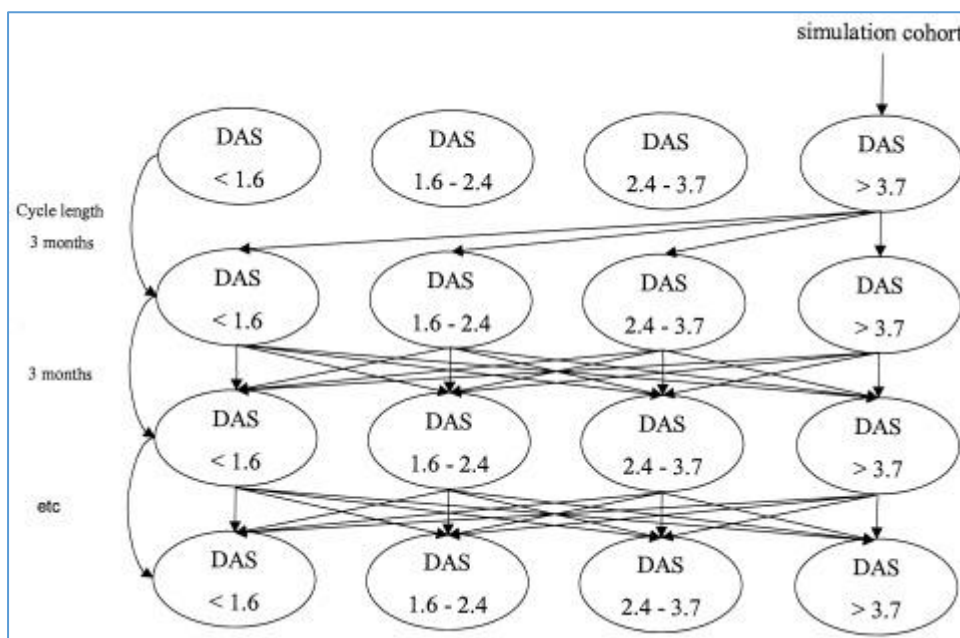
Abbreviations: DMARD disease-modifying antirheumatic drug.

Welsing model

The Welsing model was used for comparing fixed sequences of two active treatments followed by usual care (Figure 7.6).²⁷⁵ The active treatments included Leflunomide and a TNF-inhibitor. The same model was later adapted by Schipper *et al.* to compare three fixed sequences of five drugs in early rheumatoid arthritis.²⁶⁵ The initial model was not well reported. It included four health states that represented the patient's transition through different levels of disease severity based on disease activity score 28 (DAS28) thresholds (Figure 7.6). These included remission (DAS<1.6), low disease activity (1.6-2.4), moderate disease activity (2.4-3.7), and high disease activity (>3.7).²⁷⁵ The cycle length was three months, and the model ran for 20 cycles. In the initial model all patients started in the high disease activity state, reflecting the patient population starting TNF-inhibitors. In the case of non-response at three months, patients switched treatment. Each treatment sequence was modelled separately, using specific transition probabilities for the different drugs. The model outcomes included the expected percentage of time spent on each treatment and the patient years spent in the disease activity states. The percentage of time spent on the second treatment was equal to the estimate for the first treatment minus usual care, which was the same in both sequences. In the model presented by Schipper *et al.* the simulated cohort represented patients who were starting conventional

DMARDs, and were evenly distributed across all the states at the start of the model.²⁶⁵ At the end of each cycle patients could be either in remission and remain on their treatment for the next three months, or not in remission (with low, moderate, or high disease activity) and switch to the next treatment. This was implemented using a Markov cycle tree. The initial Markov tree node branched to each of the five individual treatments. Each treatment was then attached to a sub-tree representing a decision node between the four disease activity health states, three of which led to switching to the next treatment, whilst remission led to the same treatment and a repeat of the sub-tree. The percentage of patients achieving remission for each treatment was based on individual patient level data from two cohorts. Patients were assumed to sustain remission after being in remission for two cycles.

Figure 7.6: A schematic diagram of the Welsing model: a Markov cohort model for the treatment of rheumatoid arthritis



Taken from: Welsing. P. M., Severens. J. L., Hartman. M., van Riel. P. L., Laan. R. F. (2004). Modeling the 5-year cost effectiveness of treatment strategies including tumor necrosis factor-blocking agents and leflunomide for treating rheumatoid arthritis in the Netherlands. *Arthritis Rheum*, 2004, 51(6), 964-73.

Abbreviations: DAS Disease Activity Score.

Shepherd model

The Shepherd model was used to compare six fixed sequences of antiviral drugs for chronic hepatitis B, which is a fairly complex disease.²⁶⁶ Patients may remain asymptomatic for some time before developing symptoms of progressive liver disease, such as cirrhosis or hepatocellular carcinoma. They may also clear the virus spontaneously or move into remission. Patients can be either HBeAg positive or HBeAg negative, depending on the presence or absence of the e antigen. The aim of treatment differs between the two groups. Antiviral treatment for chronic hepatitis B can be divided into two classes which tend to differ in terms of their adverse effect and drug resistance profiles, and whether they are administered for a fixed period or maintained until treatment failure. The sequences

included a first-line immunomodulator followed by second-line nucleos(t)ide analogue, and in two sequences, a third-line salvage treatment with a nucleos(t)ide analogue.

Treatment sequencing was implemented using tunnel states, within a natural history model depicting the complex nature of chronic Hepatitis B. The model included patients with both HBeAg positive and HBeAg negative disease. The same model and structural assumptions applied equally to both disease variants but it was necessary to keep the two separate in the analysis as they had different distributions of age at diagnosis and transition probabilities between health states. This was also achieved using the tunnel states. The model was developed as a decision-tree Markov cycle model where all the destination states, other than death, consisted of up to 12 tunnel states, which were used to track history within the simulated patient cohort. Separate tunnels were defined for HBeAG-positive and -negative patients, which were then further subdivided to show whether the patients were resistant to either first or second-line drug, and whether they were continuing or had stopped treatment. The model had a cycle length of one year, with a half-cycle correction applied. The model included eight health states representing an asymptomatic condition, different treatment response, various progressive liver disease, and death. The Markov cycle tree included two subtrees, referred to as *clones*, which were attached to different locations, or nodes, in the tree. The 'Progression' subtree indicated all the possible states that an individual could progress to in the next cycle. The 'PreResistance' subtree showed the different management options for individuals who develop resistance, and indicated whether the patient would continue, stop or, if other antiviral agents were available, switch treatment after experiencing treatment resistance. Patients who did not develop treatment resistance during a cycle followed a branch called 'NoResist' and had outcomes evaluated as described in the progressive sub-tree. Patients who developed resistance followed the pre-resistance subtree. Each terminal branch of the pre-resistance subtree had a progressive subtree attached to it. During the progressive subtree, patients were first exposed to the probability of dying, based on age-specific all-cause mortality rates. The survivors were then exposed to the state-specific risks of seroconversion, remission, progressive liver disease, and state specific excess mortality risk. All the terminal branches, or destination states, except death were tunnel variables. Not all of the destination states were accessible from each starting state. For example, individuals with chronic hepatitis B were assumed not to progress directly to decompensated disease, and an individual with HBeAg-negative disease were not be able to undergo HBeAg seroconversion.

The model was used to evaluate up to two active lines of antiviral drugs. There are a number of reasons why patients with chronic hepatitis B would switch treatment in practice, including sub-optimal or non-response, intolerance, or cross resistance. The authors noted that switching in the model was 'due to treatment failure', but it was unclear how this was implemented as it appeared to have actually been based on treatment resistance. Furthermore, the two treatments used as first-line were immunomodulators, which are not associated with drug resistance. The clinical effectiveness estimates used to inform the model included HBeAg seroconversion rates and alanine aminotransferase (ALT) normalisation or remission rates, and the durability of these responses.

These estimates were taken from an accompanying systematic review, whilst the data on treatment resistance was not. Treatment sequencing effects were not considered, and the same effect estimate, for each treatment, was used irrespective of their positioning in the sequence. Developing resistance to a second treatment was considered independent of the fact that the patient had already developed resistance to the first treatment. If patients developed resistance to the second drug, it was assumed that they either continued, but with no therapeutic benefit assumed, or stopped all antiviral treatments, receiving best supportive care from then on.

7.3.3.4 Advantages and disadvantages of Markov cohort modelling

The Markov cohort approach provides a means of modelling treatment sequences over time, and can be easily implemented within commonly used spreadsheet or decision analytic software. It is a method that is well suited for modelling time to event and repeated events. Markov cohort modelling was successfully used for both identifying the optimum treatment sequence from all conceivable sequences, and for comparing pre-defined sequences. The former included an approach based on ordering treatments according to their net benefit per unit of time (or cost-effectiveness ratio per unit of time), estimated separately using a simple two state Markov model. An advantage of this approach is that the equation for estimating the expected net-benefit per unit time for the overall treatment strategy can be used for any given treatment sequence, thus avoiding the necessity of selecting predefined sequences in advance. It can also be used to determine the optimum strategy for an individual patient based on the treatment options that are suitable to them. However, this approach is valid for treatments that aim to provide symptomatic relief only, and do not alter disease progression. This approach is also based on the assumption that treatments only provide benefit whilst they are being administered, and that the treatment effect is independent of positioning in the sequence. It also ignores the potential progressive accumulation of patients that may not respond to any treatment.

Treatment sequencing, when comparing pre-defined sequences, were conceptualised within the Markov model in two ways. They were either implemented as a series of Markov states representing each line of treatment or, alternatively, using sub-decision-trees or algorithms, with the Markov health states used to represent the different levels of treatment response or disease severity that patients progress through. Additional health states were also used to control how patients progressed through the treatment sequences. Where certain transitions were dependent on particular attributes, such as disease progression,²⁶³ or relapse,²⁵⁰ these were incorporated as temporary or tunnel states, which patients must pass through to either re-enter the prior treatment, or enter the next line of treatment, respectively. The use of temporary or tunnel states also enabled the timing of events to be controlled. Some studies that used health states to represent each line of treatment also used additional health states to represent further attributes such as adverse effects. A potential limitation of the Markov model in implementing both treatment sequencing and additional attributes or health states is that patients *cannot be in more than one health state at the same time*. However, Wong *et al.* was able to achieve this by using a one week cycle length and the assumption that toxicity and progression were both independent and mutually exclusive events.²⁷⁷ This may not be practical for modelling a chronic

condition over a lifetime horizon. An alternative approach, which was used in the Maetzel model, is to incorporate details regarding, for example, the occurrence of adverse effects and the subsequent impact on treatment adherence or selection using a Markov cycle tree.²⁵³ Representing both the disease condition and treatment sequences as an exhaustive list of health states is likely to be challenging for modelling lengthy sequences with complex decisions relating to when and how treatment switching occurs. As the number of health states or potential attributes that need to be accounted for increases, such as the duration of treatment, different duration and levels of disease severity, treatment history, and reasons for discontinuing treatment, the less tractable the model will become. Identifying data to inform additional conditional transition probabilities will also be challenging.

The main disadvantage of using Markov modelling for evaluating treatment sequences is the Markovian assumption, which implies that patients will have an equal opportunity to respond well to each treatment they encounter, irrespective of the order in which they are used, or the *history of states visited*. Memory can be introduced by creating states that include history. Shepherd *et al.* used tunnel states to account for treatment history within a model developed for comparing treatment sequences for chronic hepatitis B.²⁶⁶ The model was primarily based on a Markov cycle tree, which consisted of two repeating subtrees. All but one of the terminal branches, or destination states, within one of these subtrees were tunnel variables. This allowed each disease state to consist of up to 12 tunnel states, which took into account previous treatment history and disease variant. The model was developed to consider a complex decision problem, but the sequences being investigated only included two lines of active treatment.²⁶⁶ The number of states required would greatly increase if more extensive sequences were considered, which would likely lead to a model that is difficult to handle.

Another challenge of using a Markov cohort modelling approach for evaluating treatment sequencing relates to how *time is incorporated in the model*. This is potentially relevant for modelling: i) changes in the responsiveness of the condition to treatment over time, ii) the time spent on individual treatments, iii) and duration of response. In a simple Markov chain model, time is represented as a constant variable, which means that it is unable to account for non-linear changes in responsiveness to treatment over time. This is adequate for modelling a stable disease or an ageing cohort but problematic for modelling a progressive condition with a varying disease course, such as rheumatoid arthritis. The duration of treatment or response, represented as the time spent within a state, is also unlikely to be constant over time. Most of the Markov models used for evaluating treatment sequences for cancer used structures based on line of therapy. In the Heeg cancer model, time-independent transition probabilities, based on the exponential survival curve, had to be applied to the second- and third-line treatments, thus assuming a constant treatment switching over time²⁴³ (I come back to this in Section 7.3.6.4 when discussing the Birmingham Rheumatology Arthritis Model). The authors noted that implementing time-varying transition probabilities in second and third line would have blown up the number of health states required, especially as the model also accounted for different levels of response, which can lead to different durations of response.²⁴³ However, the

counter-argument given to this was that, by incorporating line specific transition probabilities, time was implicitly included by defining specific transition probabilities. In the Heeg cancer model, the probabilities of moving from the health states representing response to first-line treatment were derived from a clinical trial using a Weibull survival model. However, the duration of response was assumed to be the same across all treatments.²⁴³ The probabilities for the level of response in the model (partial, complete, or non-response) were both treatment and line specific, whilst the duration of response was line specific only (Chapter 6, Section 6.5.3.1). Using simple Markov model structures based on line of therapy also potentially ignores the cross resistance between drug classes, which may need to be taken into account for some targeted therapies.²⁴³ The Markovian property *assumes homogeneity within each individual health state*, and limits the model's ability to account for any variability, such as the time spent within a state.²¹⁶ This means that the probability of treatment failure or experiencing serious adverse effects, which usually dictate transition to another state, cannot depend on the length of time spent in the current health state, whether it is used to represent treatment response or a specific treatment line. A variation in time dependency can be incorporated into the Markov cohort model using tunnel states. For example, Albert *et al.*, who used a simple three-state Markov model with treatment sequences implemented using a Markov cycle tree, revised their initial model to account for time spent in the 'improved' state. It was originally modelled as a terminal state, whilst the 'active' and 'toxic' states resulted in treatment switching. By modelling the 'improved' state as a separate tunnel for each drug, the model was able to reflect the fact that probabilities for continuing treatment and developing toxicity varied with time, and for each drug.²¹⁶ The York psoriasis model, based on the net benefit per unit of time, also provided an approach that allowed the treatment trial period to vary between treatments, and the failure probabilities to depend on the duration of treatment for individual treatments. However, both these models were based on the assumption of treatment independence. In fact, Albert *et al.*, deduced that treatment non-independence was the likely cause of their original model overestimating the effectiveness of treatment sequences.²¹⁶ The added requirement of modelling treatment sequences when incorporating time-dependency using standard tunnel states can potentially make the process complicated to programme in a spreadsheet or decision analytic software.²⁷⁶ Alternative approaches used for incorporating time dependency in a state transition cohort model included the use of a partitioned survival modelling and semi-Markov cohort modelling.

7.3.4 Partitioned survival models

7.3.4.1 Description of studies using partitioned survival

Three included studies used partitioned survival modelling, also known as an area under the curve model. Two economic evaluations that were undertaken on behalf of NICE used partitioned survival analysis for evaluating treatment sequences for advanced colorectal cancer. One study (NICE CG131), which was discussed previously (Section 7.3.2.2), used partitioned survival analysis within a decision tree modelling framework.²⁰⁷ The second study (Hind, 2008), which was a technology appraisal (NICE TA93), used the area under the curve to calculate the survival benefit of one treatment sequence over another.²⁰⁴ They limited inclusion to two prospective trials of treatment

sequences in order to obtain estimates of overall survival. One rheumatology study (Schadlich, 2005) also appears to have used an area under the curve state transition model.²⁶⁴ The model was described as a stochastic simulation model with clinical effectiveness parameters quantified by the model based on area under the curve calculations. The study included a cost utility analysis of adding leflunomide into sequential therapy consisting of the most frequently used conventional DMARDs over a three year time horizon.

A theoretical modelling approach for evaluating treatment sequencing for advanced cancer using partitioned survival within a Markov cohort framework was presented by Briggs *et al.*¹⁹⁵ The novel approach was proposed as a means of allowing time dependency in the time to event analysis within a cohort model framework, thus avoiding the disadvantages of resorting to individual simulations. They used a Weibull model to characterise time to event, but any parametric survival model can be used. The authors proposed that this provides a framework that is flexible enough to capture treatment effects that vary by line of therapy, and enables appropriate discounting that allows for differential timing to be made. However, this approach was only reported as an abstract and is yet to be applied to a real decision problem. Similar to the other included cancer studies that evaluated treatments sequences of more than two lines, the analysis was based on progression free survival for each treatment line, which is unlikely to be treatment-specific.

The partitioned survival modelling approach is illustrated here in more detail using the Schadlich model.²⁶⁴

Schadlich model

The study by Schadlich *et al.* compared fixed treatment sequences, with and without leflunomide, in terms of the time a patient benefited from treatments, which is comparable to progression free survival used in cancer studies.²⁶⁴ Treatment sequences were implemented in the model using a series of treatment-specific health states. The model structure was based on a previous unpublished model developed by Cambridge Pharma Consultancy, and was not well reported in the current publication. The patient cohort progressed through a series of six DMARDs in a fixed order over a three-year period, in six-monthly intervals. Patients switched to the next treatment, due to loss of effectiveness or adverse effects, at the start of each successive interval based on DMARD specific proportions of patients remaining on treatment at the beginning of each interval. At each treatment interval, patients were either treated initially with a given DMARD or received follow-up treatment with the same DMARD as the preceding interval. The differentiation allowed the application of quantified costs and effectiveness parameters. The respective values relating to the first six months were applied to patients assigned to initial treatment within an interval, and the values relating to the following six months were applied to patients assigned to follow-up treatment within the interval. The interval related representation was necessary for proper discounting of costs and effects in the second and third year of treatment.

Clinical effectiveness was based on response years (RYs) according to the ACR criteria for 20%, 50% and 70% improvement. Different degrees of improvement and corresponding response years (ACR20/50/70RYs) were quantified by the model using the area under the curve for the proportion of patients with each ACR response within a given interval. Quality adjusted life years (QALYs) were also quantified in a similar way. Data on ACR response rates were taken from two RCTs for leflunomide and methotrexate, and a meta-analysis for the remaining four DMARDs relative to methotrexate. Termination rates were taken from observational studies with a minimum of three years follow-up. The area under the curve survival functions appear to have been determined graphically. A survival plot, depicting the decrease in the proportion of patients remaining on treatment for each DMARD, at the beginning of each successive six-month interval over 3 years, was presented.

In the base case analysis the same treatment-specific effect estimate was used for each treatment irrespective of when they were used in the sequences. However, the impact of sequencing effects was explored in scenario analyses, as part of the sensitivity analyses. Response rates were reduced by 25% where four DMARDs, excluding leflunomide and methotrexate, were used as second or subsequent line (Chapter 5, Section 5.6). All other parameters were kept the same. Decreased treatment response rates could not be achieved for leflunomide and methotrexate due to the kind of data available and the underlying model framework, which estimated the effectiveness of four DMARDs based on data on their relative effectiveness compared with methotrexate. The retention rates, or the probability of remaining on each treatment depicted in the survival graph, were also decreased by 20% per successful interval for every DMARD given as second-line or later in the sequence.

7.3.4.2 Advantages and disadvantages of partitioned survival analysis

The partitioned survival approach provides a simple cohort framework for modelling treatment sequences, which allows the treatment effects to vary by line of treatment. It provides a non-data-intensive approach but does not allow for other separate attributes, which may need to be accounted for when modelling a complex decision problem. The Schadhlich model approach was able to account for the decreasing probability of remaining on a given DMARD with time.²⁶⁴ Treatment sequences were implemented as a series of health states, representing six consecutive DMARDs that patients could migrate through over time, which was divided into six intervals of six months. The model was also able to partially account for sequencing effects based on simplifying assumptions. However, the study was unable to account for disease duration and did not consider the impact of adverse effects or other attributes.

7.3.5 Semi-Markov cohort process model

7.3.5.1 Description of studies

An alternative approach to modelling the probability of a patient making a given transition dependent on the time spent in the current state is to do so externally. An example of this is provided by the York epilepsy model, which used a semi-Markov process to build time dependency into the Markov cohort

model by using an additional time dimension.²⁷⁶ The use of a multi-dimensional transition matrix allows the transition probabilities to depend on more than just the start and finish state, giving a non-Markov cohort state-transition model.³⁷³ The York epilepsy model was undertaken as part of a NICE technology appraisal (NICE TA76) of 'newer' anti-epileptic drugs for treating epilepsy in adults.^{276 380} The available evidence for epilepsy indicated that the probability of a patient changing treatment decreases as the time they have been on a given treatment increases^{18 27 380} (See Appendix Volume I, Section C5). This means that assuming a constant probability of treatment failure over time, based on the clinical trial data, would overestimate treatment discontinuation.¹⁸ The Markov model therefore required the transition probabilities to be a function of both the current state and the time spent in the current state.

An integrative model was developed for comparing antiepileptic drugs used for both newly diagnosed and refractory epilepsy, which were analysed separately. The treatment sequences included three active treatments, first-line monotherapy, second-line monotherapy, and third-line adjuvant therapy. Patients failing combination therapy were assumed to go on to receive maintenance therapy. The same model structure was also used for the *de novo* economic evaluation of drug sequences for epilepsy, conducted as part of NICE clinical guideline number 137, which was implemented using a Markov cohort model.¹⁸ The NICE CG137 model is summarised here in order to highlight the differences between them. Although the York epilepsy model accounted for the differential time spent on a treatment, it was based on the assumption that treatment response did not vary according to positioning in the sequence, whilst in the NICE CG137 model the response to second-line monotherapy was not independent of the response to the first monotherapy. The NICE CG137 model also accounted for four different treatment outcomes, whilst the York model was based on response or non-response to each treatment. However, both studies modelled fixed drug sequences used over a 15-year time horizon, and both models also accounted for differential epilepsy mortality linked to whether the patients were seizure free or not seizure-free. The two models are described in more detail below.

York epilepsy model

Treatment sequences were implemented in the York model as a series of health states representing starting and continuing each line of active treatment.²⁷⁶ The progress of the cohort through the model was tracked externally in the statistical software R, which can support multi-dimensional arrays. Time spent in the current state was implemented as a two-dimensional matrix, where one-dimension represented the current (*starting*) treatment state, and the other, the time or number of cycles spent in the current state. This enabled the proportion of patients in each state for each duration to be recorded, and transitions to the future state, representing the third dimension, to be conditional on both the current state and time spent in the current state.³⁸⁰

The model included eight health states in total: start monotherapy for newly diagnosed patients; continue monotherapy for newly diagnosed patients; start monotherapy for refractory patients;

continue monotherapy for refractory patients; start combination therapy; continue combination therapy; maintenance therapy; and death. Depending on the initial state, patients could move through the first seven states in sequence. The model used a cycle length of six months. Patients only spent one cycle in the three starting active treatment states, and distribution within these states was established within the first cycle. After the first six-month cycle patients either moved to the corresponding continued treatment state or the next active treatment, based on treatment response. The probability of response during the first cycle were taken from stratified meta-analyses for each of the three indications: monotherapy for newly diagnosed patients, monotherapy for refractory patients (representing second-line treatment), and combination therapy. The definition for 'response' incorporated remaining on the drug until the end of the trial. The probability of continuing treatment during subsequent cycles was based on not failing treatment. The probability of treatment failure, i.e. not remaining on treatment, was taken from observational data, which was not specific to the drug under consideration.²⁷⁶ This included a longitudinal cohort study, the National General Practice Study of Epilepsy, for monotherapy and an open label follow-up study of tiagabine for combination therapy.

NICE CG137 (2011)

The NICE CG137 model was created using TreeAge Pro 2008.¹⁸ The model was described as a multistate Markov model, which was built to reflect transitions between health states defined by the outcomes of each treatment line (starting each treatment was represented as events). First line monotherapy was associated with four treatment outcomes, whilst second-line monotherapy and adjuvant therapy had three. All three treatment lines were associated with following three outcomes: complete seizure freedom, referred to as remission; a 50-99% reduction in seizure frequency; and treatment failure. Treatment failure was further subdivided, for first-line monotherapy, into inadequate seizure control and unacceptable adverse events. This allowed the differential choice of second-line treatment depending on the reason for discontinuing first-line treatment. The probability of achieving remission or treatment failure for monotherapies was informed by a network meta-analysis of RCTs, which incorporated individual patient data, whilst treatment failure and remission of adjuvant therapies were informed by a meta-analysis of placebo-controlled trials. The probability of treatment failure in subsequent cycles were obtained from observational and open-label clinical trial data that was not specific to the drug being considered. The data were interpolated from the published graphs presented by Wilby *et al.*,²⁷⁶ and converted to specific six-month transition probabilities.

Similar to the York model, comparison of first-line monotherapies and adjuvant therapies was done separately using an integrative model. The first probability faced by the newly diagnosed hypothetical cohort was that of withdrawing due to adverse effects. Patients who did not fail treatment at the end of the first cycle next faced the probability of achieving remission. Patients who did not achieve remission or treatment failure in the first cycle were assumed to have a 50-99% reduction in seizure frequency. Patients who achieve seizure freedom were assumed to continue with the same drug treatment for subsequent cycles. Patients who experienced a reduction in seizure frequency continued for up to two cycles with the potential of achieving seizure freedom on the same drug.

Patients who experienced treatment failure at any point switched to the next treatment. Patients failing first-line monotherapy due to inadequate seizure control were assumed to be 75% less likely to achieve remission with second-line monotherapy. For patients who failed first-line due to intolerable side effects, it was assumed that response to the second monotherapy was independent of response to the first-line drug. Patients who failed treatment with a second monotherapy were assumed to move on to adjunctive therapy, with no differentiation made as to the reason for treatment failure. Treatment response for adjuvant therapy was usually defined in the clinical trials as the achievement of greater than 50% reduction in seizure frequency. This outcome was divided into patients achieving seizure freedom, and patients achieving 50% to 99% reduction in seizure frequency. Patients who achieved partial seizure freedom on adjuvant therapy could continue to respond to treatment, or fail treatment and move directly to the 'no response' maintenance treatment state, during subsequent cycles. However, they could not transit to the seizure-free state. As with the York epilepsy model, patients could transit to death at any time.

7.3.5.2 Advantages and limitations of using a semi-Markov cohort approach

The semi-Markov process model based on the use of external multidimensional matrices provides a straightforward approach for incorporate time-dependency within the cohort framework for modelling treatment sequences. This was achieved by the York epilepsy model, which represented a simple model that did not require excessive data for parameterisation.^{276 380} It was used to model fixed treatment sequences where each treatment line was characterised as two health states, an initial temporary state representing the first six-month treatment period, and a recursive state representing continued treatment use. The probability that the patient would progress to the next treatment state was allowed to vary according to how long the patient had been in their current state. However, the model was not suitable for evaluating treatment sequences where the treatment effectiveness is dependent on positioning in the sequence. It also accounted for a fairly simple decision problem based on only one level of treatment response, where the patients had complete seizure freedom or not. It did not account for the impact of toxicity and it was assumed that all patients who did not respond would progress to the next line of treatment. The selection of add-on treatment was not dependent on specific treatments previously used. The NICE CG137 model, which used the same structure as the York model, was able to account for some of these complexities. However, the authors had access to a slightly better evidence base for the clinical effectiveness of first-line monotherapies, which included individual patient level data. In theory, multidimensional matrices can also be used to reflect 'patient history' or previous treatments into cohort models. However, no included study did this.

Theoretical development: Nested Markov cohort model

An alternative approach for incorporating time-dependency within a Markov cohort framework was presented by Shah, *et al.*, based on the use of nested Markov models.³⁸¹ However, this theoretical approach was only reported in a conference abstract. The hypothetical model was originally developed using Excel, and subsequently tested and implemented as a probabilistic decision model

using R, based on the York epilepsy model and data. The model structure was identical to that of the semi-Markov York epilepsy model.^{276 380} The method involved first disaggregating the model by treatment, first-line monotherapy, second-line monotherapy, and third-line adjuvant therapy, and then calculating the net present value for each treatment. The nested model was then rolled back into the treatment sequence by combining the net present values of each treatment weighted by the proportion of time spent in the sequence. These were then combined to calculate total outcomes. The authors concluded that the nested Markov modelling approach represented a straightforward and intuitive approach to modelling a fixed treatment sequence, but it may not be suitable if the position in a sequence is interchangeable and treatment effectiveness depends on the position in a sequence, for example cancer therapies where disease progression impacts treatment effectiveness.

7.3.6. Individual patient simulation

7.3.6.1 Description of studies using Individual patient simulation

An alternative approach, to using tunnel states in Markov cohort models or semi-Markov modelling, for incorporating time dependency or patient histories is the use of individual patient simulation. Twenty rheumatology studies^{218 219 224 225 227 229 232 234 235 238 241 245 246 248 251 254 256 272-274} and four non-rheumatology^{199 233 242 244} used individual patient simulations, based on a closed cohort of patients.

Six individual patient simulation models were described as being a Markov or state transition model.^{218 232 241 246 248 256} A further seven studies that merely reported using an individual sampling model appear to have used a similar approach based on discrete time intervals, and are therefore summarised here under the state transition model heading.^{224 225 234 235 238 244 274} Ten individual patient simulation models were reported as being discrete event simulation (DES).^{219 227 229 233 242 245 251 254 272 273} One further study that used an individual sampling model and does not appear to have used discrete time cycles, but instead sampled from time to event distributions is summarised under discrete event simulation.¹⁹⁹ The categorisation of included studies according to model type is presented in Tables 7.1 and 7.2. A schematic diagram of a state transition individual patient simulation model is provided in Figure 7.7 using the Diamantpoulus model as an example. A schematic diagram of a discrete event simulation model is provided in Figure 7.8 using the Birmingham rheumatoid arthritis model (BRAM) as an example.

The implementation of treatment sequencing can be complex, with many decisions dependent on the attributes of individuals, such as response to previous treatment and disease duration. An alternative approach to account for the various pathways that can potentially be taken is to use an individual patient simulation. These can also account for a heterogeneous patient population, or be used to track the progress of specific subgroups within the model. Individual patient simulation allows the patients' history to be memorised. This can potentially be used to facilitate simulating treatment selection based on past patient characteristics, for example, previous treatment or reasons for discontinuing previous treatment.

7.3.6.2 State transition individual patient simulation model

Thirteen studies used a state transition individual patient simulation modelling approach, which is described here using the rheumatology studies. The twelve rheumatology studies (Bansback, 2005; Brennan, 2004; Brennan, 2007; Davies, 2009; Diamantopoulos, 2012; Diamantopoulos, 2014; Finckh, 2009; Hallinen, 2010; Kielhorn, 2008; Kobelt, 2011; Merkesdal, 2010; Wailoo, 2006) represented seven distinct models (Table 7.1).^{218 224 225 232 234 235 238 241 246 248 256 274} This included a series of models developed by a group of health economists at Sheffield University, which have been collectively referred to as the Sheffield rheumatoid arthritis models.³⁸² These include the Sheffield Etanercept model,²²⁴ the Sheffield BSRBR model,²²⁵ and the Sheffield AHRQ model.²⁷⁴ The Bansback model also represents consultancy work by the Sheffield team.²¹⁸ The Sheffield models were also used, or further developed by other researchers (Davies, 2009; Finckh, 2009),^{232 238} Treatment switching was implemented in these models using a Markov tree (Section 7.3.3.3). Treatment sequences were represented within two other distinct models, as a series of health states (Section 7.3.3.3) that patients progressed through in a fixed order. However the modelling concept was similar to that of the earlier Sheffield models. This included the Diamantpoulus model (Figure 7.7)^{234 235} and Kielhorn model,^{246 256} which were based on a fairly similar approach. These models are summarised here using the Diamantpoulus model (Figure 7.7) as an example. The final Kobelt model was based on health states representing the combined effect of function and disease activity.²⁴⁸ The model also accounted for reducing the dose of etanercept, for those who achieve remission. The non-rheumatology study used an individual patient simulation model to investigate the cost-utility of antiretroviral therapy (ART) strategies.²⁴⁴ This is referred to in Table 7.2 as the Holmes model. The model was not well reported and is summarised in the Appendix Volume II (Appendix E).

All the studies were cost-utility analysis and all reported probabilistic sensitivity analysis. All but one study used a lifetime perspective, whilst the Kobelt model was based on a 10-year time horizon. All the Sheffield models, the Diamantpoulus model, and the Kielhorn model were developed in Microsoft Excel. The Finckh model was implemented using R project statistical software. The software used for the Kobelt model was not reported. The Holmes model was implemented using TreeAgro software.

Treatment for rheumatoid arthritis typically involves the use of a number of different therapies over the long term, with some patients having long periods of successful improvement in symptoms, whilst others withdraw more quickly due to lack of efficacy or adverse events, and move on to an alternative treatment.³⁸² Individual patient simulation was considered the most appropriate approach for modelling treatments for rheumatoid arthritis by the Sheffield team, due to this typical sequential use of treatments over time and the uncertain duration of effect on each patient. Patients' progression in the model is conditional on their past, which can include the number of previous treatments and disease level. The most complex model structure was implemented by Finckh *et al.* who noted that, although the model could be implemented within a Markov modelling framework, an individual sampling model was chosen as the probabilities were based on the historical factors.²³⁸ A Markov

model would have required a great number of tunnel states to 'remember' history, which is simple to implement in an individual sampling model.

Most of the included individual sampling models for rheumatology were based around monitoring the patient's disease activity over their whole lifetime. Most monitored changes in the patient's health state over time using the Health Assessment Questionnaire Index (HAQ). This is a functional measure that is used as a proxy for the patient's disease activity. There is currently no specific health related quality of life measure for rheumatoid arthritis,³⁸² and the advantage of using HAQ is that it can be converted to health related quality of life using any number of published lineal regression functions. Most studies were based on the HAQ disability index sub-scale, which has been abbreviated here, for simplicity, as the HAQ score. A reduction in HAQ score represents improvement.

Sheffield etanercept model and Bansback model

The Sheffield etanercept model, presented by Brennan *et al.* (2004), represents the earliest model by the Sheffield team.²²⁴ It also represents the basic structure used by many of the rheumatology studies that used state transition individual patient simulation. It was developed to address the same decision problem as the initial rheumatology model developed by the Birmingham team (NICE TA36), which is summarised under discrete event simulation. The etanercept model was used to compare a fixed sequence of conventional DMARDs with and without etanercept added to the start. Three named conventional DAMRDs was used as an exemplar sequence. The model focused on the progression of the HAQ score over time. It represented a simple three state model implemented using a Markov cycle tree to depict whether, for each treatment, patients continued on the same treatment, switched to the next, or died. For each patient entering the model, the following baseline characteristics were sampled from appropriate distributions: age, gender, HAQ score, disease duration, and number of previous conventional DMARDs. A cycle length of six months was used, with the patient's HAQ score and mortality evaluated at the end of each cycle. After the initial treatment period, patients could be an 'initial responder', based on the ACR20 threshold, or 'non-responder'. Patients who did not respond switched to the next treatment. Initial responders remained on treatment for several six-month cycles, until subsequent longer-term withdrawal due to loss of efficacy or adverse effects. Patients could also die during any cycle, based on the age, sex, and HAQ score achieved at that point. The first treatment period was set at three months, so that the model could explicitly examine the percentage of withdrawal from etanercept at that point, based on clinical guidelines that suggest that a TNF-inhibitor should be withdrawn at three months if no response.

The mathematical approach used by Bansback *et al.* was built on both the Sheffield etanercept model and the Birmingham preliminary model.²¹⁸ The model structure was similar to that of the etanercept model, with the addition of a separate health state representing serious adverse effects. This time the sequence of three conventional DMARDs was implemented using a generic DMARD at each line (Chapter 6, Section 6.5.3.2). At the end of each six-month cycle both non-responders and those who experienced severe adverse effects withdrew from their current treatment and moved to the next one,

whilst responders continued on the same treatment. Patients who were classified as treatment success or non-responders were also assessed for the occurrence of mild to moderate adverse effects in order to account for their potential treatment cost. The model used both ACR20 and ACR50 as a response threshold for determining treatment success, which were implemented separately. Another study, reported by Davies *et al*, which used the same model, incorporated an additional stage where the initial treatment response was categorised into ACR intervals: ACR0-20, ACR20-50, and ACR70-100. Each level of response was associated with a given reduction in the patient's HAQ score from baseline. Achieving the ACR50 threshold was used to determine the acceptable response for continuing treatment. This study also differed in that patients were randomly sampled to experience several alternative sequences. The study evaluated five treatment sequences in total, one of which included sequential TNF-inhibitors (Table 6.4).

These models represent the use of a simple modelling structure, with the main advantage of using individual patient simulation being the ability to track a large number of patient histories, which vary in terms of response to treatment, risk of withdrawal, adverse events, and mortality. Subsequent models developed by the Sheffield team, including the Sheffield AHRQ²⁷⁴ and Sheffield BSRBR²²⁵ models, were able to further build on this by using a series of multivariate analyses to track individual patient pathways. This allowed model parameters to be based on covariate adjustment, which was made possible by the availability of individual patient data obtained from patient registries. This approach is illustrated here using the Sheffield AHRQ model. The AHRQ model was further developed by Finckh *et al.* to model a more complex decision problem relating to treatment sequencing for rheumatoid arthritis.²³⁸ Nick Bansback, from the Sheffield team, was a co-author and contributed to the development of the Finckh model.²³⁸ The Sheffield BSRBR model presented by Brennan *et al.* (2007)²²⁵ addressed the same decision problem as Chen *et al.*,²²⁷ which used the Birmingham Rheumatoid Arthritis Model (BRAM). It represents an independent evaluation of TNF-inhibitors submitted to NICE as part of the TA130, and was undertaken in collaboration with the British Society of Rheumatology Biologics Registry (BSRBR). The authors had access to individual patient level data from the BSRBR dataset. The use of sequential TNF-inhibitors was examined as part of further analysis. However, given the absence of correlation identified in the BSRBR data, the response and utility gain from a second TNF-inhibitor was assumed to be independent of the response to the first (Chapter 6, Sections 6.5 and 6.6). The model is described in more detail in Appendix Volume II (Appendix E).

Sheffield AHRQ model

The model presented by Wailoo *et al.*, was commissioned by the Agency for Healthcare Research and Quality (AHRQ) in the United States.²⁷⁴ It was developed primarily for comparing individual biological treatment strategies, including three TNF-inhibitors. Separate models were run for each biological agent. The model was then used to assess the cost-effectiveness of adding a second or third TNF-inhibitor, compared to the using a single TNF-inhibitor. However, this additional analysis was only presented in the AHRQ monograph (Wailoo, 2006),²⁷⁴ and not in the subsequent peer-

reviewed paper (Wailoo, 2008),³⁸³ which is generally used as a reference for the AHRQ model. The model was described in more detail for the evaluation of treatment strategies containing single biological agents. The model pathway was based on tracking the patient's HAQ score over time. The model started with the individual's baseline characteristics, from which a representative sample of 10,000 patients were simulated. The patient characteristics included, among others, duration of disease, comorbidities, baseline and current HAQ, and type and number of previous conventional DMARDs. Alternative strategies were compared using the same patients. A series of regression analyses were used to estimate the parameters that the model used for simulating the path each individual patient would take. Individual patients were followed from the time of starting treatment on a biological agent until death, with changes calculated every six months. For each individual treated with a biological agent, the model simulated the level of ARC response (statistical model 1) and the HAQ score archived after six months (statistical model 2). The HAQ improvement score for each patient was estimated as a function of the type of responder they were, their HAQ baseline value, and other covariates, including disease duration. The model then simulated the long-term HAQ score for each patient at six-month intervals (statistical model 3) until treatment withdrawal due to either loss of efficacy or adverse effect (statistical model 4, duration of treatment). At the time of withdrawal the HAQ was assumed to deteriorate by the same magnitude as the initial six month improvement. After withdrawal, the patients were assumed to return to conventional DMARD-based treatment, and their long-term HAQ score was modelled over six-month intervals for the remainder of their lifetime (statistical model 5). The HAQ score for each patient was then translated into QALYs (statistical model 6), and the cost for each patient calculated (statistical model 7).

The same model structure was then used for evaluating treatment strategies containing sequential TNF-inhibitors. At withdrawal from the first TNF-inhibitor, patients moved to the next TNF-inhibitor in the sequence, and statistical model 4 was followed by statistical models 1, 2, and 3 for the next TNF-inhibitor in the sequence until withdrawal from the final TNF-inhibitor in the sequence. At that point, HAQ progression was estimated using statistical model 5. For the analyses, it was assumed that a TNF-inhibitor would be tested for at least six months before a decision to withdraw is made. It was also assumed that treatment response was independent of the position of the TNF-inhibitor in the sequence (Chapter 6, Sections 6.5).

The statistical models governed how the HAQ score changed between each event, the time a patient withdrew from the TNF-inhibitor, the time of death, and the cost and QALYs incurred over the patient's lifetime. The statistical models took into account the characteristics of the individual patient and the TNF-inhibitor. The model was based on two key data sources, which included the National Data Bank for Rheumatic Diseases (NDB) and a meta-analysis of RCTs (Nixon, 2007, introduced in Chapter 5, Section 5.4.2).²⁰⁸ The probabilities of treatment response for the TNF-inhibitors were estimated in two ways, using the two data sources (discussed in Appendix Volume I, Section D2.1). The remaining statistical models were based on data from the NDB, with rheumatoid arthritis life tables used to estimate time of death for each patient.

Finckh model

The Finckh model was developed for comparing an early versus late introduction of TNF inhibitors in very early rheumatoid arthritis.²³⁸ Treatment strategies contained a sequence of two or three treatment groups, representing NSAIDs, conventional DMARDs, and TNF-inhibitors, with the last two groups containing three consecutive drugs represented by generic treatments (the sequences are illustrated in Table 6.4). The model was based on the AHRQ model, but was able to account for further complexities in the decision problem. Here the course of the disease was modelled using both HAQ and radiographic evidence of structural damage. The exact route a simulated patient took depended on i) the treatment strategy and the same patient characteristics used by the AHRQ model, and ii) the type of disease progression they were likely to follow. It was assumed that patients followed one of three disease courses, which could not be predicted at presentation. This included drug free remission, a mild course with slow progression, or rapid disease progression. The rate of disease progression was applied to each disease course pattern. Patient's response to treatment was categorised as excellent, good, moderate, or none. Initial HAQ improvement at six months depended on treatment used, response, disease duration and radiographic damage. The rate of HAQ score progression was dependant on both response to treatment and radiographic damage. The probabilities of treatment response for TNF-inhibitors and conventional DMARDs used as early treatments were based on RCTs in early rheumatoid arthritis. The response rates for six months after therapy were based on the regression analysis of patient registry data conducted by Wailoo *et al.*, which included covariates representing, among others, disease duration, number of DMARDs failed, and HAQ score at start of treatment. This was to account for the fact that response rates decrease with greater disease duration and number of previous treatments. However, the same generic effect estimate was used for each consecutive treatment within each treatment group, thus assuming treatment independence, and that the response to TNF-inhibitors would be the same whether the patient had previously responded to a TNF-inhibitor or not (Chapter 6, Section 6.5).

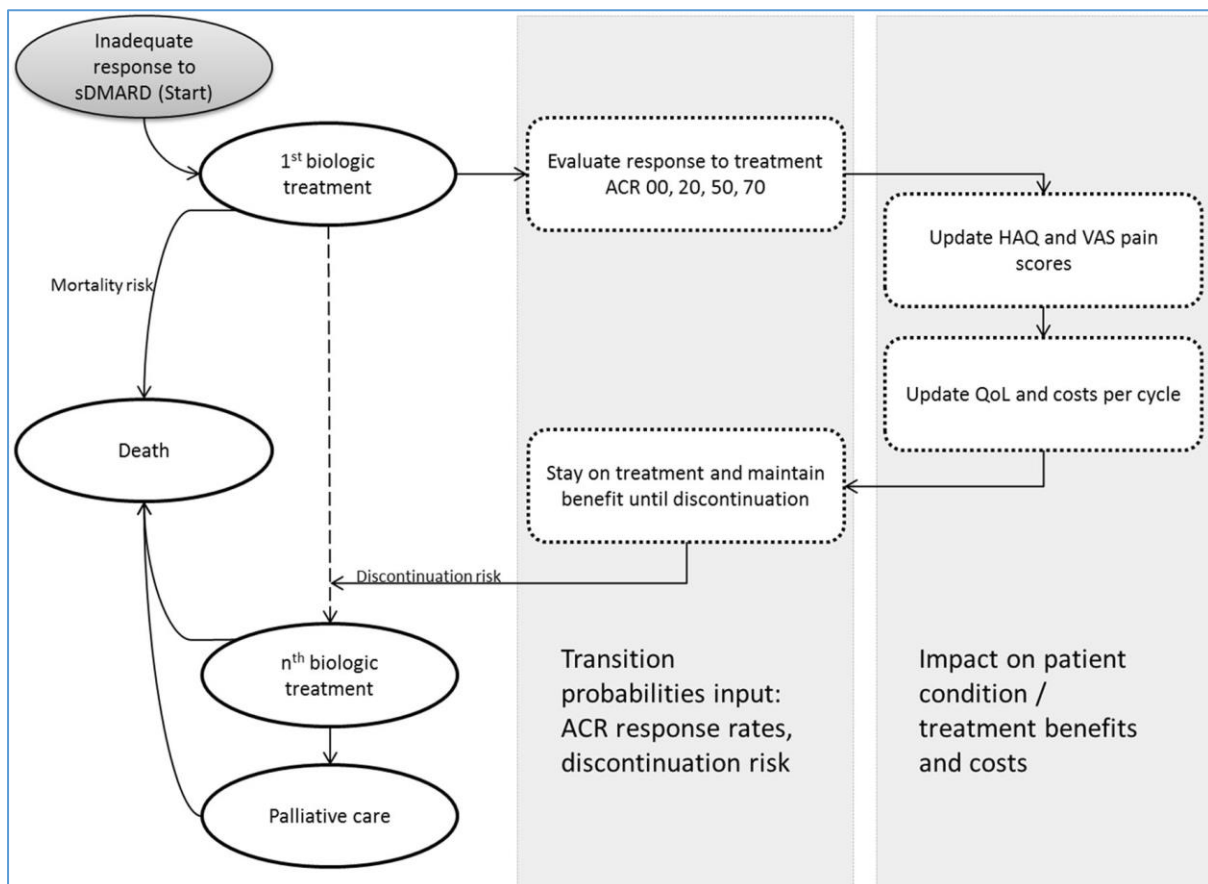
Diamantpoulus model

The Diamantpoulus model was developed to evaluate the cost-effectiveness of adding tocilizumab to the current treatment sequences, and to compare a number of different treatment sequences with and without tocilizumab (Figure 7.7).^{234 235} The patient characteristics were representative of a homogenous group of patients, which were obtained from three RCTs³⁸⁴⁻³⁸⁶ for an early version of the model (2011), and the British Society for Rheumatology Biologics Register (BSRBR)³⁸⁷ for a later version (2014). The 2014 version of the model was used to explore two separate scenarios: patients contraindicated to methotrexate and those who were methotrexate tolerant.

Treatment sequences were represented as a series of health states, which patients progressed through in a fixed order. It was assumed that the patients received all treatments in the sequence, ending with palliative care. Patients could transition to death at any time, based of mortality risk adjusted for rheumatoid arthritis. Initial treatment response was measured in terms of ACR response criteria, with simulated patients allocated to one of four categories: no response, ACR20 response,

ACR50 response, or ACR70 response. Those with no response moved to the next treatment in the sequence, whilst those with a response remained on treatment until withdrawal. Response to treatment was assumed to have an impact on disease severity, as measured by the HAQ score, in the earlier version, and both the HAQ and Visual Analogue Scale (VAS) pain score in the later version. As the patients progressed through the treatments, individual simulation was used to monitor HAQ changes. For the first six months of a new treatment, the model assumed a HAQ score (and VAS pain score) reduction according to four levels of ACR response. Although the proportion of patients achieving each level of response was treatment dependent, the initial HAQ benefit was assumed to be response, not treatment, related, and the HAQ reductions for each level of response were therefore applied universally to all treatments.

Figure 7.7: A schematic diagram of the Diamantopoulos model: a state transition individual patient simulation model for the treatment of rheumatoid arthritis



Taken from: Diamantopoulos A, Finckh A, Huizinga T, Sungher DK, Sawyer L, Neto D, Dejonckheere F. (2014) Tocilizumab in the Treatment of Rheumatoid Arthritis: A Cost-Effectiveness Analysis in the UK. *Pharmacoeconomics*, 32(8), 775–787.

Abbreviations: ACR American College of Rheumatology; HAQ Health Assessment Questionnaire; QoL quality of life; sDMARD synthetic disease-modifying anti-rheumatic drugs; VAS Visual Analogue Scale.

Kobelt model

The kobelt model was based on the combined effect of both function and disease activity to estimate costs and utilities.²⁴⁸ The health states represented disease status. The model included five main

health states based on functional capacity, which were defined using the HAQ score. Each state was further divided into high or low disease activity using the DAS28 scores. In all resulting states, patients could be on a TNF-inhibitor (as first-line, second-line, or half dose), methotrexate, or a generic conventional DMARD. Changes in disease status or treatment were modelled as transitions between the states.

7.3.6.3 Advantages and disadvantages of state transition individual patient simulation model

Treatment sequencing was generally implemented using a fairly simple modelling structure, with the complexity of the decision problem being mainly accounted for by tracking individual patient event histories. Treatment sequencing was implemented as a series of health states, or a decision tree representing the response to each treatment and the subsequent decisions. However, the rheumatology models were based on a fairly similar process, which started with an initial assessment and categorisation of treatment response at three or six months. Non-response then triggered treatment switching, whilst response led to continued treatment, based on data on long-term withdrawal. Patients would progress through the treatment sequence in a fixed order, and were generally assumed to receive all the treatments in the sequence. Individual patients could differ in the way they progressed through the model in terms of their response to treatment, risk of withdrawal, adverse events and mortality. The advantage of individual patient simulation is that this variation can be tracked using patient histories. It can also incorporate a heterogeneous patient population. However, most of the included studies used initial patient characteristics representing a homogenous group of patients, reflecting the decision problem using the average patient. The variation between individuals therefore represented random variation only. None of the included studies considered a differential selection of subsequent treatments according to the reason for quitting the current treatment or evaluated patient subgroups. In fact Diamantopoulos *et al.* modelled patients with and without a contraindication to methotrexate separately.²³⁵ Only one study (Bansback, 2005) modelled the impact of serious adverse events as a separate health state.²¹⁸

The most complex model structure was developed by Finckh *et al.*, in order to compare three different management strategies for patients with very early rheumatoid arthritis, following patients from symptom onset until death.²³⁸ The model included five disease states representing four different levels of treatment response, excellent, good, moderate, and none, and death. The model also accounted for three different potential modes of disease progression (spontaneous remission, slow progression, and rapid progression) that could not be predicted at presentation, and included two outcome measures to represent disease progression, eroded joints and HAQ score. The rate of disease progression was applied to the course of the disease. The model also monitored three other outcomes, quality of life, cost, and death. The time patients spent in a particular disease state determined the probability of transition and the ultimate outcome.

The individual patient simulation modelling approach was generally used to monitor or track changes in disease progression through the treatment sequence, and over a life time horizon. Disease

progression in rheumatoid arthritis was generally measured using HAQ disability score, whilst treatment response was frequently assessed using ARC criteria. The observed improvement in HAQ, on treatment initiation was assumed to be response, and not treatment, related. Whilst the level of response was based on the treatment used. The parameters for most of the Sheffield models were developed using individual patient level data. This meant that regression analyses controlling for other covariates, such as disease duration, could be used for estimating the patient's change in HAQ score.^{218 224 225 274} However, all the rheumatology studies used various assumptions for implementing long term changes in HAQ score.

7.3.6.4 Discrete event simulation

Eight rheumatology studies (Barton, 2004; Chen, 2006; Clark, 2004; Jobanputra, 2002; Lindgren, 2009; Malotki, 2011; Tran-Duy, 2011; Tran-Duy, 2014) reported using a discrete event simulation model, representing four distinct models: the preliminary Birmingham model, the Birmingham arthritis rheumatology model (Figure 7.8), the Tran-Duy, and the Lindgren model.^{219 227 229 245 251 254 272 273} Three non-rheumatology studies (Connock, 2006; Denis, 2008; Heeg, 2008) used a discrete event simulation model,^{233 242} or a similar approach,¹⁹⁹ for evaluating treatment sequences. These include: the Denis model,²³³ the Heeg schizophrenia model,²⁴² and the Birmingham epilepsy model.¹⁹⁹

All eight rheumatology studies were cost utility analysis with a lifetime horizon, four of which included probabilistic sensitivity analysis. One of the non-rheumatology studies used a discrete event simulation to compare the clinical effectiveness of different treatment strategies for glaucoma (Denis model).²³³ The remaining two studies were cost utility analysis; the Heeg schizophrenia model, which was used to evaluate treatment sequences of antipsychotic drugs for schizophrenia,²⁴² and the Birmingham epilepsy model for sequences of antiepileptic drugs.¹⁹⁹ The time horizon of the non-rheumatology models ranged from five^{233 242} to 15 years,¹⁹⁹ and only one study (Heeg, 2008) included probabilistic sensitivity analysis.²⁴²

Only the Heeg schizophrenia model was reported to have been developed using a dedicated discrete event simulation software package.²⁴² The Birmingham Rheumatology model was initially developed using TreeAgree and then constructed using Borla Dephi.^{219 227 229 245 254} The Tran-Duy model was also developed using Delphi language.^{272 273} The software used was not reported for remaining models.

The discrete event simulation modelling approach is illustrated here using mainly the rheumatology studies and one non-rheumatology study. The preliminary Birmingham model provides an example of a discrete event simulation adopting a simple structure for comparing fixed drug sequences. A summary of the evolution of the subsequent Birmingham rheumatology model is also provided, which was based on the random selection of pre-defined treatment sequences. This is followed by a brief summary of the Lindgren model.²⁵¹ This represents an example where data on treatment sequencing were taken from a national registry to develop a background model, which was used to investigate the

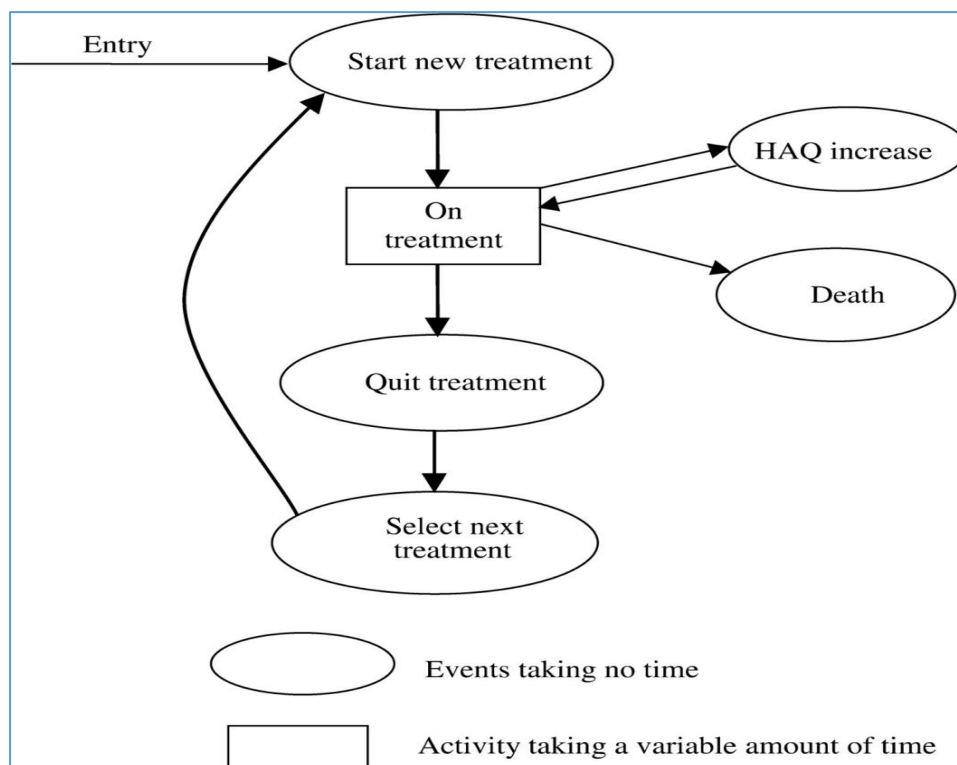
changes that would occur when a new treatment is introduced. The Tran-Duy model provides an example of a discrete event simulation where individual treatments were randomly selected for each line of treatment in order to generate the treatment sequences.^{272 273} Finally, a summary of the Birmingham epilepsy model is provided, as a continuum of the previous narrative relating to the different approaches used for modelling treatment sequences in epilepsy.¹⁹⁹ All models, including the Denis²³³ and Heeg schizophrenia²⁴² models are summarised in Appendix Volume II (Appendix E).

An individual sampling model was chosen for the Birmingham model, in order to allow for a continuous distribution of time to quitting any treatment, which was considered not follow an exponential distribution (discussed in Section 7.3.3.4).²¹⁹ The Lindgren model was programmed as a discrete event simulation because the decision problem related to different treatment courses and sequences, with varying treatment durations and intervals between them, and structuring the question into a chronological 'time-to-event' was intuitive.²⁵¹ A discrete event simulation was also considered a better candidate than a state transition individual patient simulation, in terms of computational efficiency, as fewer calculations are required. A discrete event simulation was chosen for the Tran-Duy model as it allowed specific treatment algorithms to be implemented, reflecting the recommendations made in clinical guidelines and treatment decisions made in clinical practice.^{272 273} It also allowed greater flexibility for modelling treatment selection based on the disease activity and treatment history of the patient, as patients could be in both disease activity and specific treatment health states at the same time. It also meant that various drug sequences could be selected at random, and differential treatment selection could be implemented for specific patient subgroups. An individual sampling model was chosen for the Birmingham epilepsy model because of the need to consider a sequence of drugs, a potentially long time frame, and the need to account for the fact that individuals could receive any one drug for a variable length of time.¹⁹⁹ The model was able to accommodate drugs being prescribed for variable intervals as it was not based on discrete time cycles. The modelling approach could also account for the many different treatment pathways a patient can follow, and include a patient cohort comprising of individuals with a mix of personal characteristics.

Birmingham preliminary model (BPM) and the Birmingham Rheumatology Arthritis Model (BRAM)

The Birmingham rheumatology arthritis model (Figure 7.8), which evolved from the Birmingham preliminary model, was designed to allow the comparison of a wide range of different treatment pathways for rheumatoid arthritis. It is continually being developed and updated and different versions have been used to inform a number of NICE technology appraisals.^{219 227 229 245 254} Most versions were used to investigate the addition of biological agents at various points in the baseline treatment sequence (Table 6.4 and Appendix Volume I, Section D). However, the third version was used to compare biological agents, which were added to a sequence of conventional DMARDs immediately after the failure of a previous TNF inhibitor.²⁵⁴

Figure 7.8: A schematic diagram of the Birmingham rheumatoid arthritis model (BRAM): a discrete event simulation model for the treatment of rheumatoid arthritis



Taken from: Barton, P. (2011). Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. *Rheumatology*, 50(Suppl 4), iv32-iv38.

Abbreviations: HAQ

The model starts from the point of initiating conventional DMARD therapy for early disease in most instances, with the population to which the decision problem applies being generated within the model (discussed further in Appendix Volume I (Section D2)).^{219 227 229 245} The models did not start from the point at which the treatment sequences diverged in order to avoid the requirement of knowledge of distribution of patients' age, gender and HAQ score for the separate starting populations at that point.²²⁹ This method meant that only a single data set for the starting population was required. Patients who did not reach the divergent point were excluded from the analysis.

The initial Birmingham preliminary model (Jobanputra, 2002) was structured so that a patient followed a pathway containing a fixed sequence of conventional DMARDs, both with and without the addition of a TNF-inhibitor.²⁴⁵ Patients switched to the next DMARD in the sequence when the current DMARD was ineffective or produced toxicity. Only one initial patient characteristic was required, which was the patient's remaining lifetime. Two things were sampled for each treatment; first, the time on treatment and, second, the question of whether the treatment was quit for toxicity or lack of effectiveness. The latter was used because the last active treatment included combination therapy of cyclosporin with methotrexate, which could not be given if either treatment component had previously been quit because of toxicity. The model cycles were based on the time spent on a particular DMARD. Tracker variables enabled the patient's clinical course to be influenced by how long they had been on a

particular drug and be related to their past medical history. For individual patients, the total lifetime remaining from entry into the model, accounting for age and gender, and maximum time on a DMARD were sampled from appropriate distributions. For time on DMARD a Weibull curve was fitted to the available data points. The effect of each treatment was assumed to produce a fixed improvement in the patient's health-related quality of life (HRQL), which would be lost on quitting treatment. This assumption allowed QALYs to be modelled relative to the basic curve representing natural history or deterioration over time.³⁸⁸ Deductions were made at the start and end of each treatment period representing the delay in treatment taking effect and an assumed gradual loss of benefit due to loss of effectiveness or toxicity.³⁸⁸

One of the limitations of the preliminary model was that it did not incorporate flexibility in the order in which DMARDs were used, and did not consider the effect of DMARD on disease progression. The assumed fixed pattern of effects of DMARD on quality of life also prevented the model from allowing for mortality effects of DMARD.

In the subsequent Birmingham rheumatology Arthritis Model (Barton, 2004; Clark, 2004) hypothetical patients were assigned to alternative pre-defined treatment sequences using computer-generated random numbers.^{219 229} Although the model was designed so that any desired sequences of DMARDs could be compared, a baseline sequence of conventional DMARDs was used in order to make the decision problem tractable. TNF-inhibitors were then added to this sequence to allow for the different comparisons to be made. The initial model allowed for a total of 16 treatment sequences, which were incorporated by allocating a strategy number to each sequence, with the more expensive and more effective treatments put first. However, all sequences were not simultaneously compared within the model; rather comparisons were made within different sets, for example with and without a TNF added as the third treatment.

One of the main changes made to the model was to define the patient's health state in absolute terms using the HAQ score rather than quality of life. The HAQ improvement on starting treatment was modelled as a fixed decrease in the score, based on the average decrease for all patients receiving a specific treatment. The general decline in a patient's condition over time was also modelled using stepped HAQ increases.

The model structure consisted of events, which took no time, and activities, which took a variable amount of time. The model included six events and only one activity, 'on treatment'. The main loop, 'start new treatment' – 'on treatment' – 'quit DMARD' – 'select next treatment' (Figure 7.8), was followed for each DMARD successively until no DMARDs remained and the patient then moved to palliation. The events 'HAQ increase' and 'joint replacement' interrupt the normal flow through the model whilst on treatment. Time was advanced in the model during the activity 'on treatment', which could be terminated by any of the following four competing events: death, HAQ increase, need joint replacement, or quitting DMARD. HAQ increase was not modelled as a separate event, but included

at the end of 'on treatment'. Each virtual patient was assigned the attributes, age, gender, and starting HAQ score from the appropriate distributions. The number of DMARDs left and the identity of the first DMARD were determined according to the treatment strategy being applied. The patient then moved to 'start new treatment'. Time on the DMARD was sampled from the Weibull distribution with parameters appropriate to the particular DMARD. This was added to the patient's current age to give the age at which the DMARD was quit. Using age to quit avoided the need to resample after HAQ changes or joint replacement. Time to quitting treatment was found by subtracting the patients' current age from age at which treatment was quit. Time to death, HAQ increase, and joint replacement were sampled from appropriate distributions. Each of the times sampled was calculated based on the assumption that no other event occurred first. The lowest represented the event that occurred next. If the next event was 'change in HAQ score', then the HAQ score was increased by 0.125 and the patient restored to the state 'on treatment'. Otherwise they moved to one of the remaining three events, as appropriate.

The second version of the Birmingham rheumatology Arthritis Model (Chen, 2006) included two important improvements.²²⁷ Firstly, rather than using a fixed average HAQ change for all patients, the HAQ improvement on starting treatment was allowed to vary between individuals. The HAQ on starting a new DMARD was sampled on an individual basis and took the form of a multiplier, which was applied to the patient's HAQ score on starting treatment. This allowed the model to reflect the fact that patients with a higher score on starting treatment have a greater scope for improvement. Secondly, time on treatment included explicit consideration of early quitting, with early quitting owing to lack of effectiveness being correlated with poor HAQ improvement on starting treatment. The model allowed for two stages of early quitting. The first was set at six weeks, with withdrawal assumed to be due to toxicity, and the second between 6 and 24 weeks, when withdrawal could be due to toxicity or inefficacy.

The second version of the model was also used to inform additional work commissioned by NICE looking at the sequential use of TNF-inhibitors.³⁸⁹ The analysis included changes to the input parameters to that of the initial submission by Chen *et al.*²²⁷ In the initial submission²²⁷ the effect of the second TNF-inhibitor was assumed to be the same as the first, whilst the additional work³⁸⁹ accounted for the differential effect of TNF-inhibitors when used as -first and second-line treatment. The data sources used to inform the second TNF-inhibitor are discussed in Chapter 6, Section 6.6. In modelling terms, the effect of a second TNF-inhibitor being less effective than the first was handled by allowing separate parameters for any treatment used as the first or second TNF-inhibitor.³⁸⁸ This additional work also used rituximab as the comparator for the second TNF-inhibitor, which is administered as fixed treatment courses at intervals of at least six months. This meant that time on treatment could not be modelled as following a survival curve in continuous time, and additional coding was required.³⁸⁸ It was proposed that future versions of the model should include each new prescription as a single event in the model.

The third version of the model (Malottki, 2011) included the use of probabilistic sensitivity analyses.²⁵⁴ The model also allowed different choices of treatment options depending on toxicity of previous treatments.

Software comparison for implementing a discrete event simulation

Two adaptations of the first Birmingham Rheumatoid Arthritis Model were developed using different software, one in TreeAgree DATA Pro and the other in Borland Delphi.²¹⁹ Barton *et al.* noted that:²¹⁹

“TreeAgree DATA has the advantage that the logic of the model is open to inspection by the user, whereas the Borland Delphi version runs quicker and can thus be used for extensive sensitivity analysis. In the TreeAgree DATA version of the model, the events and activities are coded as states following a Markov node Tracker variables are used to record all relevant information, including number of DMARDs remaining, total cost and QALYs to date, and time taken. The implicit time-keeping routines within DATA, which assume a constant (and unspecified) time interval between each cycle of the model, are completely bypassed. The structure of this model is substantially different from that of a Markov model; the use of a Markov node is simply the means provided by the software which allows the model to be built. In the Borland Delphi version, procedures are used for each event and activity. These are linked through further procedures which ensure that the ‘working’ procedures are called in the correct order.”

Lindgren model

The Lindgren model was developed to evaluate the introduction of rituximab, as a second-line treatment after the failure of the first TNF-inhibitor for rheumatoid arthritis.²⁵¹ The authors had access to individual patient level data on the effectiveness of up to three lines TNF-inhibitors from a National registry. This included data on functional capacity (HAQ), disease activity (DAS28), and utility (EQ-5D). Data for rituximab was based on aggregate data taken from the treatment arm of a placebo-controlled trial.

The model structure included three events: ‘start treatment’, ‘stop treatment’, and ‘die’. Patients could therefore be in one of three states: on treatment, off treatment, or dead. For on treatment, a difference was made between the first, second, and third TNF-inhibitor, but not between agents *per se*. The treatment state was further divided into high or low disease activity (using a DAS28 score of 3.2 as the cut-off point). In between treatments all patients were assumed to have a high disease activity. A change in state for each individual was triggered by treatment discontinuation, treatment re-initiation, change in disease activity or death. While in a given state, the characteristics for individual patients relating to gender, age, disease duration and function drove the time to the next event. A cox proportional hazard model was used to identify covariates with possible impact on times to event data for TNF-inhibitors, which included gender, age, disease duration, current HAQ, current disease activity and treatment line.

Patients entering the model were either starting their second TNF-inhibitor or rituximab, and stayed on these treatments until discontinuation, according to the registry data for TNF and an RCT for rituximab. Patients on a TNF-inhibitor would then re-initiate treatment with their third TNF-inhibitor according to the timings in the registry, whilst patients on rituximab would start immediately on their second TNF-inhibitor. Not all patients re-initiated treatment as treatment intervals could be longer than the simulated time, representing the data in the registry. Where patients failed again they switched to another TNF-inhibitor, which was assumed to be the same as the third-line. When patients withdrew from treatment it was assumed that the treatment effect would be lost immediately and patients would return to their HAQ score at baseline.

Tran-Duy Model

A discrete event modelling framework was chosen by Tran-Duy *et al.* (2011) for simulating long-term outcomes of sequential treatment strategies for ankylosing spondylitis.²⁷² The same modelling approach was subsequently used (Tran-Duy, 2014) for evaluating treatment sequences for rheumatoid arthritis.²⁷³

The models were used for comparing two alternative strategies, with and without the use of biological agents. Patients entering the model were newly diagnosed, with the decision population developed as part of the modelling process. For rheumatoid arthritis, the treatment strategy without biological agents included eight conventional DMARDs, where the first two were pre-specified and the remaining five drugs were selected at random. The alternative strategy included the same eight conventional DMARDs plus four biological agents, which included two consecutive TNF-inhibitors followed by two biological agents using a different mode of action. TNF-inhibitors were initiated after the failure of the first two conventional DMARDs and randomly selected from a list of five available drugs. Non-TNF-inhibitors were randomly chosen from a list of three available drugs, with the selection taking into account the patients rheumatic factor status.

The model included seven patient attributes, four DAS28-related states, eight treatment-related states, three DAS28-related events, five DAS28-neutral events, and eight procedures. These are summarised in the Appendix Volume II, (Appendix E). Disease activity was characterised by changes in DAS28. The authors noted that if no toxicity occurred, the DAS28 status of a patient could be conceptualised as undergoing three phases whilst on active treatment: a decreasing phase characterised by a steady decrease in DAS28, a maintenance phase characterised by small fluctuations in DAS28, and increasing phase where there is a continuous increase in DAS28 after a patient stops responding to treatment. Within this third phase the patient's DAS28 score was assumed to return to the baseline value, known as rebound. The treatment-related states included: on methotrexate, on second conventional DMARD, on first TNF-inhibitor, on second TNF-inhibitor, on first non-TNF-inhibitor, on second TNF-inhibitor, and on 'palliative' treatment. The DAS28-related states were used to determine the events that may occur given a trend of change in DAS28. Whilst the treatment states were used to determine changes in DAS28, times to DAS8-related events, and a

new treatment when the current drug failed. Three events (severe toxicity, select new treatment, start new treatment) only arose when a visit to a rheumatologist occurred. The remaining seven events were competing events. For competing events the patient 'jumped' to the event to which the sampled time was shortest. When the event occurred an associated procedure was invoked for implementation, where the patient characteristics were updated and times to next events computed.

The treatment algorithm in the model was based on clinical guidelines and rheumatologist opinion. Within this algorithm, methotrexate was started as soon as the patient was diagnosed. The next treatment was considered when a drug failed primarily or secondarily, or caused severe toxicity. Primary failure was assumed if DAS28 was still higher than 3.2 after 3-6 months since the start of the treatment. Secondary failure was assumed when DAS28 went back to a level less than 3.3 after primary response to the treatment. Biological agents were combined with methotrexate if patients did not experience severe toxicity when receiving methotrexate monotherapy after diagnosis. Differential treatment selection was also implemented according to whether the patient was Rhesus factor positive or negative.

Treatment effects on disease activity, toxicity, and duration of effect and timing of toxicities were based on individual patient-level data obtained from observational studies. These were based on an inception cohort for conventional DMARDs and a national patient registry for biological agents. A number of treatment-related assumptions were made due to insufficient data. It was assumed that the effectiveness of a specific drug was independent of the identity and the cause of failure of the drugs that had been given previously. The absolute changes in DAS28 were sampled for each drug, or drug class, distinguishing the first and second biologic, using a statistical linear model with DAS28 at the start of the treatment as an explanatory variable. Estimates for the intercept and slope were obtained from patient registry data. Long term progression of physical function, which was measured by HAQ, was simulated based on the longitudinal relationship between DAS28 and HAQ. The study accounted for the fact that QALYs and costs are non-linearly related to DAS28 and HAQ.

Birmingham epilepsy model

The Birmingham epilepsy model presented by Connock *et al.* was used for comparing fixed drug sequences of up to four treatment lines for patients with newly diagnosed partial epilepsy.¹⁹⁹ The model represented the use of antiepileptic drugs during childhood, and patients therefore exited the model on becoming 18 years of age. The baseline treatment sequence included monotherapy for the first- and second-line treatment, followed by either a monotherapy or adjuvant therapy for third- and fourth-line, depending on treatment outcome. For each patient in the model a tracker variable monitored whether or not the drugs used as monotherapies earlier in the sequence were partially effective and, if so, the drug would be available for use as combination therapy.

The included patients varied in terms of the age at diagnosis, gender, and whether or not they experience learning difficulties. As patients progressed through the model their treatment pathways

and experience of epilepsy were monitored. Treatment sequences were compared in terms of the average time spent in each of four main treatment outcome states, with the assumption that longer durations in states with reasonable efficacy and side-effect profiles represent a positive outcome. The four treatment outcome states included: i) intolerable adverse effects, ii) lack of effect on seizure rate, iii) partial efficacy with tolerable or no adverse effects, and iv) complete seizure freedom with tolerable or no adverse effects. The last two states, 'partial efficacy' and 'seizure freedom', also included secondary states relating to successful or unsuccessful drug withdrawal, depending on whether the patient had seizure freedom, or not seizure free but preferred to remain untreated, respectively. The first two treatment outcomes lead to early discontinuation of treatment. Patients who experienced these outcomes progressed to the next choice monotherapy or opted to discontinue drug treatment. Those who entered the third outcome could stay on the current drug, try next choice monotherapy, try next choice add-on therapy, or discontinue treatment. It was assumed that the patients' willingness to try an alternative treatment depended on the number tried at this point and that, as the number of drugs tried increased, the patient was more likely to try add-on therapy and less likely to try further monotherapy. Patients who achieved outcome four were assumed to withdraw from the drug after a given period, which was sampled in the model, or remain on current drug if reluctant to withdraw. It was assumed that the proportion discontinuing due to late toxicity or reduction in efficacy over time was negligible.

The Birmingham model represents the most sophisticated model of treatment sequences for epilepsy. However, it was still restricted by the available evidence. Transition probabilities, or the likelihood of a patient reaching a particular outcome for each drug were based on a systematic review of RCTs, and the use of a reduction factor and assumptions to account for treatment sequencing effects. This is discussed in more detail in Chapters 5 (Section 5.6) and 6 (Section 6.5.3). Data for other clinical parameters, such as proportions discontinuing treatment, time to discontinuation or withdrawal, and likelihood of moving on to add-on therapy were based on epidemiological studies and clinical advice. Data on the proportion of patients achieving secondary outcomes were not treatment-specific, but were specific to each line of treatment. It was approximated that 10% of patients went on to have successful surgery, and did not have any further drug treatment after first-line treatment. One of the main challenges of this approach was the limited data on time to event from the available evidence. Where time to treatment withdrawal was reported, it could not be disaggregated according whether treatment was discontinued due to adverse effects or lack of effect.

7.3.6.5 Advantages and disadvantages of discrete event simulation models

Discrete event simulation appears to provide a flexible approach for modelling treatment sequences and can handle complex sequencing decision problems intuitively. It was successfully used for implementing treatment selection and cessation based on algorithms reflecting clinical guidelines and practice. This required the implementation of stopping rules and specific sequence of treatments to be followed based on the different reasons for quitting treatment. It was able to account for different treatment selection for patient subgroups, such as patient with rhesus positive or negative rheumatoid arthritis. It was also able to accommodate the unpredictable nature of disease progression, multiple

treatment outcomes, and the fact that not all patients go on to receive subsequent treatments in the sequence.

Treatment sequencing was implemented within the model based on either the random selection of pre-defined sequences, or developing the sequences by selecting individual drugs using a random process at specific points in the sequence. Discrete event simulation was also used for implementing fixed treatment sequences, whilst allowing for treatment duration to be modelled as a continuous distribution.

The main advantage of discrete event simulation over state transition models for modelling treatment sequences is the way it handles time, as it is not restricted to either the use of equal time periods, or the Markovian assumption. In fact, the discrete event simulation approach was frequently chosen in order to implement the variable time to quitting treatment, which may not be constant and will differ between treatments. Another advantage is the ability to accommodate competing risks. Patients moving through the discrete event simulation can experience events at any time period after the previous event. The analysis of the model is triggered by the occurrence of an event, at which point the model asks what and when is the next event for this patient, rather than a Markov process, which asks what events are occurring at regular intervals.³⁶⁶

The main disadvantage is the extensive data required for model parameterisation, including time to event data. The flexibility of the discrete event simulation allows the modelling of complex treatment and disease process, but treatment independence was often assumed due to insufficient data, or simplifying assumptions used for treatment sequencing effects. Patient level data are preferred for implementing discrete event simulation, but they can also be based on aggregate data. Only two included studies used individual patient level data. Tran-Duy *et al.* used data from an inception cohort (Nijmegen Inception Cohort) and a patient registry (Dutch Rheumatology Arthritis Monitoring, DREAM).²⁷³ However, despite this the effectiveness of a specific drug was assumed to be independent of the identity and the cause of failure of the drugs that had been given previously. For Lindgren *et al.*²⁵¹ individual patient data were only available for usual care from a patient-based registry (Southern Sweden Antirheumatic Therapy Group, SSATG), whilst the new treatment was based on aggregate data from a published RCT (REFLEX).³⁹⁰ Another important limitation is the added skills required for implementation. Specialist software may be preferable, in terms of ease of implementation and time commitment, to trying to adapt commonly used decision analytic modelling software. However, none of the studies reported problems with computing and implementation, and although earlier versions of the Birmingham Rheumatoid arthritis model did not include probabilistic sensitivity analysis, many of the included studies using discrete event simulation did (Table 7.1).

The relationship between the patient characteristics and the outcomes, costs and health utilities are non-linear. Using linear regression with constant error terms may lead to biased expected outcomes of a population based on individual patient simulation. Tran-Duy *et al.* reported using a different

approach from linear models used in existing discrete event simulation models,²⁷² where long-term progression of physical function in rheumatoid arthritis was quantified using a linear mixed-effects model.²⁷³

7.3.7. Population based models or open cohort models

Open models allow new cohorts to enter over time and are sometimes referred to as population models.³⁶⁹ Two included studies (Erhardt, 2012; McEwan, 2010) were described as population models, which were based on an individual patient simulation.^{237 255} Both studies used the same model, known as the Cardiff Diabetes model, and are therefore described together here. However, one of the studies (Erhardt, 2012) was in fact a closed model.²³⁷ One further included study (Launois, 2008) used an open cohort model. The Launois model, which was described as a dynamic Markov cohort model, was used in order to consider the budget impact of treatment sequencing.²⁴⁹

7.3.7.1 Non-terminating and terminating population based models

Cardiff Diabetes Model

The Cardiff type 2 diabetes mellitus model (T2DM) was developed for evaluating a wide range of health economic issues. It was used in one study (McEwan, 2010) for evaluating treatment sequences representing treatment escalation, and in the other study (Erhardt, 2012) for comparing two second line-treatments within a fixed sequence of three treatments. For the evaluation of treatment escalation a previous version of the Cardiff model was adapted to operate as a non-terminating simulation (McEwan, 2010), whilst for the comparison of fixed sequences it was implemented as a terminating simulation (Erhardt, 2012). A terminating simulation is one in which the model either has a natural run length, for example, the model terminates when the last simulated individual dies, or is run for a fixed amount of time, for example, 40 years. The terminating model was described as a 'discrete event simulation', and as a 'fixed-time-increment stochastic simulation'.²⁵⁵

The Cardiff model uses the UK Prospective Diabetes Study (UKPDS) Outcomes Model equations to predict macrovascular and microvascular complications in subjects with type 2 diabetes mellitus. The model is continually being updated in order to incorporate the most recent risk equations and input parameters from UKPDS, including UKPDS 82 outcome study equations in 2014. It can be implemented using mean values, probabilistic inputs, or user-supplied patient level data. Standard model outputs include time-dependent event rates, and total cost and utility decrements associated with all predicted events.

Treatment pathways for type 2 diabetes tend to advocate a failure-driven algorithm for lowering blood-glucose, which leads to the sequential addition of treatments, or treatment escalation. The addition and combination of multiple blood-glucose lowering agents may be associated with significant side effects resulting in detrimental quality of life.²⁵⁵ The Cardiff model was therefore used in treatment escalation study (McEwan, 2010) to examine the effects of treatment strategies on health related quality of life (HRQoL) improvements associated with different hypoglycaemia profiles, rather than the

efficacy variables, such as change in blood glucose levels (HbA1c). However, the latter outcome was used as an indicator for treatment switching. In the second study (Erhardt, 2012) the Cardiff model was adapted to accommodate treatment effects, as well as complications resulting from disease progression. As well as producing the standard model outputs developed by the core equations, the model was also used to estimate changes in HbA1c and body weight, and occurrence of hypoglycaemia.²³⁷ Treatment switching to second- and third-line was either time dependent or controlled via pre-specified glycated haemoglobin (HbA1c) thresholds.

The non-terminating simulation model (McEwan, 2010) was initialised using a prevalent patient population profile and utilised the annual type 2 diabetes incident rate to allow new cases to enter the model each year. Patients exited the model through diabetes-specific or all-cause mortality. The baseline cohort characteristics were drawn from the UK Prospective Diabetes Study (UKPDS) 68 outcome study, and time-dependent evaluation of risk factor profiles were implemented using equations reported in the same study. Where appropriate, each simulated cohort had a treatment effect profile applied, which was consistent with that reported for each treatment. Treatment profiles were taken from a systematic review of first-line metformin, and selected RCTs for second and third-line. These treatment profiles were fully applied in the first cycle of the model. In all subsequent cycles, the risk factor trajectories were updated according to the natural history progression specified by the UKPDS 68 panel equations for HbA1c, systolic blood pressure (SBP), and total cholesterol.

Patients started on first-line treatment as they entered the simulation. Following the application of a treatment effect modification to each patients' baseline HbA1c, the model used dynamic equations to project HbA1c over time. Pre-specified HbA1c threshold values were used to invoke an escalation in treatment to second- or third-line. To control for the rate at which simulated subjects progressed through the therapy escalations, the slope coefficient that controlled the change in HbA1c over time was recalibrated for each treatment-line. This ensured that the model predicted a constant proportion of subjects on first-, second-, and third-line treatment to that seen in the UK when applying the specific thresholds to second- and third-line.

The core Diabetes model was coded in C++ and linked to a Microsoft Excel front end. Non-terminating simulations require a 'run in' period to achieve a steady state prior to collecting summary statistics.²⁵⁵ The model was run over 100 years and data collected over last 10-year period. The terminating model was run over 40 years using yearly cycles in which treatment-dependent risk factor profiles, including glycated haemoglobin and body weight, were modelled dynamically.²³⁷

7.3.7.2 Dynamic Markov multi-cohort model

Launois model

The Launois model was used to estimate the budget impact implied by the introduction of rituximab after failure of one or more TNF-inhibitors for rheumatoid arthritis.²⁴⁹ A Markov model was developed to reproduce the course of patients treated by either infliximab, etanercept, adalimumab or rituximab

after the failure of one or more TNF-inhibitors. Sensitivity analysis was conducted to account for patients on third and subsequent lines of treatment who were expected to consume more healthcare resources. The model period extended over 4 years. The model was run by having the current patient cohort progress through the model, accompanied at each six-month cycle by a cohort of newly diagnosed patients. The model comprised of four states corresponding to the four biologics, and one additional entry/exit state, where patients switch to a new biologic or die. At the end of each cycle a patient could either continue with the original treatment if ACR20 criterion was fulfilled, or switch to a new biologic treatment, or exit the model (switch to a non-biologic strategy or die). New patients on second line treatment were generated from the 'switch' state to exactly compensate for patients exiting the model. In other words, it was assumed that the target population was constant over time. The main weakness of the dynamic cohort approach is that a patient who failed a biologic treatment could receive exactly the same treatment after a switch. However, the authors were not interested in individual patient patterns, but in the global budget impact of rituximab for the whole target population.

The model was used to test three different hypotheses for rituximab penetration into the market.

These represented the following situations:

- i. TNF-inhibitors are the only available treatments. This was based on the market shares of: infliximab 16%, etanercept 38% and adalimumab 46%
- ii. Rituximab penetrates the market progressively. This was based on previous market shares for the first cycle. Then, each patient who failed a TNF-inhibitor switched to rituximab and each new patient in second-line biologic treatment began with rituximab.
- iii. Rituximab is the only available treatment.

The overall findings showed that rituximab is expected to produce important savings when used after the failure of one or more TNF-inhibitors, but this is mainly due to its lower drug acquisition cost.

7.3.7.3 Advantages and disadvantages of open or population based models

The included studies did not show any clear advantage of using an open model for modelling the clinical effectiveness of treatment sequences, compared to other included modelling approaches. The included studies that used an open approach were also dynamic models. As with previously reported discrete event simulation models, the Cardiff model was able to account for a dynamic disease process whilst incorporating treatment sequencing. The Cardiff model also provides another example of a model that utilised the natural history and time to event data from an observational study, and treatment effect profiles based on clinical trials. The model is capable of running using various levels or types of data, and can potentially be built upon in the future.

7.4 DISCUSSION

7.4.1 Summary of the findings

An initial evaluation of the key features of different modelling approaches was conducted based on published taxonomies.^{344 391} This was then used for categorising the different modelling approaches

used for evaluating treatment sequences. The included modelling approaches were discussed under three broad headers, depending on whether they simulated a closed cohort or a dynamic population and, for closed models, whether they simulated all individuals simultaneously or one at a time. The included models were also grouped according to whether they were state transition models, such as Markov models, or conceptualised around events, such as decision trees and discrete event simulation. Individual-level simulation models were grouped according to whether they allowed transitions to occur only at specified time intervals or not. Those that did were considered as state-transition models, and those that did not were categorised as discrete event simulation.

The different modelling approaches were assessed in terms of their advantages and disadvantages for modelling treatment sequences. The rule of thumb generally used for choosing a modelling approach is to select the simplest approach required, which will be dependent on the complexity of the decision problem and the extent of the treatment sequences being investigated.^{370 392} One important question is whether a simple modelling approach can accommodate the added complexity sometimes required for modelling treatment sequences. This not only includes the need to consider different algorithms that dictate the choice of subsequent treatments, but also the need to simultaneously capture other important elements relating to the treatment and natural history of the disease. Included models were therefore evaluated in terms of what types of additional attributes or complexity in the decision problem they also accounted for.

Seventy modelling studies were reviewed and fifty distinct models identified. A wide range of modelling approaches were used for investigating treatment sequences, which included deterministic decision tree, stochastic decision tree, Markov cohort model, partitioned survival cohort model, semi-Markov cohort model, individual-patient simulation (IPS) state transition models, discrete event simulation, non-terminating population-based simulation, terminating population-based simulation, and dynamic Markov cohort model. No study compared different approaches for evaluating treatment sequences. Thirty-six included studies (51%) were rheumatology and ten (15%) were oncology. Most non-rheumatology studies used a Markov cohort modelling approach, whilst most of the rheumatology studies used individual patient simulation. The extent of the sequences being modelled varied quite considerably. The number of treatment lines considered by the rheumatology studies ranged from two to 12, whilst the number of treatment lines modelled by the oncology studies ranged from two to five, and for the remaining studies ranged from two to six.

A summary table that provides a quick reference and allows for the comparison of all the different modelling approaches used, the individual attributes they accounted for, and their advantages and disadvantages in modelling treatment sequences was developed (Table 7.3).

7.4.2 Advantages and disadvantages of different modelling approaches

7.4.2.1 Cohort models

Decision trees

Decision trees are the most transparent and simplest modelling approach to implement. They enable the decision problem to be structured in a meaningful and visual manner. Treatment sequences can be compared simultaneously within the same model, or implemented as a separate model for each sequence. Their potential limitations for modelling treatment sequencing is that they have a finite time horizon, only allow one way progression, and cannot handle recursive events very easily, without potentially becoming bushy and unwieldy.^{391 393} However, they were in fact successfully used for modelling treatment sequences that represented the option of treating a relapse with the same treatment that was initially found to be successful.²³⁹ They were also used to model the overall time spent in a health state by incorporating assumptions regarding the timing of events or chance nodes.^{236 257} An alternative approach used to account for the timing of long term outcomes was to implement a decision tree model alongside a portioned survival analysis.²⁰⁷ The decision tree was used to account for the fact that not all patients went on to receive a second active treatment. The included decision tree models were also able to allow further complexities in the decision problem, such as different levels of response that lead to different durations of response, different reasons for discontinuation that impact on the selection of subsequent treatments, and different reasons for mortality that will have different probabilities and timing, for example toxic death and all-cause mortality. However, despite this, decision trees are likely to be inefficient at modelling treatment sequencing for a chronic disease, which generally includes a series of repeatable events and evolves over time. The fact that they are governed by fixed timing of the treatment outcomes means that, for example, events such as recurrence or intolerance to treatment are based on the simplifying assumption that these would occur at a fixed time in the future.

Markov cohort models

A Markov cohort model allows events to occur at any time, which is why it is often used for modelling treatments for a chronic disease. Within a Markov model, the decision problem is conceptualised as a series of mutually exclusive health states, with transitions between states based on events. It has the advantage of being easily implemented within commonly used spreadsheet or decision analytic software, and is well suited for modelling time to event and repeated events. However, the occurrences of these events are analysed at fixed intervals, or model cycles, and, in the commonly used Markov chain, are assumed to be constant over time. Other potential limitations for modelling treatment sequencing relate primarily to the Markovian property. The Markov model requires state transition probabilities to be independent of previous states of the model, and asserts that patients can only be in one state at any given time.³⁹⁴ This means that a Markov model cannot account for patient's history, time spent in a specific states, or the occurrence of multiple events within a single cycle, all of which are relevant for modelling treatment sequencing. However, the Markov cohort model is commonly used in health economics in general and a number of solutions have been

developed over the years to overcome these limitations, which could also be useful for modelling treatment sequences.

Most Markov models were used for comparing pre-defined sequences, based on modelling a series of individual treatments or treatment lines. However, one distinct Markov cohort model was developed for identifying the optimal treatment strategy out of all conceivable sequences.²⁷⁸ A very simple two state Markov model was run for each individual treatment and the optimal treatment based on ordering the treatments according to their net-benefit or cost-effectiveness per unit of time. The advantage of this approach is that it can incorporate all conceivable sequences, does not require the treatment sequences to be defined in advance, and can also be used for individualised treatment decision making. However, this approach was based on the strong assumptions that the efficacy of individual treatments are independent of treatment history, and that the treatment does not impact disease progression, and is therefore only suitable for certain health conditions and treatment types.

The Markov framework was used for comparing predefined sequences by running a separate model for each sequence, generally using the same structure and an identical cohort of patients. Three different approaches were used to conceptualise treatment sequencing within the Markov cohort model. This included:

- i. Treatment sequences implemented as a series of treatment-specific Markov states that patients progressed through in a forward motion
- ii. Markov states used to represent both treatment-line and other relevant factors or attributes
- iii. Markov cycle tree used to implement a treatment switching algorithm or event, with the Markov states being used to represent the patients' transition through different levels of disease activity, or to model the natural history of the condition.

Most of the rheumatology studies implemented treatment sequences using Markov cycle trees, whilst all the cancer studies used the first two approaches.

The first approach represents the simplest structure for modelling treatment sequences. It can easily allow for treatment skipping and can be expanded to account for different levels of response (to each treatment), which are likely to have different progression rates. This approach was frequently based on line of therapy. A potential limitation of modelling lines of therapy, rather than specific treatments, is that it ignores potential cross resistance between classes, which may be an issues for some targeted therapies.

The first approach does not account for additional factors, for example, different reasons for quitting treatment (adverse effect, lack of response, or loss of efficacy), which are likely to influence the choice of subsequent treatment. For this reason some studies used the second approach to represent other attributes. However, the challenge here is that the Markovian property only allows patients to be in one state at a time.³⁹⁴ Treatment sequencing was generally based on a simple structure, with Markov states representing response and non-response to each line of treatment, and with treatment

response being modelled as a recursive state.²⁶⁹ Additional states were usually implemented as temporary or tunnel states, which allowed the model to overcome the Markovian assumption of no memory and make certain transitions dependent on particular attributes such as relapse or disease progression.^{216 263} These were also used for modelling treatments administered for a fixed period, where patients who respond to treatment, or enter a state of clinical remission, could withdraw from treatment.²⁵⁰ Temporary states such as relapse were then used to trigger the next treatment, or even the re-use of a prior treatment. Additional states were also used to monitor adverse effects. One cancer study incorporated toxicity as a separate health state, allowing toxic death and all-cause mortality to be modelled separately.²⁷⁷ This also enabled the model to account for the fact that patients who develop toxicity can continue on the same treatment at a lower dose. However, toxicity and progression are not mutually exclusive events, and therefore a short cycle length of one week was used to overcome the Markovian property prohibiting multiple events to occur within a cycle. This solution, however, may not be feasible for most complex sequencing decision problems. An alternative approach used was to include additional health states and stratification, for example, modelling response to each line of treatment with and without adverse effects as a series of mutually exclusive states.

The second approach can potentially lead to state explosion, and a complicated model. Some studies therefore used the third approach, where the occurrence of multiple events such as treatment response and adverse effects were implemented using a Markov cycle tree. The Markov cycle tree consists of a series of chance nodes representing the events that can occur within a model cycle.³⁷² The terminal branches represent the resulting health states, and the distribution of the cohort among these states at the end of the cycle. This approach was used in models where there was a need to account for different treatment response, fluctuating levels of disease activity, and fairly complex treatment pathways.^{253 265 266} It also enabled the model to be based on the natural history of the disease condition.²⁶⁸ One model (Maetzel model) used the Markov cycle tree to account for what happens to patients who continue on treatment who may not achieve a clinical response but still choose to continue the same drug, or may experience mild adverse effects, which have cost implications.²⁵³ It also enabled the model to allow patients who experienced intolerance to skip some of the subsequent treatments.

The third approach, of using cycle trees, was used for implementing both simple and complex models. One model (Shepherd model) provides an example of a complex model that also used tunnel states to track treatment history.²⁶⁶ Tunnel states were used to account for potential cross-resistance from previous treatment and to monitor two patient subgroups with a different disease variant. The model was implemented as a Markov cycle tree, which included a subtree with terminal states representing multiple tunnel variables. However, the treatment sequences being investigated only included two lines of active treatment. Modelling more extensive and complex treatment sequences would potentially lead to state explosion or require the use of simplifying assumptions that may not hold. Another model (Albert model) provides an example of a simple model using the cycle tree, which also

used tunnel states to allow transition probabilities to vary with time and for each treatment.²¹⁶ The model used a cycle tree to implement treatment sequencing within a simple three state model representing the treatment outcomes improvement, active disease, and toxicity. The improved state was modelled as a separate tunnel state for each treatment, which allowed the model to reflect the fact that probabilities for continuing treatment and developing toxicity varied with time, and for each drug. Another approach used was to make the duration of treatment dependent on the level of response, rather than treatment, whilst the level of response was treatment-specific. This allowed generic estimates of duration of treatment to be used. An example of this was provided by one model (Heeg cancer model), which assumed that the probability of treatment switching was dependant on the level of response and line of treatment, but independent of the actual treatment used, whilst treatment response was dependent on both the actual treatment used and the line of therapy.²⁴³ This assumes a homogenous duration of response across different treatments, but not line of therapy.

Partitioned survival

The partitioned survival model, also referred to as area under the curve, is another example of a cohort model with a finite number of exhaustive and mutually exclusive health states. However, in this type of model, the number of patients in each health state at any time is determined directly from the underlying survival curves,³⁹⁵ and based on continuous time. This approach was used in studies of advanced cancers for modelling overall survival for specific sequences of two to three-lines of treatment, with data on survival obtained from prospective sequencing trials. One theoretical study, presented as an abstract only, proposed using partitioned survival analysis based on the progression free survival for each treatment line for modelling more extensive treatment sequences for advanced cancer.¹⁹⁵ A partitioned survival approach was also used by one included rheumatology study (Schadlich model) to model fixed sequences of five or six treatments.²⁶⁴ Treatment sequences were implemented as a series of health states that patients could migrate through over time. However, the base case analysis was based on the assumption of treatment independence, the impact of which was explored in sensitivity analysis (Chapter 5, Section 5.6.2 and Chapter 6, Section 6.5.3.1). The model was also unable to account for disease duration, and did not consider the impact of adverse effects, or other additional attributes.

Semi-Markov cohort model

An alternative approach to incorporate time dependency in specific Markov states is to use a semi-Markov process model. An example of this was provided by a model (York epilepsy model) of antiepileptic drug sequences of three treatment lines.²⁷⁶ In epilepsy the probability of treatment failure decreases with increased time spent on an antiepileptic drug (Appendix Volume I, Section C5). The progress of the cohort through the model was tracked externally using a three-dimensional transition probability matrix. Time spent in the current state was implemented as a two-dimensional matrix, where one-dimension represented the current treatment state and the other represented the time, or number of cycles, spent in the current state. This enabled the proportion of patients in each state for each duration to be recorded. The transition probability matrix included a third dimension, represented

by the potential future state. This meant that transitions to the future state were conditional on the current state and time spent in the current state.³⁸⁰ A fairly simple structure was used to model fixed treatment sequences, which were implemented as a series of health states. Each active treatment line was characterised by two health states. An initial temporary state representing the first six-months of treatment, and a recursive state representing continued treatment use. The use of multidimensional matrices allowed the probability of the patient progressing to the next treatment state to vary according to how long they had been in their current state. However, the model was based on the assumption of treatment independence, whilst an alternative model (NICE CG137),¹⁸ which implemented the same antiepileptic drug sequencing structure, but within the conventional Markov cohort framework, applied a reduction to the efficacy of treatments used as second line (Chapter 5, Section 5.6). The York epilepsy model only accounted for one level of treatment response, complete seizure freedom or not, whilst the NICE CG137 also incorporated partial response. In theory, multidimensional matrices can also be used to allow cohort models to reflect 'patient history', or previous treatments; however, no included study did this.

7.4.2.2 Individual sampling models

One of the reasons why implementing of treatment sequencing can be complex is that many decisions are dependent on the attributes of individuals, such as response to previous treatment and disease duration. An alternative approach to account for the various potential individual pathways that can be taken is to use an individual patient simulation. The use of individual patient simulation allows the models to be based on both disease activity and the treatment history of patients.³⁷³ Virtual patient histories are constructed by first developing a set of baseline patient characteristics sampled from appropriate distributions or a multivariate distribution. The patient's history is then allowed to develop over time, taking into account the properties of the various treatments applied. This development can be done either for each patient's full remaining lifetime or with a maximum time horizon in the model.³⁸⁸ Relevant outcomes are then developed from the patient histories using a computer random number generator. Some of the proposed advantages of using individual patient simulation include the ability to account for a heterogeneous patient population, and that the progress of specific subgroups can be tracked within the model.^{344 373} The fact that patients' history is memorised means that it can also potentially be used to facilitate simulating treatment selection based on past patient characteristics, for example, previous treatment or reasons for discontinuing previous treatment.^{344 361}

382

State transition individual patient simulation

The individual patient simulation state transition model is similar to that of the Markov cohort model except that it simulates each member of the cohort one at a time. The number of states required are greatly reduced as the model keeps track of each individual's history. Tracker variables can include past states, risk factors, time in states, and time since last event. The individual-based model is also not limited by the Markovian property. However, its potential limitations are that it is less straightforward to implement, and not transparent. It is also still based on the use of mutually

exclusive states, and transitions between states can only occur at specified or fixed time intervals. Included modelling studies used this approach to investigate treatment sequencing within a homogenous group of patients, and none investigated patient subgroups. This approach was also used to evaluate fixed sequences, based on specific treatments or class of treatments used at a specific line. They did not include treatment skipping, or a differential selection of subsequent treatment due to reason for discontinuation.

State transition individual patient simulation modelling was mainly used in rheumatology studies. For example, a series of models commonly referred to as the Sheffield models³⁸² specifically chose this approach to model the sequential use of treatments over time and the uncertain duration of effect on each patient, which is typical for rheumatoid arthritis. Most models were set up to measure the patients' disease activity over their lifetime. This was generally done using the HAQ-disability score as a proxy for disease activity, which was measured at six-monthly intervals. However, three studies were able to model disease activity using multiple outcomes. A number of studies, including most of the Sheffield models, had access to individual patient level data, which was generally taken from a patient registry. This enabled changes in the HAQ score and other parameters or key events in the patient histories to be calculated using a series of multivariate analyses, allowing adjustment for important covariates. The regression analyses provided the parameters that the model used for simulating the path each individual patient would take. Baseline characteristics included the HAQ score, and for some models also included disease duration and number of previous conventional DMARDs used, as well as generic patient characteristics such as age and sex. The models themselves were based on a fairly simple structure, with health states representing either treatment response, non-response, or death.^{218 224 225 274} Patients who did not respond switched to the next treatment, whilst those who responded continued treatment for a predefined period. One of the earlier Sheffield models included serious adverse effects as an additional state, which, similar to non-response, led to treatment switching.²¹⁸ Patients who were classified as 'treatment success' or 'non-responders' were also assessed for the occurrence of mild to moderate adverse effects in order to account for their potential treatment cost. Treatment sequencing was represented as a series of health states in one model (Diamantpoulous model), but the model was essentially still based on the same structure as other models.^{234 235} Treatment response was measured at the end of a cycle, and those with no response moved to the next treatment, whilst those with a response remained on treatment until withdrawal. The majority of models also included allocating patients, after initiating treatment, to different response categories or thresholds, with only those who achieved a specific threshold (representing remission) continuing the same treatment. The observed improvement in the HAQ score during the first six months of treatment was then assumed to relate to the level of response, not treatment, allowing universal HAQ reductions for each level of response to be applied across all treatments. The model (Finchkh model), which accounted for the most complex decision problem, was based on a series of statistical regression analyses.²³⁸ It was developed to investigate three different management strategies for patients with very early rheumatoid arthritis: a stepped care approach starting with non-steroidal anti-inflammatory drugs and other symptomatic treatments, the

use of early conventional DMARD, and the use of early biological therapy. The model also accounted for the fact that the patient could follow one of three different disease courses, which could not be predicted at onset. This allowed the model to account for the fact that patients with rapidly progressing disease cannot be reliably identified in early disease, and that expensive biological therapy may be given to patients who would have responded satisfactorily to conventional DMARDs.

The series of Sheffield models provides an example of where an initial model was modified and built upon with time, and used to address different decision problems.^{218 224 225 274} Other examples of this includes the Tran-Duy model^{272 273} and the Birmingham Rheumatoid Arthritis Model,^{219 227 229 245 254} both of which used discrete event simulation. Patients entering these models were newly diagnosed patients, with the decision population developed as part of the modelling process. This meant that, for the Birmingham model, only a single data set for the starting population, or knowledge of the distribution of patient baseline characteristics was required, irrespective of which decision problem the model was used for.

Discrete event simulation

In a discrete event simulation the course of the disease and its management are represented as a chronological sequence of events that can happen to individuals over time, and the subsequent effects of these events on their current and future health.²⁷² This was conceived within the included models using both a fairly simple structure,^{219 245} similar to that used in the state-transition models, and a very complex one.^{199 273} As a complex structure it was also able to accommodate the unpredictable nature of disease progression, multiple treatment outcomes, and the fact that not all patients go on to receive subsequent treatments in the sequence. It was also used to investigate treatment sequencing within a heterogeneous patient group, and to account for different treatment selection for patient subgroups, such as patient with rhesus positive or negative rheumatoid arthritis.

Treatment sequencing was implemented within the discrete event simulation model in three different ways. This included:

- i. Using fixed treatment sequences (for example the initial Birmingham Preliminary Model);
- ii. Based on the random selection of pre-defined sequences (for example the subsequent Birmingham Rheumatoid Arthritis model)
- iii. Developed as part of the modelling process by selecting individual drugs, using a random process, at specific points in the sequence (for example in the Tran-Duy model)

As a complex structure it was successfully used for implementing treatment selection and cessation based on algorithms reflecting specific clinical guidelines and practice. This required the implementation of stopping rules and specific sequences of treatments to be followed based on the different reasons for quitting treatment. Similar to the state transition models described previously, states were used to characterise changes in disease activity and treatments. However, unlike a Markov model, the states in a discrete event simulation are not necessarily mutually exclusive.²⁷³ This

allowed greater flexibility for modelling treatment selection based on the disease activity and treatment history of the patient. Conceptually, patients could be in one of multiple disease activity phases whilst being treated on a specific drug used at a particular point in the pathway.

The discrete event simulation approach was frequently chosen in order to implement the variable time to quitting treatment, as this may not be constant and differed between treatments. For example, a discrete event simulation was chosen for this specific reason in one model (Birmingham epilepsy model) of treatment sequences for epilepsy.¹⁹⁹ The model was able to account for different levels of treatment response, and the fact that patients with either complete or partial response may then progress to a successful or unsuccessful drug withdrawal. They may also have different preferences about trying monotherapy or adjuvant therapy, based on prior treatment experience. However, the main challenge was the limited data on time to event from the available evidence. For example, where time to treatment withdrawal was reported, it could not be disaggregated according whether treatment was discontinued due to adverse effects or lack of effect. The discrete event simulation was also chosen as the preferred approach for the Birmingham rheumatology models as it allowed treatment duration to be modelled as a continuous distribution. Time on each DMARD was sampled from a Weibull distribution, with parameters appropriate to the particular treatment. The earliest preliminary version of the model represented one of the simplest structures used by included discrete event simulation models.²⁴⁵ In the Birmingham preliminary model only one initial patient characteristic was required, which was the patient's remaining lifetime, and only two things were sampled for each treatment, which included the time on treatment, and whether the treatment was quit due to toxicity or lack of effectiveness.²⁴⁵ The latter was used because the last active treatment could not be given if either components of the combination therapy had previously been quit due to toxicity. The effect of each treatment was assumed to produce a set improvement in the patient's health related quality of life (HRQL), which would be lost on quitting treatment. One of the main changes in the subsequent Birmingham rheumatology Arthritis Model was to define the patient's health state in absolute terms using the HAQ score, rather than quality of life.²¹⁹ The HAQ improvement on starting treatment was modelled as a fixed decrease in the score, based on the average decrease for all patients receiving a specific treatment. The model structure was still fairly simple though, with the main loop consisting of 'start new treatment' – 'on treatment' – 'quit DMARD' – 'select next treatment'. The second version of this model incorporated two important improvements, which included allowing the HAQ improvement on starting treatment to vary between individuals, and for time on treatment to explicitly consider early quitting.²²⁷ The model allowed for two periods of early quitting, where withdrawal was assumed to be due to toxicity in the first, and could be due to toxicity or inefficacy in the second. The third version of the model included the use of probabilistic sensitivity analyses.²⁵⁴ The model also allowed different choices of treatment options depending on toxicity of previous treatments.

The main advantage of a discrete event simulation over an individual patient state transition is that it requires fewer calculations as it is not governed by the analyses of the data at set intervals, or time cycles. Chrosny *et al.* have shown that Markov models can introduce bias to the absolute costs and

QALYs due to the simplifying assumptions of fixed cycle length and half cycle corrections, which discrete event simulation models do not suffer from.³⁹⁶ In a discrete event simulation the analysis of the model is triggered by the occurrence of an event, at which point the model asks what and when is the next event for the patient, unlike a Markov process, which asks what events are occurring at regular intervals. Each event occurs at a specified time, and where there are competing events, the patient will 'jump' to the event to which the sampled time is shortest.³⁶⁴ The main disadvantage of using a discrete event simulation is the extensive data required for model parameterisation, including time to event data.^{367 373} Patient level data are also preferred for implementing discrete event simulation, but they can also be based on aggregate data.³⁷³ Only two models were based on individual patient level data, both were used in rheumatology studies. One of these models (Tran-Duy model) was able to incorporate a complex treatment and decision process, but still had to use simplifying assumptions regarding treatment sequencing effect, due to limited data. The data source included an inception cohort and a patient registry.²⁷³ It was assumed that the effectiveness of a specific drug was independent of the identity and the cause of failure of the drugs that had been given previously.

7.4.2.3 Dynamic population models

Open models allow new cohorts to enter over time, and are sometimes referred to as population models.³⁶⁹ An open cohort model and individual patient simulation were identified that evaluated treatment sequences. This included a budget impact study²⁴⁹ that used a dynamic Markov cohort model, and the Cardiff diabetes model²⁵⁵ based on discrete event simulation. The included studies did not show any clear advantage of using an open model, over the other included modelling approaches, for modelling the clinical effectiveness of treatment sequences. However, the Cardiff model was able to account for a dynamic disease process whilst incorporating treatment sequencing. The Cardiff model also provides another example of a model that utilised the natural history and time to event data from an observational study, and treatment effect profiles based on clinical trials. The model is continually being developed and updated and is capable of running using various levels or types of data.

Table 7.3: Summary of the different modelling approaches used and their advantages and disadvantages for evaluating treatment sequences

Description of modelling approach	How treatment sequences are conceptualised in the model	Further complexities in the decision problem that were allowed for	Advantages	Disadvantages	Examples
COHORT-BASED MODELS: Simulates a closed group of individuals					
Decision tree (DT)					
Depicts the risk of events over a fixed time period.	Naturally provides a diagrammatic representation of treatment sequences, possible outcomes, and events.	Relapse treated with a previous successful treatment. Duration of response differs according to levels of response. Reason for discontinuation impacts selection of subsequent treatments. Some treatments administered for a fixed period only. Toxic death and all-cause mortality have different probabilities and timing. Not all patients receive all treatments in sequence (<i>implemented in conjunction with portioned survival</i>).	Can be relatively straightforward to develop and not computationally intensive. Can be easy to interpret and transparent. Can include a large number of different treatment sequences within the same model. Can be used in conjunction with other methods.	No explicit time component; governed by fixed timing of outcomes and events. Only allows one way progression. Cannot handle looping/recurring events easily. Can become exponentially complex with additional events and disease states. Poorly suited for complex scenarios.	Dranitsaris model NICE CG81 NICE CG152 Sciatica model Frankum model Knoster model
Stochastic decision tree (DT)					
A type of DT which allows the comparison of variable distributions	Same as DT	Same as DT	Same as DT	Same as DT	Advanced simulation model Greenhalgh model
Markov cohort					
Simulation of a hypothetical cohort through a set of health states over time, which is divided into equal intervals (cycles). Involves time-dependent transition between states. In Markov chain transition probabilities (TPs) are constant over time.	Implemented using three different approaches: i) a series of treatment-specific Markov states; ii) as a series of treatment-specific states (or lines) along with additional temporary states representing e.g. adverse effects, relapse; iii) as a Markov cycle tree, with Markov states used to represent different levels of disease activity or natural history.	Not all patients receive all treatments in sequence. Duration of response differs according to levels of response and treatment line. Probability of continuing treatment and developing toxicity varies with time and for each treatment (<i>using tunnel states</i>). Reason for discontinuation impacts selection of subsequent treatments. Some treatments administered for fixed period. <i>Cycle trees used to account for:</i> Consequence of adverse effects. Different levels of treatment response Some patients continue treatment despite not achieving full/clinical response. Fluctuating disease activity. Complex treatment pathways.	Can be relatively straightforward to construct and communicate. Has a time component; events can occur at any time. Allows looping/recurring events. Transitions can be unidirectional or bidirectional. Can be used in conjunction with decision tree (Cycle tree). The use of cloned subtrees enables ease of update.	Markov assumption (<i>memoryless</i>): prohibits TPs being dependent on time spent in the state, or previous states visited (<i>can be overcome using additional states and stratification</i>). Patients can only be in one state at a time. Transitions limited to fixed intervals defined by cycle length. Cannot account for multiple events within one cycle (<i>can be overcome using short cycles or cycle trees</i>). Occurrence of events assumed to be constant over time (Markov chain). Exponential increase in complexity with increasing number of states.	Albert model Maetzel model Welsing model Tanno model Wu model York psoriasis model Beard model Cameron model Davies model Heeg cancer model Lee model Orme model Sawyer model NICE CG137 Shepherd model Smith model Soini model Tebas model Wong model
Semi-Markov cohort					
Incorporates the use of a multiple-dimension transition matrix. Assumes TPs depend on the current state, and the time	A series of treatment-specific health states.	Probability of treatment failure decreases with time on a specific drug.	<i>Same as Markov cohort</i> Reduced impact of Markovian assumption (not memoryless; incorporates time dependency).	Patients can only be in one state at a time. Transition limited to fixed time intervals defined by cycle length. Transitions can only occur at fixed intervals. Only one transition allowed per cycle.	York epilepsy model

spent in each state depends upon the current and next state.				Becomes more complex with added states.	
Partitioned survival					
Simulation of a hypothetical cohort through a set of exhaustive and mutually exclusive health states over time. Time spent in each health state calculated from the area under the curve of survival functions.	A series of treatment-specific health states.	Decreasing probability of remaining on a given treatment with time.	Can be relatively straightforward to develop and not computationally intensive. Non-data intensive. Transparent. Area under the curve can be calculated continuously over time; no cycles required. Can be used in conjunction with DT.	Cannot account for complex treatment sequencing algorithms or additional attributes (e.g. adverse effects, disease duration).	Schadlich model NICE CG131 (with DT) Hind model
INDIVIDUAL SAMPLING MODELS: Simulates one individual at a time. Tracks the past health states of each individual and stochastically models the risk of future events					
State transition (STM)					
Simulates each individual through a set of exhaustive and mutually exclusive health states over time, which is advanced in fixed intervals.	Fixed treatment sequences; disease activity monitored for each individual over time. Health states generally represented response or non-response to each treatment, with the addition of adverse effects as a separate state in some models.	Duration of response differs according to levels of response. Fluctuating disease activity (changes in disease activity assumed to relate to level of response, not treatment). Complex treatment algorithms. Patients follow different disease courses, which cannot be predicted at the onset.	Not limited by Markov assumption (eliminating need for excessive number of states). A large number of characteristics can be ascribed to individually simulated patients. Access to individual patient data enabled key parameters and events in patient histories to be calculated using multivariate regression, allowing adjusting for important covariates. Can account for heterogeneous population.	Transition limited to fixed time intervals defined by cycle length. Cannot account for multiple events in one cycle. Can be computationally intensive.	Sheffield Etanercept model Bansback model Sheffield BSRBR model Diamantpoulos model Sheffield AHRQ model Kielhorn model Kobelt model Holmes model
Discrete event situation (DES)					
Simulates time to an event and subsequent events for each individual. Probability of the occurrence and timing of an event is determined by random sampling of a probability distribution. Simultaneously varies multiple variables; inputs vary over time.	Implemented in three different ways: i) fixed treatment sequences; ii) random selection of pre-defined sequences; or iii) developed as part of the modelling process by selecting individual drugs, using a random process, at specific points in the sequence.	<i>When conceived as a simple structure:</i> Variable time to quitting treatment. Duration of response differs according to levels of response. Fluctuating disease activity. Reason for discontinuation impacts selection of subsequent treatments. <i>When developed as a more complex structure:</i> Treatment selection and cessation based on algorithms reflecting specific clinical guidelines. Unpredictable nature of disease progression. Multiple treatment outcomes. Not all patients go on to receive subsequent treatments in the sequence. Differential treatment selection for subgroups.	Can ascribe a large number of characteristics to individually simulated patients. Can account for heterogeneous population. Not limited by the use of fixed time advancement (cycles). Patients can simultaneously be in multiple states, and experience different events. Allows for modelling of complex scenarios and treatment algorithms. Computationally more efficient than STM. Can be easily adapted to incorporate additional events or patient attributes.	Model structures can be difficult to communicate and interpret. Computationally challenging in terms of model design and running it.	BPM / BRAM Tran-Duy model Lindgren model Birmingham epilepsy model Denis model Heeg schizophrenia model
(OPEN) POPULATION-BASED MODELS: Allows new cohorts to enter over time					
Non-terminating population based simulation					
Individual sampling model (DES)	Pre-specified clinical thresholds used to invoke escalation to next treatment.	Dynamic disease process (dynamic equations used to project haemoglobin levels over time).	Same as DES	Same as DES	Cardiff T2DM model
Markov multi-cohort model					
Cohort model (Markov)	Markov cycle tree (Markov states represented individual treatments and 'switching' [entry/exit state]).	Impact of adding a new drug on health care budget assessed using prevalence approach (target population kept constant over time - entry of newly diagnosed cohort at each cycle).	Same as Markov Cohort	Same as Markov Cohort	Launois model

BPM Birmingham preliminary model; BRAM Birmingham Rheumatoid Arthritis model; NICE National Institute for Health and Care Excellence; T2DM type 2 diabetes mellitus

7.4.3 Summary of software and implementation issues

Potential limitations for using individual sampling models include the computational burden, expertise required to implement them, and the need for specialist software.³⁷³ A number of studies mentioned that implementing the model could be time consuming, but this was also reported for implementing probabilistic decision trees²²³ and not just the more sophisticated approaches such as discrete event simulation.²⁷³ A number of studies reported choosing a discrete event simulation over a state transition model due to the improved computational efficiency,^{199 251 272 273} but none of the included studies compared the two approaches for modelling treatment sequences.

The most commonly used software packages for modelling treatment sequences were Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) and TreeAge Pro (TreeAge Software, Inc., Williamstown, MA, USA). TreeAge Pro supports the development of Markov models through the decision tree structure. The use of specialist software was reported to be an advantage, especially for implementing Markov cycle trees and discrete event simulation. One of the limitations of implementing an individual patient level state transition model within a spread sheet package using programming language (such as Microsoft Visual Basic) is that as the complexity of the model increases it becomes harder to detect programming errors.³⁷³ When implemented within TreeAge Pro using the 'microsimulation' function, the tracker variables can be used to capture significant prior events and drive the transition matrix in a more transparent manner.³⁷³ TreeAge Pro (from 2014 version) also includes time to event functionality, which means that it can be used to implement a discrete event simulation. One discrete event simulation model (Birmingham Rheumatoid Arthritis Model) was developed using both TreeAge DATA Pro (2004) and Borland Delphi (Borland Software Corporation, Scotts Valley, CA, USA), and the two software platforms were compared in terms of efficacy and transparency.²¹⁹ The use of TreeAge DATA allowed the model to be constructed transparently, with the logic open to inspection, whereas the use of Borland Delphi programming language sped up the analysis allowing extensive sensitivity analysis to be conducted, but it required programming skills on behalf of the modeller, and led to the loss of transparency in the model.²¹⁹ There are also bespoke software packages available for developing a discrete event simulation model that can be used to guide the model development.³⁶⁴ Davis *et al.* (2014) suggest that using a bespoke simulation package such as Simul8 (Simul8 Corporation, Boston, MA, USA) has a number of benefits including: graphical representations of the simulation, easy debugging and validation, quick and easy development of new models, random number control, and easy sampling of time-to-event variables from commonly used distributions.³⁷³ However, Simul8 is not currently included in NICE's list of recommended software packages for conducting technology appraisals.^{373 397} Only one included study reported using a dedicated discrete event simulation software package, with most models having been implemented using Delphi language. It could be argued that the use of specialist software is likely to increase the skillset required to implement the model and potential cost of purchasing, but the use of common software packages for implementing complex models would also require extensive programming skills. A promising new addition to the choice of approaches for conducting economic modelling for health technology assessment is the discretely integrated condition event (DICE) simulation,³⁹⁸ which

can incorporate multiple different modelling approaches within the same framework and is implemented in the commonly used spreadsheet package Microsoft Excel. However, this is yet to be used for modelling treatment sequences.

In the past, the need to conduct probabilistic sensitivity analysis has been viewed as a barrier to using patient-level simulation due to the computational requirements of simulating both a large number of patients to limit stochastic uncertainty, and a large number of parameter samples to evaluate parameter uncertainty.³⁷³ Overall, this did not appear to be problematic, or to affect the choice of modelling approach used in the current review, with all modelling approaches having a number of examples where probabilistic sensitivity were both included and not included.

7.4.4 External validation of a sequential treatment model

The lack of sequencing studies means that data for performing external validation of a treatment sequencing model may not be available. Modelled estimates could also be compared to those produced in real life (clinical experience or real-world data), or previously developed models.⁸¹ However, observational studies are likely to suffer from bias (Chapter 5, Section 5.9.9), which could result in the model being validated against the type of flawed estimates that it was designed to replace.⁸¹

The type of data required for validating the model outcomes will also be dependent on the model structure.²⁴³ One study (Heeg, 2015) that developed a Markov model (Heeg cancer model) for predicting overall survival for treatment sequences for multiple myeloma compared the predicted findings of the model with those reported in RCTs of first-line treatment in order to assess external validity.²⁴³ This was considered appropriate by the authors because the model used transition probabilities, representing treatment switching or mortality, that were response and line specific, not treatment-specific. The findings showed that the median survival predictions for one treatment sequence underestimated the median survival reported in the VISTA trial (of newly diagnosed patients), which may be considered unusual, as the model input for first-line treatment was largely derived from VISTA.^{243 377} The main reason for this underestimation was the use of the APEX trial (of relapsed patients) data to estimate the overall survival and time to next treatment transition probabilities for second and third-line treatments. Patients who went on to have second line treatment after VISTA would have received on average better treatment compared to APEX patients, given that the VISTA trial is more recent than the APEX trial.^{243 378} Hence the second and third line transition probabilities derived from APEX are likely to be higher for patients after VISTA, causing an underestimation of the modelled median overall survival.²⁴³

7.4.5 Structural complexity of the modelling approach

The decision problem relating to treatment sequencing can be complex, especially when considering its impact on a chronic condition such as rheumatoid arthritis. The extent of the structural complexity of the model will in turn also depend on the clinical scenario being modelled, the extent of the

treatment sequences being considered, the need to incorporate algorithms or various logical rules on which treatment to use next, and whether there is a need to account for variation in treatment duration, which may not be constant. There are three important questions regarding model complexity that needs to be considered when choosing which modelling approach to use:

- i. Is there a point at which the level of complexity required means that a specific modelling approach is preferable to another?
- ii. Does the decision problem necessitate a complex modelling structure?
- iii. Is there sufficient data to parametrise a complex model?

The findings of the review of modelling studies shows that it is too simplistic to think of cohort models such as decision trees and Markov models as examples of simple modelling approaches, and that individual patient simulation approaches represent complex modelling approaches. According to Barton *et al.* the complexity of the model is dependent on the size and not the technique used,³⁶¹ which was also reflected by some of the included studies. The review included a discrete event simulation model²¹⁹ that had a much simpler structure than many of the included Markov cohort models and some of the decision tree models. Both the Markov cohort modelling approach and individual patient simulation were used to implement complex decision problems relating to treatment sequencing. Simpler models are usually easier to understand than complex models, and thus easier to validate.³⁶¹ Cohort models, such as decision trees and Markov models, can be more transparent than individual patient simulation approaches, but this advantage can soon be lost when they are adapted to account for further complexities in the decision problem.²⁶⁶ The inherent flexibility of individual patient simulation enables a simpler model to be constructed, compared with using a Markov cohort model, where the need to overcome some of the limitations of the modelling approach in order to allow treatment sequencing and specific characteristics of the disease to be modelled adds another layer of complexity to the structure. The use of a discrete event simulation, which involves codifying the behaviour of a complex system as an ordered sequence of well-defined events in time, provides even greater flexibility than an individual patient level state transition model, especially in terms of intuitively modelling treatment sequencing algorithms.^{272 273}

Treatment sequencing does not always mean that the decision problem is necessarily complex, and all models essentially represent a simplification of reality. The extent or level of complexity in the decision problem, which was accounted for in the models, varied quite considerably within the included studies, even within the same disease condition. For example, three studies evaluating similar antiepileptic drug sequences chose different modelling approaches and incorporated different levels of the disease process.^{18 199 276} (This is also described in the Appendix Volume I, Section C5, as all three studies were conducted on behalf of NICE). The decision problem can be simplified by modelling a limited number of treatments, streamlining the disease process, and using a short time horizon.

The greater the complexity and sophistication of the model, irrespective of the modelling approach used, the more extensive the data and resources required for developing and parameterising them. One of the reasons given for simplifying certain aspects of the decision problem regarding treatment sequences was due to limited data. For example, a few studies reported having to use a short time frame for modelling treatment sequences for rheumatoid arthritis, because a longer time horizon implied too many assumptions.^{222 223 228 260-262 275} Complex models may be able to represent treatment sequencing and related issues better, but if the data required to parameterise the model is not available, or limited, then a simpler approach may be preferable. An important but unaddressed question is, at what point will a complex model be unrealistic because of the use of too many simplifying assumptions, and the model output becomes non-informative or even misleading? The data on sequencing effects, or treatment effects that are conditional on the previous treatments used, was missing or limited in most cases. However, on the other hand, a model based on an over simplification of the decision problem and clinical practice is also unlikely to be useful for decision makers.

7.4.6 Using a conceptual framework to inform the structural complexity of the model

The process of developing a model structure requires the modeller, in conjunction with other stakeholders, to make a number of structural model development choices regarding what is relevant and what can be considered as irrelevant to the decision problem.³⁹⁹ The development of a qualitative conceptual model as a pre-cursor to the quantitative decision analytic model provides a powerful tool to help understand of the nature of the decision problem, to inform choices about the level of structural complexity required by the proposed model, and assist communication regarding its intent and structure.³⁹⁹ A conceptual model provides a visual picture of the decision problem and all relevant aspects of the disease. Conceptual modelling frameworks have been developed, which guide modellers step-by-step through the process of developing a model structure¹⁸³ and encompass both the process of conceptualising the problem and disease, and conceptualising the structure of the subsequent mathematical model. For example, Squires *et al.* have recently published a conceptual modelling framework for developing the structure of public health economic models.⁴⁰⁰ The ISPOR-SMDM Joint Modelling Good Research Practices Task Force have also developed guidance to inform conceptual modelling for health economics.³⁷⁰ The guidance recommends that the initial model conceptualisation process should:

- i. Be linked to the problem and not based on data availability
- ii. Be used to identify key uncertainties in the model structure where sensitivity analysis could inform the impact of structural choices
- iii. Follow an explicit process (expert consultations, influence diagrams, concept mapping, or similar method) to convert the conceptualisation of the decision problem into an appropriate model structure to ensure that the model reflects current theory of disease or the process being modelled

The guidance also advocates simplicity in the model structure for transparency, ease of validation, but also note that the structure must be sufficiently complex to answer the question, and should maintain face validity.³⁷⁰

7.4.6.1 INTEGRATE-HTA

The recently completed EU-funded research project INTEGRATE-HTA, which ran from January 2013 to December 2015, can provide some very useful information here. (The INTEGRATE-HTA project was first discussed in Chapter 5, Section 5.9.10.3). One of the project's main aims was to develop a process that supports asking questions that are relevant and finding answers that fit the questions, rather than substituting difficult to answer questions with those that *can* be answered.⁴⁰¹ The main product of the project was a series of seven publicly available guidance documents on the integrated assessment of complex health technologies.⁴⁰² A themed issue presenting the results of the project, as well as the application of some of the guidance by the Canadian Agency for Drugs and Technologies in Health (CADTH) was published in the International Journal of Technology Assessment in Health Care in December 2017.⁴⁰³

The INTEGRATE-HTA project included developing guidance on the use of logic models in health technology assessments or systematic reviews of complex interventions.⁴⁰⁴ Logic models encompass any of the following: conceptual frameworks, analytic frameworks, concept maps, or influence diagrams.⁴⁰⁴ Logic models were described as providing a key means for integration across different parts of the health technology assessment,⁴⁰² can be used to 'think through' the multiple components of a complex intervention in context, enhance the transparency of underlying assumptions, and can assist in communication within the technology assessment / systematic review author team and with a range of stakeholders.⁴⁰⁴ Four different types (generic logic model, *a priori* logic model, iterative logic model, and staged logic model) and a number of sub-types were identified, each with different strengths and limitations. It was also acknowledged that the process of developing any type of logic model can take a significant amount of time, potentially delaying subsequent stages of the already time-consuming health technology assessment / systematic review process.⁴⁰⁴

7.4.7 Developing a flexible model that can be re-used or further developed

The time and resources available for conceiving and implementing a complex decision analytic model within a health technology assessment is limited. It may therefore be tempting to try to develop a simplified model. An alternative approach to overcome this would be to re-use or further develop an established model. In health technology assessment a new model is generally developed for each decision problem. However, the review identified some good examples where an existing model was further developed over time. This included the Tran-Duy, Birmingham, and Sheffield rheumatology models, and the Cardiff Diabetes model. Most of these models were developed using a discrete event simulation. The discrete event simulation, which was originally developed in industrial engineering and operations research, is generally used to model physical systems where an initial model is refined and built upon over time.³⁶⁸ Discrete event simulations can be easily adapted to incorporate additional

events or patient attributes and, as such, may lend themselves to iterative decision making processes or repeated use, whilst adapting decision tree or state-transition models to include additional health states or patient attributes can be time consuming, particularly if the model is implemented within a spreadsheet package.³⁷³ However, the Shepherd model provides an example of how a complex Markov model can be made easier to update.²⁶⁶ The model was implemented as a Markov cycle tree, which included two subtrees referred to as clones that were attached to different locations, or nodes, in the tree. The use of cloned subtrees meant that only one 'master' copy needs to be maintained rather than requiring maintenance of numerical identical trees. This would also aid updating the model.

The extent of the available data to inform treatment sequencing effects may improve over time, which is another reason for developing a model that can be easily updated. This is especially true bearing in mind the current trends in developing large datasets,³²⁷ and the requirement of many funders that data arising from their grants are shared.⁷⁹ Alternatively, an existing model can be re-run using a different modelling approach, or a simple model further developed to incorporate further complexities in the decision problem that may have evolved over time as more treatments become available. This provides a case for publishing the pre-cursor conceptual framework (Section 7.4.6) alongside the modelling data.

7.4.8 Potential decision analytic modelling approaches not covered by included studies

A number of new modelling approaches that were not used by the included studies evaluating treatment sequences are now being used by health economists. These include attributes that may be useful for modelling treatment sequences in the future.

7.4.8.1 Agent-based simulation

An example of an alternative modelling technique that has not yet been used for evaluating treatment sequences is agent-based simulation.⁴⁰⁵ It has many similarities to discrete event simulation, but unlike the system-based rules used in discrete event simulation, the agent-based simulation applies rules to agents (individuals), or group of agents, and their responses, which depend on the individual characteristics of the agents and can change either over time or due to interactions with the environment.^{368 406} Agent-based simulation is more complex than discrete event simulation and can require considerably more data to represent the heterogeneous population.⁴⁰⁶

An example of an agent based simulation used for evaluating complex multicomponent public health interventions is Archimedes,⁴⁰⁶ which was introduced in Chapter 1, Section 1.4.2.2. Archimedes is an object-orientated, continuous-time simulation model designed to be both comprehensive and deep, and covers the anatomy, pathophysiology, tests, treatment, and outcomes pertaining to a specific condition, for example diabetes.⁹⁶ It can be applied to a wide variety of clinical and administrative decision problems and can be used to simulate RCTs of interventions.⁹⁶ However, it relies on a large amount of processing power and data. For example, the level of detail in the diabetes model

corresponds to that found in general medical textbooks, patient charts, clinical practice guidelines, and designs of clinical trials.⁹⁵

7.4.8.2 Whole disease modelling

Another potentially useful development in health economic modelling that incorporates the use of a single comprehensive model to address both current and future decision problems is 'whole disease modelling' (Section 1.4.2.2).⁹²⁻⁹³ This represents a system or disease-level modelling approach that simulates treatment pathways from preclinical disease through to death. Whole disease models are developed using the discrete event simulation framework, and can be used to inform a range of different decision problems relating to any part of the patient pathway, including screening, diagnosis, active treatment, and palliative care. Examples have been developed to inform NICE clinical guidelines for colorectal cancer,⁹³ prostate cancer,⁴⁰⁷ and atrial fibrillation.⁴⁰⁸ These were not considered in the review of modelling studies, as treatment sequencing was not extensively evaluated; they were based primarily on the assumption of treatment independence, and therefore did not contribute any new information. For example, due to the lack of randomised clinical trials that evaluated planned sequences of treatments for advanced prostate cancer, it was assumed that first-line palliative treatment was the sole determinant of overall survival due to prostate cancer.⁹⁴

7.4.8.3 Markov influence diagrams

Influence diagrams are a graphic representation of the causal relationships between decisions, external factors, uncertainties and outcomes.⁴⁰⁹ They provide a way of representing decision problems that are mathematically equivalent to decision trees and Markov Models.³⁷¹ However, unlike decision trees, which are open graphs, and Markov models, which are partially cyclic graphs, influence diagrams are closed, directed graphs without recursion.³⁷¹

Markov influence diagrams are a new type of probabilistic graphical model that extends influence diagrams in the same way that Markov cycle trees extend decision trees.⁴¹⁰ They can be used to easily build and evaluate complex models within the paradigm of state-transition models, in which many health economists feel more comfortable.⁴¹⁰ Their potential advantages include the fact that they can contain several variables per cycle using a causal graph, they can model various patient characteristics without multiplying the number of states, and they can represent the history of the patient without using tunnel states.⁴¹⁰ If implemented using the OpenMarkov software (<http://www.openmarkov.org/>), probabilistic sensitivity analysis can be conducted without writing any code.⁴¹⁰

7.4.9 Comprehensive decision analytic modelling

The comprehensive decision analytic modelling approach, which was introduced in Chapter 1, Section 1.4.2.1, is where both the meta-analysis (including network meta-analysis and meta-regression) and the economic model are conducted within a single framework, or one coherent model.⁸⁸ This approach allows the probabilistic sensitivity analyses to draw directly from the posterior

distribution output of the meta-analysis. This approach was used by a number of included studies using Markov cohort modelling as part of the economic evaluation for a NICE technology appraisal or clinical guideline.^{207 257 276 278} However, none have yet used meta-analysis that accounted for treatment sequencing effects.

7.5 CHAPTER SUMMARY AND NEXT-STEP

7.5.1 Chapter summary

A wide range of modelling approaches have been used for evaluating treatment sequences. The decision problem regarding treatment sequencing can be complex, especially when extensive treatment sequences are considered. There may be a need to account for differential algorithms that dictate the choice of subsequent treatments, as well as additional aspects of the natural history of the disease that need to be considered. More sophisticated models based on individual sampling are better able to accommodate the complexity in the decision problem than the simpler cohort models. Examples where cohort models were successfully adapted to accommodate additional complexity in the decision problem were identified, but these were no longer simple models. The discrete event simulation approach appeared to provide the optimum approach in terms of intuitively modelling treatment sequencing algorithms, computational efficacy, and ease of updating, however, they require more extensive modelling skills and specialist software. The added advantage of using a state transition individual patient model, over a complex Markov cohort model structured around a Markov cycle tree is unclear. A discrete event stimulation appears to have some advantages to that of an individual patient state transition model, especially when implemented using specialist software. The greater the complexity and sophistication of the model the more extensive the data requirement were for implementing them.

The limited evidence base regarding treatment effects that are conditional on the previous treatments used meant that these complex models were frequently supplemented with simplifying assumptions, even when an individual patient level data source such as a national patient register was available. In fact, the overall findings tended to indicate that a single data source is unlikely to be sufficient for modelling treatment sequences. The list of simplifying assumptions developed in Chapter 6 will provide a useful tool to clarify what assumptions were made regarding treatment sequencing effects in implementing a complex or sophisticated models. Another potential disadvantage of developing complex and sophisticated models is the time required to develop and implement them, which is generally limited for health technology assessment. One potential solution to this would be to use or further adapt an established model which has already been developed and tested. Another potential option would be to re-build an existing model using a different, more flexible, modelling approach.

The choice of an existing model to re-use or further develop is likely to be based on those developed for a similar decision problem, however, models used for other health conditions may also be informative. The table (Table 7.3) summarising the existing models developed for evaluating

treatment sequences, and the type of additional features that they accounted for, could help to choose relevant models. Information on any accompanying conceptual or logic models would also be useful. A number of models were developed with the intention of being able to use them to inform future decisions, most of which were developed using a discrete event simulation.

7.5.2 Next-step

This chapter provides a summary table (Table 7.3) that allows the comparison of different modelling approaches used and the individual attributes they accounted for. The next step is to incorporate the data identified in Chapter 6 on the type of simplifying assumptions made when modelling treatment sequences to this table. This can then be used as part of a framework for assisting future researchers to choose an appropriate modelling approach for a specific treatment sequencing decision problem, or identify a published model that can be developed or further adapted for their own purposes.

CHAPTER 8: FRAMEWORK FOR EVALUATING TREATMENT SEQUENCES

8.1 CHAPTER SUMMARY

The aim of the thesis was to develop a framework for conducting quantitative evidence synthesis of the effectiveness of sequential treatment options within the context of undertaking evidence reviews for informing clinical and policy decision making. The current chapter presents the framework for evaluating treatment sequences, and is one aspect of the novel contribution of the thesis. The framework considers the decision problem from the perspective of the policy maker and, although it is primarily intended for analysts (e.g. evidence review groups), it is also appropriate for appraisers (e.g. NICE committees). It aims to contribute to the evaluation of treatment sequences within a health technology assessment or similar process, such as comparative effectiveness research, technology appraisal, and evidence-based clinical guideline development. The framework is underpinned by my comprehensive review of previously-used methods for evaluating treatment sequences or estimating the treatment effects for an intervention conditional on the previous treatments administered, and my investigation of the added value of using complex evidence synthesis methods over more simplistic approaches.

The findings of the series of integrated literature reviews conducted within the thesis showed that there is no optimal way of evaluating treatment sequences. A range of different approaches have been used, each with advantages and disadvantages. The choice of approach is dependent on a number of issues, such as the extent and type of the sequences being evaluated and the available evidence base. The review of modelling studies revealed that, even when accounting for the limitations of the evidence, the best approach was not necessarily adopted by modellers. It is my intention that the framework and the simplifying assumptions work will aid future modellers to improve the modelling of treatment sequences and hence reduce the uncertainty and bias in the results generated.

The framework can be used to help choose the best approach, and consists of a series of recommendations developed in response to the key issues that emerged from the review.

8.2 DEFINITION AND SCOPE OF THE FRAMEWORK

There is very little guidance for conducting quantitative evidence synthesis of treatment sequences to inform clinical and policy decision making. Furthermore, there is nothing to help modellers recognise when treatment sequence issues are pertinent and should be included in a model. The evaluation of treatment sequencing is bound to be more complex than synthesising the evidence of the clinical effectiveness of single (or discrete) treatments. A framework for evaluating treatment sequences would be useful here. The Collins dictionary defines a framework as 'a particular set of rules, ideas, or beliefs which you use in order to deal with problems or to decide what to do'.⁴¹¹

The framework incorporates recommendations for key stakeholders (reviewers, statisticians, modellers, health economists, commissioners of health technology assessment or evidence reviews) and covers the entire health technology assessment process, from the initial scoping of the research question and articulating the decision problem, to the development of a decision-analytic model. It is based on the premise that there is no single best way of evaluating treatment sequences, and that a range of approaches and methods are currently in use, each with advantages and disadvantages. It also incorporates the key challenges identified for evaluating the evidence for treatment sequences. The framework has been developed to use alongside general best-practice guidance for undertaking systematic reviews, meta-analyses, and decision-analytic modelling. It is based on the guidance and guidelines development processes and methods used by NICE, which was used as an exemplar commissioner (of health technology assessment) and policy maker.

The decision problem relating to treatment sequencing can be complex and requires different types and levels of evidence to address it. Decision-analytic modelling can be used to synthesise different types of data. Health economic models often draw together evidence concerning the natural history of a disease, epidemiology, treatment effectiveness, health state utilities, adverse events, resource use and costs.³⁴⁶ A review of modelling approaches was undertaken to inform the use of decision modelling. However, it was not the intention of the review to give an in-depth evaluation of how to actually build and run a model, rather it focused on the overall approach for incorporating or evaluating treatment sequences within the model. The review addressed important questions such as, how was this done within current modelling studies? What were the simplifying assumptions that had to be made to achieve this? How does each approach compare with others? Of note here also is that the framework (and the thesis) focuses on the impact of treatment sequencing on effectiveness and not costs.

Treatment sequences are generally represented within a decision-analytic model as a series of discrete treatments, each requiring a summary effect estimate that is conditional on positioning in the treatment pathway. The scarcity of data to inform such estimates means that simplifying assumptions are frequently used in conjunction with the available data on discrete treatments. The framework also incorporates a coding scheme for simplifying assumptions used to inform treatment sequencing effects. This can be used to support the choice of modelling approach, but also to assess existing models.

8.3 THE FRAMEWORK TO INFORM THE EVALUATION OF TREATMENT SEQUENCES

In summary, the overarching framework has three components:

- i. Framework to guide practice presented as two related tables outlining:
 - a. Recommendations for practice (Figure 8.1 and Table 8.1)
 - b. Advantages and disadvantages of different modelling approaches that can be used to help guide modellers to choose which approach to use (Table 8.2)

- ii. A list of publications that can be used as a resource to support the use of best-practice guidance for undertaking meta-analysis and decision modelling in general (Figure 8.2)
- iii. A coding scheme for simplifying assumptions applied to treatment effects when modelling treatment sequences (Table 8.3)

8.3.1 Framework component 1: Framework to guide practice in the conduct of health technology assessment

The recommendations presented in Table 8.1 are currently ordered in line with the health technology assessment process and its reporting, which starts with developing the research question, then the clinical evaluation, followed by the economic evaluation. The recommendations are categorised according to:

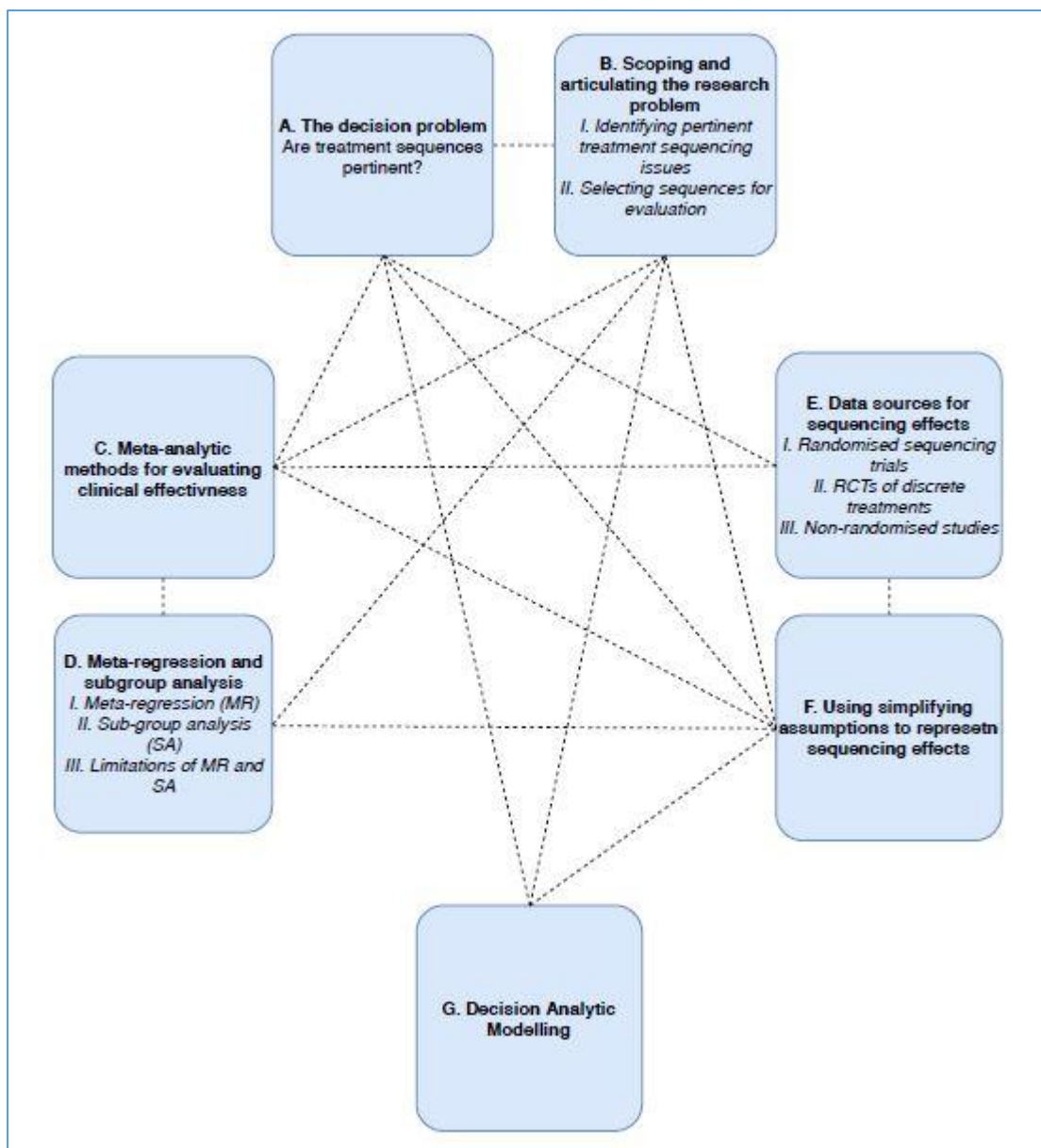
- i. How to assess the importance or need to consider treatment sequences, which corresponds to developing the research problem (Sections A and B)
- ii. Meta-analytic methods (Sections C and D)
- iii. Data sources used to inform sequencing effects (Section E)
- iv. Simplifying assumptions regarding sequencing effects (Section F)
- v. Decision-analytic modelling approaches (Section G)

These are illustrated in Figure 8.1.

Each recommendation is designed to be 'stand-alone', this allows re-arrangement or deletion for future development. Consequently, there is some repetition with a few recommendations being relevant under more than one header.

The relevant stakeholder(s) for each recommendation are listed in Table 8.1. The recommendations pertaining to decision-analytic modelling approaches are supplemented by Table 8.2, which summarises the advantages and disadvantages of different modelling approaches and the simplifying assumptions made for evaluating treatment sequences as identified in the review of modelling studies (Chapters 6 to 7).

Figure 8.1: A schematic of the categorisation scheme for practice recommendations



The dotted lines indicate where there are recommendations which relate to more than one category

Table 8.1: Framework for evaluating treatment sequences: recommendations for practice

	STATEMENT BASED ON REVIEW FINDINGS (RELATED CHAPTER)	DRAFT RECOMMENDATION FOR PRACTICE*	RELEVANT GROUP**
A	THE DECISION PROBLEM – ARE TREATMENT SEQUENCES PERTINENT?		
1A	Health technology assessments and clinical guidelines may need to consider treatment sequences.	<p>The potential importance of treatment sequencing needs to be considered prior or at the very start of the scoping process. This should include asking ‘Is there a reason why treatment sequencing issues are not relevant at all?’</p> <p>The need to consider treatment sequencing will likely depend on a range of factors including i) the underlying disease condition, ii) the care pathways, iii) the aim of the treatment, iv) type of treatment, v) range of existing treatments, and vi) the fixed term or continuous administration of treatments.</p>	Policy maker Reviewer Modeller

2A	The assessment of heterogeneity and inconsistency within pairwise MA and NMA of discrete treatments may provide important information on whether treatment sequencing effects needs to be considered. This would be enhanced in a MA based on individual patient data (see 9C)	Consider whether previous treatment is an effect modifier, and the impact of which needs to be accounted for. <i>Assessment of heterogeneity:</i> <ul style="list-style-type: none">• Elicit expert opinion on potential clinical heterogeneity.• Use consensus-based recommendations for investigating clinical heterogeneity in systematic reviews (Gagnier, 2013).• Use MR to explore whether 'previous treatments' is an important effect modifier. Where the findings are not statistically significant, consider whether this is due to insufficient power. (see also 2C)• If there are insufficient studies (reporting on previous treatment) for MR use subgroup analysis to explore the impact of previous treatments. This should include the use of appropriate statistical tests for interaction. (see also 2C)	Reviewer Policy maker
3A	When comparing the effectiveness of single treatments, subsequent treatments can have a confounding effect on the long-term outcome overall survival.	Consider whether the impact of subsequent treatment needs to be considered. (see also 3C)	
B SCOPING AND ARTICULATING THE RESEARCH QUESTION			
BI <i>Identifying pertinent treatment sequencing issues</i>			
1B	A number of factors can influence both the choice and the effects of each treatment used at different points in the pathway. These factors, which are both treatment- and time-dependent and independent, need to be accounted for when evaluating treatment sequences.	Elicit expert opinion on all potential factors that are likely to impact treatment sequencing effects, and whether each is likely to have a dependent and/or independent effect. (See also 5F) Consider using a conceptual framework/logic model to inform the process. Guidance is available on using conceptual framework and logic models for health technology assessment (Roberts, 2012; Rohwer, 2016; Squires, 2016). (See also 3G) The assessment of clinical and statistical heterogeneity in previous MA of single treatments is likely to provide useful information here. (See also 2A)	Policy maker Reviewer Modeller
2B	Treatment may be discontinued for various reasons, which will have a differential impact on the effectiveness of subsequent treatment.	Consider from the onset whether there is a need to differentiate between discontinuation due to lack/loss of effectiveness and adverse effects/intolerance, or non-adherence, which may be independent of both of these. However, implementation may be hampered by the reasons for switching treatment being poorly reported in the primary studies. (See 6C and 5F)	Policy maker Reviewer Modeller
3B	Disease duration could be considered as a surrogate outcome for previous treatments, but treatments may also become less effective with time due to disease progression (independent of treatment).	There is a need to ascertain whether disease duration and previous treatments are independent predictors of treatment response. Disease duration is a crude measure and therefore should only be used if no other relevant data is available. (See also 2B and 3-4D)	Policy maker Reviewer Modeller
BII <i>Selecting sequences for evaluation</i>			
4B	Selecting appropriate sequences for evaluation may be challenging. In some clinical areas, for example oncology, the introduction of new therapies means that the treatment sequences used in clinical practice change rapidly.	The choice of sequences for inclusion needs to be justified. The involvement of key stakeholders for selecting pre-defined sequences for inclusion in the analysis is essential. The sequences may be defined by the policy maker (commissioner). When conducting a NMA of complete sequences, the inclusion of further comparator sequences, outside the commissioning brief, may be necessary in order to develop a closed network.	Policy maker Reviewer Modeller

		Consider the pace at which the treatments available in practice are likely to change, and the usefulness of developing a model that can be updated easily. (See Section G)	
5B	Selecting a manageable number of treatment sequences for inclusion in the evaluation may not be easy. Alternative modelling approaches have been developed that aim to identify the optimal ordering of treatments, and are thus able to consider any conceivable sequence.	The <i>net benefit per unit time</i> modelling approach can be used for identifying the optimal ordering of treatments. However, it should not be used where treatment sequencing has an effect, as it requires the simplifying assumption of treatment independence to be made. <i>This approach is based on the premise that the earlier in the sequence a treatment is tried the greater the proportion of patients who receive it and respond to it will be.</i> (See Section F, and 3rd component of the framework: CODING SCHEME for simplifying assumptions, Table 8.3)	Policy maker Reviewer Modeller
6B	Identifying where the 'decision node' lies within the treatment pathway, and whether both upstream and downstream effects need to be considered is important.	A conceptual/logic model should be developed to identify the decision node, and whether upstream effects are important or influential, or only the downstream effects are relevant. The involvement of key stakeholders is also important. (See also 1B and 3G)	Policy maker Reviewer Modeller
C	META-ANALYTIC (MA) METHODS FOR EVALUATING CLINICAL EFFECTIVENESS		
	This section covers methods used for evaluating clinical-effectiveness and develop clinical effect estimates to parameterise the decision-analytic model (Section G) .		
1C	Prospective sequencing trials are few in number, and may not cover the breadth of decision making needed. They generally evaluate a limited number of treatment-lines.	A MA of sequencing studies is likely to be hampered a lack of available studies, which can cause problems with establishing a closed network of treatments, evaluating extensive treatment sequences, and considering important sub-populations. In order to be useful for clinical decision-making, the evidence review will need to consider a broad evidence base (and meta-analytic approaches) even when prospective treatment sequencing trials exist. (See also Section E1)	Reviewer
2C	MR and subgroup analysis are generally used to provide evidence on the optimal treatment for patients who have failed previous treatment (or assess the impact of previous treatment).	MR can be used for adjusting for previous treatment (and or disease duration), and subgroup or stratified analysis can be used to assess whether the treatment effect differs according to the specific previous treatments used or line of therapy. (See Sections DI and DII) MR and subgroup analysis are likely to be limited to the evaluation of the impact of immediate prior treatment, a specific series of prior treatments, or the number and type of previous treatments. MR and subgroup analysis will not be useful for evaluating the optimal initial treatment in a sequence or accounting for the effect of subsequent treatments on long term outcome measures.	Reviewer
3C	To inform decision making, methods are needed to assess the impact of subsequent treatments.	Methods that have been used to identify the optimal initial treatment or account for the impact of subsequent treatments on long term outcomes include: limiting inclusion to sequencing studies, partitioned survival analysis, or decision-analytic modelling. (See also 2A) Recent reviews of the latest methodological developments in mathematical modelling methods (Panayidou, 2016) and using partitioned survival (Woods, 2017) can be used to inform the implementation of these methods.	Reviewer Modeller

4C	To inform decision making, treatment effect estimates that are conditional on positioning in the treatment pathway are needed.	One approach for adjusting a summary treatment effect to make it conditional on positioning in the pathway is to apply a reduction or multiplication factor. (See also 5E and 5F) A specific reduction, or multiplication factor, can be developed and then applied to the treatment effects obtained from RCTs of first-line use to represent their use later in the treatment pathway. However, developing the multiplication or reduction factor will not be trivial or straightforward, as the evidence base is likely to be very limited.	Reviewer Modeller
5C	There is no established method for developing a reduction or multiplication factor, and the evidence base available for developing such estimates are likely to be limited.	The methods and data sources used to develop the reduction or multiplication factor should be clearly stated. Current methods include: i) the comparison of Hazard ratios of treatment withdrawal for a class of drugs used as first vs subsequent line (based on an observational study of registry data); and ii) comparing the findings from 2 RCTs of the same drug used at different points in the treatment pathway (1 st -line and last resort). The beta-coefficient taken from MR can potentially be used as a multiplying factor, but this approach has not yet been used in practice (See also 1D) . The time and resource implications of using this approach means that it will need to be planned for from the onset and be part of both the clinical and economic evaluation. (See Section B) The impact of using a reduction or multiplication factor needs to be explored in sensitivity analysis. (See Section F)	Reviewer Modeller
6C	The reason for discontinuing previous treatments can have a differential impact on the clinical effectiveness of subsequent treatments. However, reasons for switching treatment are often poorly reported by primary studies.	When adjusting a treatment effect conditional on positioning in the pathway, if feasible, consider the impact of discontinuing previous treatment due to inefficacy separately to that of switching due to intolerance or adverse effects. (See also 2B and 5F) A separate modifying factor should be developed and used for discontinuation due to lack of effect and intolerance. (See also 5E)	Reviewer Modeller
7C	One approach used for identifying the optimum sequence is to order treatments according to the absolute treatment effect estimates of single treatments.	This approach represents a naive method, and should not be used unless there is clear evidence that treatment effect is independent of positioning in the sequence.	Reviewer Modeller
D	META-REGRESSION AND SUBGROUP ANALYSIS		
	This section provided further detail on the use of meta-regression and subgroup analysis as part of the meta-analytic technique used (Section C)		
DI	Meta-regression (MR)		
1D	MR provides the most useful approach for developing summary estimates for discrete treatments that allow for previous therapies. Individual regression coefficients (from MR analysis) describe how the intervention effect changes with each unit increase of the covariate, which may be useful for evaluating treatment sequences.	MR can be used to develop 'adjusted' treatment effect estimates (plus an estimate of the uncertainty) according to the number and type of previous treatments used. (See also 4-5C)	Reviewer Modeller
2D	Poor reporting of previous treatment within primary studies is likely to contribute to non-statistical or false positive/negative findings due to	When selecting covariates for inclusion in the MR consider the clinical (not just statistical) significance of previous treatment (and disease duration).	Reviewer Modeller

	insufficient power. But previous treatment may still be an important covariate despite non-significant findings.		
3D	Disease duration may be better reported within primary studies than previous treatment, and could be used as a surrogate outcome for previous treatments.	Consider using disease duration as a surrogate for previous treatments, especially when evaluating whether to consider treatment sequencing by assessing heterogeneity. However, this is a crude measure and therefore should only be used if no other relevant data is available. (See also 2A and 3B) The likely scenario is that some data on both disease duration and previous treatments will be reported (inconsistently) by relevant studies. Both outcomes are correlated.	Reviewer Modeller
4D	Both previous treatment and disease duration are likely to be important treatment effect modifiers.	The MR will likely need to include the covariates representing both previous treatment and disease duration, which are correlated. A potential limitation is the limited number of covariates that can be considered in a MR. (See also 2A and 3B)	Reviewer Modeller
5D	The availability of RCTs reporting IPD, including full treatment histories, would greatly enhance the usefulness of MR (and subgroup analysis) as a method of developing sequence specific effect estimates.	Consider using an IPD-MA, which has much higher power if patient level covariates are of interest. However, access to IPD without adequate reporting of previous treatment is insufficient. The evidence synthesis is likely to be based on both aggregate and individual participant data. There is guidance on conducting IPD-MA recent review methods outlining the challenges (Debray, 2015; Tierney, 2015) .	Reviewer Modeller
6D	For MR to be useful in informing treatment sequencing decision making the findings need to be reported in full irrespective of statistical significance, especially when non-significance is likely to be due to lack of power.	The method used for selecting covariates in the MR needs to be reported in full. The findings or prior univariate analysis and non-significant MR analysis also need to be made available. (See also 5F)	Reviewer Modeller
DII	Sub-group analysis		
7D	Subgroup analyses are easier to implement than MR and can provide useful information on potential trends.	Subgroup (or stratified) analyses can be used to assess whether treatment effect varies according to positioning in the pathway (treatment-line) or population (treatment history), but need to be analysed using appropriate statistical tests for interaction.	Reviewer Modeller
DIII	Limitations of MR and subgroup analysis		
8D	The methods of MR and subgroup analysis have a number of generic limitations.	Potential limitations of using MR and subgroup analysis include: poor reporting of previous treatment; susceptibility to type I error (false negative) due to a small number of studies; and potential for ecological fallacy. They are also inherently observational in nature, and will have the same inherent limitations associated with non-randomised studies, including selection bias and confounding. Potential solutions include conducting an IPD-MA and including a broader evidence base (e.g. non-randomised studies; multiple outcome measures). The latest methodological developments in multi-parameter MA will aid the implementation of these solutions (Efthimiou, 2017; Debray, 2015; Cooper, 2015; Tierney, 2015) .	Reviewer Modeller
9D	The use of subgroup analysis has similar limitations to MR, as well as other drawbacks.	Further potential limitations of using subgroup analyses include:	Reviewer Modeller

		<p>i) they can only consider one covariate at a time, and each analysis will be confounded by other variables (e.g. disease duration);</p> <p>ii) they cannot provide an estimate of the extent of interaction;</p> <p>iii) it is difficult to interpret the findings when a series of subgroup analyses are used to assess multiple covariates, especially when low power produced insignificant findings;</p> <p>iv) the evaluation of numerous variables increases the risk of chance findings (type II error, false negative).</p>	
E	DATA SOURCES FOR SEQUENCING EFFECTS		
E1	Randomised sequencing trials		
1E	Treatment sequences can be conceived as a series of decision rules that specify how the treatment should change depending on the patient's needs (also referred to as dynamic or adaptive treatment regimens).	The sequential, multiple assignment, randomised trial (SMART) design allows for the testing of multiple potentially adaptive interventions along with tailoring variables that trigger change in (or switching) treatments (Almirall, 2014). SMART designs can provide good empirical evidence on the optimal treatment sequence, but are unlikely to cover all conceivable sequences and relevant patient subgroups. They are also designed to develop an adaptive (dynamic) treatment rather than confirm that the adaptive treatment is better than an alternative (control).	Reviewer Modeller
2E	Well conducted RCTs provide the most valid data for causal inferences. Pragmatic RCTs have been developed for evaluating both fixed sequences, dynamic/adaptive treatment regimens.	<p>Three types of RCTs have been used in used to inform treatment sequences:</p> <ul style="list-style-type: none"> • RCTs of pre-defined (fully formed) sequences; • RCTs of dynamic treatment sequences, e.g. SMART, which allow for the comparison of different treatment options within the context of what happens in later stages (patients randomised at each stage); and • Quasi-sequencing trials: RCTs of discrete treatments that incorporate subsequent treatment in the trial protocol. <p>Other RCT designs that are potentially useful for evaluating dynamic treatment sequences:</p> <ul style="list-style-type: none"> • N-of-1 trials; and • Adaptive trial designs <p>The benefit of randomisation is lost when making inferences about the causality of whole sequences based on RCTs of discrete treatments used at single points in the pathway. (See RCTs of discrete treatments)</p>	Policy maker Reviewer Modeller
3E	RCTs of predefined sequences can be synthesised using established meta-analytic techniques, but the synthesis of dynamic/adaptive treatments or N-of-1 studies may not be so straightforward.	<p>No methods were identified for synthesising multiple SMART trials or N-of-1 studies for evaluating treatment sequences.</p> <p>The latest methods and guides on synthesising complex interventions and meta-analysis to inform personalised medicine may provide useful information here (Pedder, 2016; Melendez-Torres, 2015; Punja, 2016a, Punja, 2016b, Tate, 2016). The methods used to analyse the findings of a single SMART, based on machine learning approaches, may also be useful for developing future MA methods.</p>	Reviewer Modeller
4E	Large RCTs take years to complete, and treatment sequences used in practice are likely to change continually with the ongoing introduction of new treatments and	There will never be sufficient sequencing RCTs that cover the breadth of clinical and policy decision making questions relating to treatment sequences. Alternative sources of evidence will need to be considered and meta-analytic methods for developing summary effect estimates that are conditional on positioning in the sequence.	Reviewer Modeller

	alterations to licence indication of exiting treatments.		
EII	Randomised controlled trials of discrete treatments		
5E	The available RCT evidence is often limited to discrete treatments used at a single point in the pathway (e.g. first-line), which are frequently used to inform the effects of the same treatment (or class) at other points in the sequence.	<p>Current methods used for adjusting the effect estimates taken from studies of discrete treatments used early in the sequence in order to represent their use later in the treatment pathway include: applying a modifying (reduction or multiplication) factor; decrementing the relative effect by a set amount at each treatment line; using expert opinion; using the summary effect from an alternative similar treatment evaluated at the relevant point in the pathway; and using the naive estimates of single treatments.</p> <p>No hierarchy can be established, but the use of a reduction or multiplication factor is considered the preferred option (See also 4-5C), and the use of a naïve approach should not be used unless there is evidence to show that treatment effect is independent of position in the sequence. (See Sections A and F)</p> <p>Sensitivity analysis should always be used with these methods.</p>	Reviewer Modeller
6E	Providing evidence to show that treatment effects are not dependent on positioning in the sequence or disease duration may not be straightforward, especially in the presence of a limited evidence base.	<p>Evidence that treatment effect is independent of position can be provided by: i) a MA showing homogeneity between studies of treatments used as different therapy lines (or positioning in the treatment pathway), ii) non-significant findings for covariates representing previous treatment (and disease duration) within a MR. However, both scenarios can be the result of low power.</p> <p>These approaches are not useful for treatments that have only been evaluated in RCTs as first-line (or at a single point in the treatment pathway).</p>	Reviewer Modeller
EIII	Non-randomised studies		
7E	Non-randomised studies may provide important information on sequencing effects. For example, registry data may be able to provide comparative effectiveness data for a specific line of treatment, which is adjusted for patient characteristics at the time of initiating the treatment, including the treatments received in previous lines.	<p>The evidence review (and MA) will likely need to include both randomised and non-randomised studies. The meta-analysis will need to incorporate methods that adjust for known biases and over-precision. However, not all biases and confounding factors will be known (or measured) and some, for example confounding by indication, may not be controlled for.</p> <p>Several approaches can be used to jointly synthesise data (and adjust for bias and over-precision) from randomised and non-randomised studies. The choice will likely be driven by data availability, time and resources, and technical expertise available in the research team (Efthimiou, 2017).</p> <p>An overview of different approaches used for bias adjustment is provided by NICE DSU TSD3 (Dias, 2012). A number of subsequent methods reviews and guidance for incorporating non-randomised data are also available, produced by DSU (Faria 2015; Bell, 2016), ISPOR (Berger, 2017), and others (Verde, 2015). The recent systematic review of the latest developments in NMAs conducted by the GetReal project also includes an overview of the latest methods for adjusting for study limitations and possible sources of bias (Efthimiou, 2016). Studies have also compared different methods for</p>	Reviewer Modeller

		<p>combining randomized and non-randomized evidence in NMA (Schmitz, 2013; Efthimiou, 2017).</p> <p>Whichever method is used to synthesise randomised and non-randomised studies, the quality of the evidence should be assessed, and sensitivity analysis should be used (Efthimiou, 2017).</p>	
8E	<p>There are potential biases that are specific to the evaluation of treatment sequencing effects.</p>	<p>Potential biases specific to the evaluation of treatment sequences include: selection (allocation) bias, channelling bias, regression to the mean, immortal time bias when restricting inclusion to participants who have received a specific late line of treatment; confounding by disease duration; enrichment with refractory patients or patients susceptible to adverse effects; class effect bias or confounding by the type of patients treated with a specific treatment class; aggregate data collection (treatments being reported at class rather than drug-level); and potential for missing or inaccurate data.</p> <p>Quality appraisal needs to encompass the extent to which primary studies have adjusted for these biases.</p>	<p>Reviewer Modeller</p>
F	<p>USING SIMPLIFYING ASSUMPTIONS TO REPRESENT SEQUENCING EFFECTS</p>		
	<p>A coding scheme for simplifying assumptions relating to treatment sequencing effects used in decision-analytic modelling is presented in the third component of the Framework (Table 8.3)</p>		
1F	<p>The scarcity of data to inform sequence-specific effect estimates within decision models means that simplifying assumptions are frequently used in conjunction with the available data on discrete treatments. No consistent or objective methods are generally used for selecting the simplifying assumptions used or the available evidence.</p>	<p>The choice of assumptions should be based on a clear rationale and evidence that they are appropriate. Where direct evidence is not available, this could include: i) citing expert literature/opinion; or ii) providing theoretical reasoning. Where relevant, appropriate methods for expert elicitation should be used.</p> <p>Commonly used assumptions (See the CODING SCHEME Component of the Framework) include:</p> <ul style="list-style-type: none"> • treatment effect is independent of positioning in the treatment sequence; • treatment effect is dependent on treatment line, but independent of the type of treatments used; • treatment effect is the same as a substitute treatment used at the same point in the sequence; • treatment effect is reduced, in line with a reduction factor, when used at a later point in the sequence; and • treatment effect decrements with each successive treatment. <p>The validity and reasonableness of the assumptions used in representing reality should be assessed using, for example real world data, other modelling studies, or expert consensus. The assumption of treatment independence should only be used if there is clear evidence to show that it is valid, which will need to be reported appropriately. (See Sections A and B)</p>	<p>Modeller</p>
2F	<p>The application of simplistic assumptions regarding sequencing effects will result in significant uncertainty around the effectiveness and cost-effectiveness estimates, the extent and impact of which should be explored as part of the analysis of structural uncertainty.</p>	<p>The evaluation of structural uncertainty is much less common than that of parameter uncertainty in health economics, but can have a greater impact on the results (Stevenson, 2014). The most common approach used is scenario analysis.</p> <p>The impact of the simplifying assumptions made regarding treatment sequences should be explored using a range of</p>	<p>Modeller</p>

		plausible (or extreme) scenarios, which need to be justified. Where these are based on clinical opinion, appropriate methods for expert elicitation should be used. Scenario analysis will not provide an indication of the most credible scenario. Other methods for handling structural uncertainty exist, but they are underdeveloped and research in this area is still ongoing (Stevenson, 2014; Briggs, 2012).	
3F	The data sources used alongside the simplifying assumptions vary, even when considering the same decision problem and addressing the same evidence gap.	The reasons for choosing certain data sources over other possible data needs to be justified. This should ideally be planned from the onset of the health technology assessment and be part of the clinical evaluation. (See Section B and E)	Reviewer Modeller
4F	The choice of data to inform the efficacy of subsequent treatments (beyond the decision point) are not consistent and not always based on a review of the evidence. Priority is often given to matching the evidence for the decision point rather than considering the treatment sequence as a whole.	Methods used to develop sequence specific treatment effect estimates to parameterise the model should be based on a comprehensive review of the evidence base and not selected single studies. A clear declaration of the data sources used and justification for their selection is required. (See Section E) The need to consider treatment sequences should be identified at the scoping stage of the review and considered as part of both the clinical and economic evaluation. (See Section A) Developing a conceptual model would help inform this process. (See 1B and 3G)	Policy maker Reviewer Modeller
5F	The assumption that a treatment becomes less effective when used later in the sequence, or with increased disease duration is sometimes employed. But methods used to inform the decrement or amount by which the treatment effects are reduced are often not reported or based on selected single studies.	The methods and evidence used to develop an estimate to downwardly adjust the efficacy of discrete treatments should be clearly reported (See also 4-5D and 5E). The evidence should take into account the disease and known pharmacology, e.g. whether non-response (requiring a move to the next treatment) is due to pharmacogenetic factor influencing absorption, distribution, metabolism, and excretion (ADEM), disease pathogenesis, or poor adherence. The methods and evidence used should ideally be based on a comprehensive review of the evidence base, which is likely to time consuming and need to be planned for from the onset of the technology assessment. (See also 1B) The impact of discontinuation of previous treatments due to lack of efficacy and toxicity should be considered separately. (See also 2B and 6C)	Policy maker Reviewer Modeller
G	DECISION-ANALYTIC MODEL		
	This section is linked to TABLE 8.2 , which outlines the advantages and disadvantages of current modelling approaches used for evaluating treatment sequences.		
1G	Appropriate recognition of treatment sequencing is crucial to many policy decisions, but developing a related economic model may not be considered and is not straightforward.	Prior to developing any economic model it is important to consider whether there is good reason why some treatment sequencing issues are NOT pertinent, i.e. is there scope to simplify to discrete treatments. (See Section A)	Policy maker Reviewer Modeller
2G	A wide range of decision-analytic modelling approaches have been used to evaluate treatment sequences. The most appropriate modelling approach should reflect the simplest model to answer the question, but it needs to adequately reflect the decision problem and its complexity in order to produce realistic results.	Table 8.2 provides a summary of the different modelling approaches used in existing studies (<i>and includes the indexing codes for the 'simplifying assumptions' that were applied by included studies using each approach</i>). Developing the simplest model required will depend on the complexity of the decision problem and the extent of the treatment sequences being modelled. (See Section B) Additional complexity (or key features) in the decision problem that the modelling approach may need to account for include: heterogeneity in the target population, patient history or previous treatments, the need for differential	Policy maker Reviewer Modeller

		treatment selection based on reasons for treatment discontinuation, the fact that not all patients go on to receive the subsequent treatment; outcomes of different subgroups, differential time on treatment; time dependency of certain parameters, repeated events, competing risks, parameter uncertainty, dynamic decision making.	
3G	The extent of the treatment sequences to be modelled and pertinent key features, both of which will impact the choice of modelling approach, need to be identified in advance.	Use a conceptual framework (or logic model) to identify the extent of the complexity of the decision problem. This process should involve all key stakeholders and follow best practice guidance. (See also 1B.)	Policy maker Reviewer Modeller
4G	The question of when a simple model would suffice is not straightforward, and the use of commonly used cohort approaches may not represent the use of a simple model structure.	<p>The question of when a simple model would suffice is dependent on the decision problem and the ease with which the modelling techniques is able to account for the treatment sequences and additional complexities (or key features) required. It is also dependent on the value of reducing uncertainty in the cost-effectiveness.</p> <p>Existing modelling studies provide examples of where complex modelling structures have been implemented using both commonly used cohort approaches and advanced methods based on individual patient simulation. Similarly, simple modelling structures have also been implemented using different approaches.</p> <p>Existing modelling studies also provide examples of where the decision problem has been simplified to allow a simple (cohort) model to be implemented, or used because the limited available evidence would render an extensive model unrealistic. The appropriateness of using this approach should be informed by an a-priori conceptual model and involve relevant stakeholders.</p>	Policy maker Modeller
5G	The feasibility of a modelling approach is dependent on the available data.	The trade-off between the details accounted for in the model structure and the available data needs to be explicit.	Policy maker Modeller
6G	Modelling treatment sequences may require a complicated or a sophisticated model, which will be time consuming and resource intensive to develop and implement.	Consider adapting an existing model or developing a model that can be further developed for future use (rather than developing for single use models).	Policy maker Modeller
7G	A range of modelling approaches have been used to evaluate treatment sequences and, although there is no clear evidence that any one type is superior to another, some may have advantages (or disadvantages) for considering different types of decision problems.	<p>Table 8.2 summarises the advantages and disadvantages of the different modelling approaches used in existing studies, the different ways treatment sequences have been conceptualised within the model, and the extent of the decision problem complexity that the different approaches were able to account for. No study has systematically tested different approaches for modelling treatment sequences.</p> <p>Cohort models can be simple, and easy to implement, but these advantages are lost when accommodating extensive sequences and additional complexities in the decision problem. Individual sampling models are more sophisticated, better able to accommodate greater decision problem complexity, and provide more flexibility. DES appeared the optimum approach in terms of intuitively modelling sequencing algorithms, computational efficacy, and ease of updating, but requires more extensive modelling skills, specialist software, and likely to be data and time intensive.</p>	Modeller

		A review of current guidelines for decision-analytic modelling is provided by Ramos, 2015	
8G	Markov cohort modelling is the most popular approach for evaluating treatment sequences, but can be limited by the Markovian property (assumption) that patients may only be in one state at a time and transition between states cannot depend on previous states visited, or time in a state.	<p>The Markovian property means that the model cannot account for patient's history, differential time on treatment, or the occurrence of multiple events within a single cycle, unless commonly used adaptations are implemented.</p> <p>The Markov cohort model can be successfully adapted for evaluating treatment sequences, using for example the Markov cycle tree, tunnel states, and 3-D matrix (semi-Markov model), but this is likely to be at the expense of having a simple model structure or accounting for the full extent of the decision problem. If the decision problem cannot be depicted in an appropriate way or requires an excessive number of health states, then a DES is preferable as it would enable a much simpler model structure to be implemented. However, DES models are expensive in terms of data and long running times, which limit the practicability of the model.</p>	Modeller
9G	The limited evidence base often precluded modelling extensive treatment sequences and lengthy time horizon, but better evidence may become available at a later date.	The trade-off between the details accounted for in the model structure and the available data needs to be explicit. Consider using a modelling approach (model technique and structure) that will optimise the feasibility of updating as and when new evidence appears.	Policy maker Reviewer Modeller
10G	Economic modelling undertaken to inform reimbursement decisions are generally required to include probabilistic sensitivity analysis (PSA) to account for the uncertainty in the decision problem.	A cohort-based model may be chosen as it is considered easier to implement PSA within this approach. However, uncertainty in the decision regarding which treatment sequence to choose should incorporate the distribution of the treatment sequencing effects.	Modeller

* The details of references highlighted in bold are presented separately Figure 8.2.

** This relates to the stakeholder group for whom the recommendations are likely to be relevant to. The term 'reviewers' refers to researchers involved in conducting the clinical evaluation (and includes systematic reviewers and statisticians, as well as health economists conducting a meta-analysis to inform the economic model). 'Modellers' refers to those involved in developing and implementing a decision-analytic model, and 'policy makers' refers to those who commission or interpret the evidence reviews in order to make policy decisions.

Abbreviations: DES discrete event simulation; IPD individual patient (level) data; MA meta-analysis; MR meta-regression; NICE DSU NICE Decision Support Unit; ISPOR International Society For Pharmacoeconomics and Outcomes Research; NMA network-meta-analysis; PSA probabilistic sensitivity analysis; RCT randomised controlled trial; S section; SMART sequential, multiple assignment, randomised trial.

Table 8.2: Advantages and disadvantages of current modelling approaches used for evaluating treatment sequences

How treatment sequences are conceptualised in the model	Sequencing decision problems evaluated* Maximum number of treatment lines (time horizon range)	Simplifying assumptions made**	Further complexities in the decision problem that were allowed for in the model	Advantages of modelling approach	Disadvantages of modelling approach
COHORT-BASED MODELS					
Decision tree (DT)					
Treatment sequences, possible outcomes, and events depicted as decision tree.	Adding treatment Optimum sequence Predefined sequences Treatment approach - 4 (1 year - lifetime)	IP, NPT, PGE, ST, RF, TD, DI, UOBS, EXC, RDD, LR-	<ul style="list-style-type: none"> Relapse treated with a previous successful treatment. Duration of response differs according to levels of response. Reason for discontinuation impacts selection of subsequent treatments. Some treatments administered for a fixed period only. Toxic death and all-cause mortality have different probabilities and timing. Not all patients receive all treatments in sequence (<i>implemented in conjunction with portioned survival</i>). 	<ul style="list-style-type: none"> Can be relatively straightforward to develop and not computationally intensive. Can be easy to interpret and transparent. Can include a large number of different treatment sequences within the same model. Can be used in conjunction with other methods. 	<ul style="list-style-type: none"> No explicit time component; governed by fixed timing of outcomes and events. Only allows one way progression. Cannot handle looping/recurring events easily. Can become exponentially complex with additional events and disease states. Poorly suited for complex scenarios.
Markov cohort					
i) a series of treatment or line-specific Markov states; ii) above with additional temporary states representing e.g. adverse effects, relapse; iii) Markov cycle tree, with Markov states used to represent different levels of disease activity or natural history.	Adding treatment Different points Optimum sequence Predefined sequences Single point Treatment approach - 7 (1 year - lifetime)	IP, NPT, PGE, ST, RF, TD, DI, UOBS, EXC, LR.	<ul style="list-style-type: none"> Not all patients receive all treatments in sequence. Duration of response differs according to levels of response and treatment line. Probability of continuing treatment and developing toxicity varies with time and for each treatment (<i>using tunnel states</i>). Reason for discontinuation impacts selection of subsequent treatments. Some treatments administered for fixed period. <i>Cycle trees used to account for:</i> <ul style="list-style-type: none"> Consequence of adverse effects. Different levels of treatment response Some patients continue treatment despite not achieving full/clinical response. Fluctuating disease activity. Complex treatment pathways. 	<ul style="list-style-type: none"> Can be relatively straightforward to construct and communicate. Has a time component; events can occur at any time. Allows looping/recurring events. Transitions can be unidirectional or bidirectional. Can be used in conjunction with decision tree (<i>Cycle tree</i>). The use of cloned subtrees enables ease of update. 	<ul style="list-style-type: none"> Markov assumption (<i>memoryless</i>): prohibits TPs being dependent on time spent in the state, or previous states visited (<i>can be overcome using additional states and stratification</i>). Patients can only be in one state at a time. Transitions limited to fixed intervals defined by cycle length. Cannot account for multiple events within one cycle (<i>can be overcome using short cycles or cycle trees</i>). Occurrence of events assumed to be constant over time (Markov chain). Exponential increase in complexity with increasing number of states.
Semi-Markov cohort					
A series of treatment specific health states. Multiple-dimension transition matrix used to allow time spent in each state to depend on current and next state.	Single point - 3 (15 years)	IP.	<ul style="list-style-type: none"> Probability of treatment failure decreases with time on a specific drug. 	<ul style="list-style-type: none"> <i>Same as Markov cohort</i> Reduced impact of Markovian assumption (not memoryless; incorporates time dependency). 	<ul style="list-style-type: none"> Patients can only be in one state at a time. Transition limited to fixed time intervals defined by cycle length. Transitions can only occur at fixed intervals. Only one transition allowed per cycle. Becomes more complex with added states.
Partitioned survival					
A series of treatment specific health states. Time spent in each state calculated from the area	Predefined sequences Adding treatment Different points	NONE, IP, DI, RF.	<ul style="list-style-type: none"> Decreasing probability of remaining on a given treatment with time. 	<ul style="list-style-type: none"> Can be relatively straightforward to develop and not computationally intensive. Non-data intensive. 	<ul style="list-style-type: none"> Cannot account for complex treatment sequencing algorithms or additional

under the curve of survival functions.	- 6 (3 years - lifetime)			<ul style="list-style-type: none"> • Transparent. • Area under the curve can be calculated continuously over time; no cycles required. • Can be used in conjunction with DT. 	attributes (e.g. adverse effects, disease duration).
INDIVIDUAL SAMPLING MODELS					
State transition (STM)					
Fixed treatment sequences; disease activity monitored for each individual over time. Health states usually represent response or non-response to each treatment, with the addition of adverse effects as a separate state in some models.	Adding treatment Treatment approach Predefined sequence Single point - 7 (10 years - lifetime)	IP, NPT, PGE, RF, DI, UOBS.	<ul style="list-style-type: none"> • Duration of response differs according to levels of response. • Fluctuating disease activity (changes in disease activity assumed to relate to level of response, not treatment). • Complex treatment algorithms. • Patients follow different disease courses, which cannot be predicted at the onset. 	<ul style="list-style-type: none"> • Not limited by Markov assumption (eliminating need for excessive number of states). • A large number of characteristics can be ascribe to individually simulated patients. • Access to individual patient data enabled key parameters and events in patient histories to be calculated using multivariate regression, allowing adjusting for important covariates. • Can account for heterogeneous population. 	<ul style="list-style-type: none"> • Transition limited to fixed time intervals defined by cycle length. • Cannot account for multiple events in one cycle. • Can be computationally intensive.
Discrete event situation (DES)					
i) fixed treatment sequences; ii) random selection of pre-defined sequences; or iii) developed as part of the modelling process by selecting individual drugs, using a random process, at specific points in the sequence	Adding treatment Different points Predefined sequences Single point - 13 (5 years - lifetime)	IP, NPT, PGE, RF, DI, UOBS, RDD.	<p><i>When conceived as a simple structure:</i></p> <ul style="list-style-type: none"> • Variable time to quitting treatment. • Duration of response differs according to levels of response. • Fluctuating disease activity. • Reason for discontinuation impacts selection of subsequent treatments. <p><i>When developed as a more complex structure:</i></p> <ul style="list-style-type: none"> • Treatment selection and cessation based on algorithms reflecting specific clinical guidelines • Unpredictable nature of disease progression. • Multiple treatment outcomes. • Not all patients go on to receive subsequent treatments in the sequence. • Differential treatment selection for subgroups. 	<ul style="list-style-type: none"> • Can ascribe a large number of characterises to individually simulated patients. • Can account for heterogeneous population. • Not limited by the use of fixed time advancement (cycles). • Patients can simultaneously be in multiple states, and experience different events. • Allows for modelling of complex scenarios and treatment algorithms. • Computationally more efficient than STM. • Can be easily adapted to incorporate additional events or patient attributes. 	<ul style="list-style-type: none"> • Model structures can be difficult to communicate and interpret. • Computationally challenging in terms of model design and running it.
(OPEN) POPULATION-BASED MODELS:					
Non-terminating population based simulation (DES)					
Pre-specified clinical thresholds used to invoke escalation to next treatment (dynamic equations used to project clinical measures over time).	Single point	IP	<ul style="list-style-type: none"> • Dynamic disease process. 	<ul style="list-style-type: none"> • <i>Same as DES</i> 	<ul style="list-style-type: none"> • <i>Same as DES</i>
Markov multi-cohort model					
Markov cycle tree; Markov states represented individual treatments and 'switching'.		NA	<ul style="list-style-type: none"> • Impact of adding a new drug on health care budget assessed using prevalence approach (target population kept constant over time - entry of newly diagnosed cohort at each cycle). 	<ul style="list-style-type: none"> • <i>Same as Markov Cohort</i> 	<ul style="list-style-type: none"> • <i>Same as Markov Cohort</i>

***Different treatment sequencing decision problems considered by included studies:**

Adding treatment: adding a new drug to a pre-defined sequence;

Comparing sequences: comparing predefined sequences;

Different points: evaluating the optimum positioning of a treatment within a sequence;

Optimum sequence: identifying the optimum sequence from all conceivable sequences;

Single point: comparing different treatments used at the same point within a sequence;

Treatment approach: comparing different management approaches, e.g. 'step-up' or 'step-down' treatment approaches.

****Simplifying assumptions made regarding treatment effects (see Table 8.3):**

IP Treatment effect is independent of positioning in the sequence;

NPT Treatment effect is dependent on the number of previous treatments (treatment line), but independent of the type of treatments used;

PGE Treatment effect is the same as an alternative treatment (*substitute from same class*) used at the same point in the sequence;

GE Treatment effect is the same as an alternative treatment (*substitute from same class*) irrespective of positioning in the sequence;

ST Treatment effect is the same as an alternative treatment (*substitute from different class*) used at the same point in the sequence;

RF Treatment effect is reduced, in line a reduction factor, when used at a later point in the sequence;

TD Treatment effect decrements with each successive treatment;

DI Treatment effect does not change when the treatment is displaced by adding a new prior treatment (displacement ignored);

Abbreviations: .NA not applicable

8.3.2 Framework component 2: List of references to related methodological development

The thesis focused on methods used for evaluating treatment sequences. However, the framework is intended to be used alongside best-practice guidance standards for a health technology, incorporating systematic reviews, meta-analysis, and decision-analytic modelling. The framework also refers to the use of recent methodological advances in evidence synthesis and modelling methods to aid assessment of treatment sequences in a number of recommendations. These were informed by several recent relevant methodological reviews. Some key papers summarising recent methodological developments are listed below as a supplementary resource for the framework. These have been derived, primarily, from three important information sources, which can also be used to support the framework and the implementation of a health technology assessment of treatment sequences:

- i) The series of Technical Support Documents produced by NICE Decision Support Unit (DSU). These documents provide detailed guidance on appropriate methodology for specific issues in health technology assessment and economic evaluation. The DSU also developed the NICE methods guide update and conducted other relevant methods work. (<https://scharr.dept.shef.ac.uk/nicedsu/>).
- ii) Methodological reviews and cases studies conducted within the Get Real project. (<http://www.imi-getreal.eu/>)
- iii) A series of consensus guideline reports on good practice standards for outcomes research (clinical, economic, and patient-reported outcomes) and on the use of this research in healthcare decision making developed by the ISPOR-SMDM Good Research Practices Task Force. These are produced collaboratively by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and The Society for Medical Decision Making (SMDM). (<https://www.ispor.org/taskForces/TFIndex.asp>)

The reference list below is not intended to include all guidance documents relating to evidence synthesis or decision-analytic modelling developed by the NICE DSU centre, the GetReal Project, or ISPOR Good Practice Task force. Rather it lists the references referred to in Table 8.1 to indicate the type of resource that may be used and further developed as part of the framework. Comprehensive lists of publications relevant to network meta-analysis methods have been developed and maintained, by a team of researchers from the Universities of Leicester and Bristol (<http://www.bristol.ac.uk/social-community-medicine/projects/mpes/courses/treatmentcomparisons/>), and Georgia Salanti (<http://www.mtm.uoi.gr/index.php/tutorial>). The DSU also provides Technical Support Documents on evidence synthesis (<http://www.nicedsu.org.uk/Evidence-Synthesis-TSDseries%282391675%29.htm>).

Figure 8.2: List of references relating to recent methodological development referred to in the framework

- Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA (2014). Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Transl Behav Med*; 4(3): 260–274.
- Bell H, Wailoo AJ, Hernandez M, Grieve R, Faria R, Gibson L, Grimm S (2016). The use of real world data for the estimation of treatment effects in NICE decision making. [Available from <http://www.nicedsu.org.uk>]
- Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, Madigan D, Makady A, Schneeweiss S, Tarricone R, Wang SV, Watkins J, Daniel Mullins C (2017). Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Value in Health*; 20(8): 1003-1008
- Briggs AH, Weinstein MC, Fenwick E, et al (2012). Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM modeling good research practices task force-6. *Value Health*; 15:835-42.
- Cooper NJ, Sutton AJ, Achana F, Welton NJ (2015). RFP Topic: Use of Network Meta-analysis to Inform Clinical Parameters in Economic Evaluations. [Available from: http://www.nihrcrsu.org/guidance/network_meta-analysis/]
- Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, et al. (2015). Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research Synthesis Methods* 6(4): 293-309.
- Dias S, Sutton AJ, Welton NJ, Ades AE (2011). NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment; last updated April 2012 [Available from: <http://www.nicedsu.org.uk>]
- Efthimiou O, Mavridis D, Debray TP, Samara M, Belger M, Siontis GC, Leucht S, Salanti G; GetReal Work Package 4 (2017). Combining randomized and non-randomized evidence in network meta-analysis. *Stat Med*; 15;36(8):1210-1226.
- Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, Reitsma JB, Shang A, Salanti G; GetReal Methods Review Group (2016). GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods*;7(3):236-63.
- Faria R, Hernandez Alava M, Manca A, Wailoo AJ (2015). NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. [Available from <http://www.nicedsu.org.uk>]
- Gagnier JJ, Morgenstern H, Altman DG, Berlin J, Chang S, McCulloch P, Sun X, Moher D; Ann Arbor Clinical Heterogeneity Consensus Group (2013). Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. *BMC Med Res Methodol*;13:106.
- Melendez-Torres GJ, Chris Bonell C, Thomas J. Emergent approaches to the meta-analysis of multiple heterogeneous complex interventions. *BMC Med Res Methodol*. 2015; 15: 47.

Panayidou K, Gsteiger S, Egger M, Kilcher G, Carreras M, Efthimiou O, Debray TPA, Trelle S, Hummel N; GetReal methods review group (2016). GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world. *Res Synth Methods*; 7(3): 264–277.

Pedder H, Sarri G, Keeney E, Nunes V, Dias S (2016). Data extraction for complex meta-analysis (DECiMAL) guide. *Systematic Reviews* Dec 13;5: 212.

Punja S, Xu D, Schmid CH, Hartling L, Urichuk L, Nikles CJ, Vohra S (2016a). N of 1 trials can be aggregated to generate group mean treatment effects: A systematic review and meta-analysis. *J Clin Epidemiol* 76:65–75.

Punja S, Schmid CH, Hartling L, Urichuk L, Nikles CJ, Vohra S (2016b). To metaanalyze or not to meta-analyze? A combined meta-analysis of N of 1 trial data with RCT data on amphetamines and methylphenidate for pediatric ADHD. *J Clin Epidemiol* 76:76–81.

Ramos MC, Barton P, Jowett S, Sutton AJ (2015). A Systematic Review of Research Guidelines in Decision-Analytic Modeling. *Value Health*;18(4):512-29.

Roberts M, Russell LB, Paltiel AD, et al. Conceptualizing a model: A report of the ISPOR-SMDM modeling good research practices task force-2. *Value Health* 2012;15:804-11.

Rohwer, A., Booth, A., Bfadenhauer, L., Brereton, L., Gerhardus, A., Mozygema, K., Oortwijn, W., Tummers, M., Van der Wilt, G.J., Rehfuss, E (2016). Guidance on the use of logic models in health technology assessments of complex interventions. [Available from: <http://www.integrate-hta.eu/downloads/>]

Schmitz S, Adams R, Walsh C (2013). Incorporating data from various trial designs into a mixed treatment comparison model. *Stat Med*; 32(17):2935-49.

Squires H, Chilcott J, Akehurst R, Burr J, Kelly MP (2016). A Framework for Developing the Structure of Public Health Economic Models. *Value Health*;19(5):588-601.

Stevenson M, Tappenden P, Squires H (2014). Methods for handling uncertainty within pharmaceutical funding decisions. *International Journal of Systems Science*;45(1): 60-68.

Tate RL, Perdices M, Rosenkoetter U, Shadish W, Vohra S, Barlow DH, Horner R, Kazdin A, Kratochwill T, McDonald S, Sampson M, Shamseer L, Togher L, Albin R, Backman C, Douglas J, Evans JJ, Gast D, Manolov R, Mitchell G, Nickels L, Nikles J, Ownsworth T, Rose M, Schmid CH and Wilson B (2016). The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Statement. *J Clin Epidemiol* 73:142–152.

Tierney JF, Vale C, Riley R, Smith CT, Stewart L, et al. (2015). Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLoS Medicine* 12(7): e1001855.

Verde PE, Ohmann C (2015). Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Res Synth Methods*; 6(1):45-62.

Woods B, Sideris E, Palmer S, Latimer N, Soares M (2017). NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review. [Available from <http://www.nicedsu.org.uk>]

8.3.3 Framework component 3: Coding scheme for simplifying assumptions used to represent treatment sequencing effects

Simplifying assumptions relating to treatment sequencing effects used by existing modelling studies, along with their indexing codes, and grouped into six broad types, are listed in Table 8.3. The list of assumptions can be used to aid the consideration of treatment sequences within a decision problem, to inform the choice of approach to use within an economic model by clarifying what has actually been done previously, or to highlight whether modellers have used the same or different approaches within health technology assessments of a similar decision problem. Different assumptions are likely to be applied at different positions in the treatment pathway. The coding scheme (as indexing codes) is used in Table 8.2 to indicate the type of simplifying assumptions applied by included studies using each modelling approach.

Table 8.3: Coding scheme for simplifying assumptions relating to treatment sequencing effects

SIMPLIFYING ASSUMPTIONS	CODE
<i>Treatment independence</i>	
Treatment effect is independent of positioning in treatment sequence.	IP
Treatment effect is dependent on the number of previous treatments used (or line of therapy) but independent of the specific previous treatment used.	NPT
<i>Substitution with another treatment effect</i>	
Treatment effect is the same as an alternative treatment from the same class, or a generic class effect - matching the same position in the sequence.	PGE
Treatment effect is the same as an alternative treatment from the same class, or a generic class effect - irrespective of positioning in the sequence.	GE
Treatment effect is the same as a substitute treatment taken from an alternative but related class of treatments, matching the same position in the sequence.	ST
<i>Reduction of treatment effect</i>	
Treatment effect is reduced in line with a specific multiplier or reduction factor, when used at a later point in the sequence. (<i>Reduction or multiplication factor is informed by the available evidence that is also relevant to the treatment of interest.</i>)	RF
Treatment effect decrements by the same pre-set amount (proportion) at each point in the sequence, or with each successive treatment. (<i>The proportion is not necessarily based on a specific evidence base.</i>)	TD
Treatments become less effective with increased disease duration.	RDD
<i>Impact of time since previous treatment</i>	
Treatment effect is not affected by previous treatments if patients have been in long term remission, and thus can re-use the same treatment(s)/class of treatment(s) as that which achieved the prior remission.	LR
<i>Displacement effect ignored</i>	
The effect of a treatment displaced further down the sequence by the introduction of a new treatment is unaffected.	DI
<i>The use of uncontrolled/observational studies without bias adjustment</i>	
Non-randomised studies provide an un-biased estimate of treatment sequencing effects.	UOBS
Expert consensus provides an un-biased estimate of treatment sequencing effects.	EXC

8.4 FUTURE DEVELOPMENT OF THE FRAMEWORK

The framework presented here provides evidence-based guidance for conducting or using quantitative evidence synthesis of treatment sequences. It was informed by an extensive search and

in-depth evaluation of the international literature covering multiple disease and treatment types. The next stage is to incorporate the views of key stakeholders, validate the recommendations, and test its use in practice. Future development could include qualitative interviews with experts in the field to corroborate and build on the list of recommendations. A modified Delphi consensus process could also be used to assess agreement with the subsequent list and identify any recommendations that may not be important. A similar process to that used to develop the ISPOR good practice guidance could be used. The thesis used NICE as the exemplar policy maker, and feedback from NICE on the usefulness of the framework and its potential adaptation would enhance future implementation. Researchers working on NICE technology appraisals and clinical guidelines could also be recruited to test the use of the framework in practice.

The next stage of developing the framework needs to consider how best to incorporate public involvement in the evaluation of treatment sequences. The views of both patients and clinicians will be important in developing the scope of the review and informing the extent and type of treatment sequences that need to be evaluated. Public involvement will also be important in designing and conducting the review, as well as writing and disseminating the findings. The future development and refinement of the framework needs to involve patients as advisors.

8.5 CHAPTER SUMMARY

This chapter presents a framework for conducting quantitative evidence synthesis of the effectiveness of sequential treatment options within the context of the evidence review for informing clinical and policy decision making. The framework now requires further input and refinement by experts in the field (key stakeholders, clinicians, and patients), to include selecting the most relevant recommendations, to be tested for usefulness in informing practice.

9. DISCUSSION AND CONCLUSIONS

9.1 CHAPTER OVERVIEW

This chapter provides a discussion of the research presented in the thesis and outlines the main conclusion of the work. It starts with a summary of the key findings, and their importance. The thesis chapters represent a series of integrated literature reviews. Each chapter contains a summary of how the findings of each review fits into the context of existing research. This is not presented again in detail in this discussion. Rather, this chapter outlines the strengths and limitations of the overall research and summarises the implications of the findings for policy, practice, and research.

The thesis identified a wide range of quantitative evidence synthesis methods used for evaluating treatment sequences. However, all were hampered by the limitations of the primary studies, and would benefit from access to individual patient data. An important outcome of the thesis is the recommendations for future research corresponding to the identified gaps in the research evidence. These are outlined here.

9.2 SUMMARY OF THE MAIN FINDINGS

The aim of the thesis was to develop a framework for conducting quantitative evidence synthesis methods to estimate the effectiveness of treatment sequences within a health technology assessment or similar process. This includes methods for developing summary estimates of clinical effectiveness or the clinical inputs to the cost-effectiveness assessment. The framework (Chapter 8) was developed through an in depth evaluation of current approaches. It consists of a series of recommendations developed in response to the key issues that emerged from the review. It also includes a summary table to support modellers to choose an appropriate modelling approach, a bibliographic resource to support the use of best-practice in quantitative evidence synthesis, and a coding scheme of simplifying assumptions applied to treatment effects when modelling treatment sequences.

The thesis focuses on the evaluation of treatment sequences to support clinical and policy decision making. Where multiple treatments are available, the use of network meta-analysis is required to inform decision making, which also enables the inclusion of a broader evidence base. A health technology assessment of the clinical and cost-effectiveness of treatments for sciatica demonstrates my experience gained in conducting a network meta-analysis (Chapter 2).

The National Institute for Health and Care Excellence (NICE), which provides funding recommendations for England and Wales, was used as an exemplar policy maker. A review of NICE guidance (Chapter 4) demonstrated that appropriate recognition of treatment sequencing is crucial to many policy decisions, but not always considered during the scoping stage of the evidence review or the clinical evaluation.

A comprehensive review of current quantitative evidence synthesis methods (Chapter 3) was conducted that considered:

- i. Meta-analytic techniques used to develop summary effect estimates of treatment sequences, or effect estimates that are conditional on previous treatment used (Chapter 5)
- ii. The range of simplifying assumptions made by decision analytic modelling studies in the absence of an adequate evidence base on treatment sequencing effects (Chapter 6)
- iii. The actual decision analytic approaches used for modelling treatment sequences (Chapter 7)

The findings demonstrated the following:

- i. Reviewing the evidence on treatment sequencing is neither trivial nor straightforward.
- ii. There is no single best way to evaluate treatment sequences, rather there is a range of approaches that have been used.
- iii. Each approach has advantages and disadvantages and is influenced by both the evidence available and decision problem.
- iv. Previous treatment is an important effect modifier, and subsequent treatments can confound long term outcomes such as survival.
- v. The reason for discontinuing treatment has a differential effect on the effectiveness of subsequent treatment.
- vi. Prospective sequencing trials are few in numbers and do not cover the breadth of decision making needed.
- vii. The extent and type of sequences being evaluated tended to reflect the available research evidence rather than clinical practice.

The evidence used to inform treatment sequencing was broadly considered in two ways, representing a one-step-at-a-time evaluation based on series of single treatments or the comparison of whole sequences. No novel meta-analytic methods were developed for evaluating treatment sequences, and none were directly aimed at developing a conditional summary estimate of effect.

The current meta-analytic approaches, which can be used in a clinical evaluation (Chapter 5) include:

- i. Network meta-analysis of whole sequences (Section 5.3). This approach is hampered by the limited number of available sequencing trials, which made it difficult to establish a closed network. There is likely to be a benefit from extending the evidence base to include non-randomised studies. However, the analysis would need to incorporate methods that adjust for known biases. A number of potential biases were identified that are specific to the evaluation of treatment sequences (Section 5.9).
- ii. Stratified meta-analysis of single treatments according to the line of therapy (Section 5.4). This approach does not account for the specific prior treatments used or consider the impact of whole sequences in any depth.

- iii. Network meta-analysis of all single treatments irrespective of where they are used in the pathway (the meta-analysis of sciatica treatments in Chapter 2 provides an example of this). An important limitation of this approach is that previous treatment can have both an impact on treatment effect, acting as an effect modifier resulting in heterogeneity, and be associated with the type of treatment comparison, acting as a confounding factor. For example, in the sciatica review non-invasive treatments were more likely to be used as initial treatments, and invasive treatments used after the failure of other treatments. However, the investigation of clinical heterogeneity within this approach can be used to assess whether treatment sequences need to be considered. This is dependent though on previous treatment, or positioning in the pathway, being clearly reported in the primary studies. Disease duration could potentially be used as proxy measure for the number of previous treatments used.
- iv. Meta-regression, or the combined use of network meta-analysis and meta-regression, to adjust for the previous treatment used (Section 5.5). This approach was not generally used for the sole purpose of evaluating treatment sequences, rather it was used to account for the heterogeneity within the meta-analysis. The covariate representing previous treatment was often dropped from the final analysis due to non-significant findings, but this may have been due to lack of power as previous treatment was often poorly reported in the primary studies. The initial covariates selected for evaluation sometimes included both previous treatment and disease duration, which are correlated, with the effect of one likely to be confounded by the other. There may be justification for the inclusion of both covariates in the meta-regression analysis. Meta-regression can be used to estimate the effect of previous treatment whilst adjusting for the effect of disease duration.
- v. Network meta-analysis based on multivariate analysis of both first and second-line treatments, as opposed to stratified analysis for first and second-line (Section 5.8). The biggest challenge here was developing an estimate of the correlation between the first and second-line treatment in order to conduct the analysis. Real world data from patient registries can potentially be used to provide this estimate. This approach was not developed for evaluating treatment sequences as such, but rather to evaluate the methods for incorporating real-world data in the evidence synthesis of second-line treatment. This included the issue of how to connect disconnected networks. The network meta-analysis did not directly compare first versus second-line treatment but provided estimates for all treatments in first and second-line, by using the correlation between them to predict estimates in second-line (or first-line) where these estimates did not exist previously.
- vi. The development of a multiplication factor, which can be applied to the summary effect of a treatment used as first-line in order to represent its use at a later point in the pathway (Section 5.6). The optimal approach for developing a multiplication factor is yet to be established.

Treatment sequences were often represented within the economic model as a series of single treatments, each requiring a summary treatment effect estimate conditional on positioning in the treatment pathway (Chapter 6). The findings of the clinical evaluation within a health technology

assessment were generally used to inform the effectiveness of interventions used at a single decision point in the economic evaluation. The scarcity of data on sequencing effects meant that simplifying assumptions were often applied to the available data on discrete treatments. A novel coding scheme was developed based on all the simplifying assumptions made (Chapter 6), which can also be used to assess future models. The most common assumptions were that treatment effect is independent of positioning in the sequence, or that treatment effect is dependent on the number of previous treatments (treatment line), but independent of the type of treatments used. These assumptions were frequently not validated, nor their impact on the overall results assessed. The data sources used alongside the simplifying assumptions for treatments used beyond the decision point varied, even when considering the same decision problem and addressing the same evidence gap.

A decision analytic model structure is generally kept as simple as possible in order to aid implementation, transparency, and understanding by decision makers.^{370 392} The advantages and disadvantages of different modelling approaches were reviewed whilst considering the structural complexity required by the clinical scenario, and the extent and type of treatment sequences being modelled (Chapter 7). The range of treatment sequencing decision problems investigated included:

- i. Identifying the optimum sequence
- ii. Adding a new drug to an established sequence
- iii. Comparing 'step-up' or 'step-down' treatment approaches
- iv. Comparing different treatments used at the same point within a sequence
- v. Evaluating a drug used at different points within a sequence
- vi. Comparing predefined sequences

Examples of some of the additional attributes that were accounted for in the model structure included:

- i. Different treatment selection for patient subgroups
- ii. Reason for discontinuation impacts subsequent treatment
- iii. Some treatments administered for fixed period, others until failure
- iv. Early recurrence treated with a prior treatment
- v. Not all patients receive all treatments in the sequence
- vi. Variable time to quitting treatment
- vii. Duration of response differs with type, and treatment line
- viii. Fluctuating disease activity
- ix. Unpredictable nature of disease progression

The absence of data was sometimes used as a justification for simplifying important issues.

A wide range of modelling techniques were identified, which fall under three main headers:

- i. Cohort-based models (deterministic and stochastic decision trees, Markov, semi-Markov, partitioned survival)
- ii. Individual sampling models (state transition and discrete event simulation)
- iii. Open population-based models (discrete event simulation and Markov cohort)

No study systematically tested modelling approaches for treatment sequences. The most popular approach was a Markov cohort. Cohort models have the advantage of being simple and easy to implement. Examples of cohort models that were successfully adapted to accommodate additional complexity in the decision problem were identified, but these were no longer based on a simple structure, which impacts transparency and implementation. Individual sampling models are more sophisticated, better able to accommodate greater decision problem complexity, and provide more flexibility. However, they are likely to be more resource-intensive and less transparent. The discrete event simulation appeared the optimum approach in terms of intuitively modelling sequencing algorithms, computational efficacy, and ease of updating, but requires more extensive modelling skills, specialist software, and is data and time intensive.

9.3 STRENGTHS AND LIMITATIONS OF THE RESEARCH

This is the first review of methods to investigate the evaluation of treatment sequencing across all clinical scenarios, and to include both meta-analytic techniques and decision analytic modelling. It represents an extensive in-depth review of current methods used to evaluate the clinical effectiveness of treatment sequences, representing a broad and disparate area of research.

Most of the included studies investigated methods for evaluating treatments for inflammatory arthritis or advanced cancer. It is unclear whether the clustering of included studies around these two chronic conditions was due to either of the following:

- i. The 'information scent searching' approach used, which also involved scanning the reference list of existing reviews that tended to be limited to these conditions
- ii. The fact that these are the only conditions for which treatment sequencing has been explored in any depth

I suspect that the latter is true, and that this has been influenced by the fact that NICE have highlighted the importance and challenge of investigating treatment sequencing for both inflammatory arthritis and advanced cancers. The review also identified other chronic conditions, such as depression, epilepsy, nerve pain, and viral infections e.g. immunodeficiency virus (HIV) and hepatitis, for which treatment sequencing is an important issue for informing clinical practice and decision making.

The review of methods, and the related searches focused on studies that specifically aimed to evaluate treatment sequences, the vast majority of which were economic modelling studies. It also focused on treatment sequencing effects, and not costs. However, some of the included modelling studies reported that treatment sequences were modelled in order to reflect clinical practice and capture the downstream cost of subsequent treatment; some aimed to evaluate the cost-effectiveness of single treatments. A potential limitation of the review is that the searches, in particular those of the reference databases, may have missed some relevant studies that aimed to evaluate single treatments, but developed a model of treatment sequences in order to account for the downstream

cost of subsequent treatment. This is because their published title, abstract and indexing terms are unlikely to have included sequencing-related terms. In fact, this may have been one of the reasons why so many relevant modelling studies were identified via the hand search (Chapter 3, Figure 3.1), rather than the reference databases. Another potential contributing factor to this is that the reference database searches, completed in 2013, were not updated, whilst hand searching continued throughout the review process. New literature was also picked up during the writing process, and I am therefore fairly confident that no new methods were missed. The search strategy did not cover NICE single technology appraisals, as discussed in Chapter 4 (Section 4.5). However, it is unlikely that any health technology assessment or economic evaluation that did not specifically aim to compare treatment sequences used a novel method or one that is not covered by the included studies. The findings of the review of methods demonstrated that modelling studies focusing on the evaluation of single treatments tended to either ignore potential sequencing effects or only consider them in a simplistic way. This is corroborated by the findings of a published review of NICE appraisals evaluating treatment sequences, which did not identify any new methods not already included here.¹⁰⁴ This was discussed in more detail in Chapter 4, (Section 4.5) and Chapter 6 (Section 6.7.2). I come back to this in Section 9.4.3.

The methodology review was based on objective and transparent methods. However, it was conducted by a single reviewer, and the data extraction was not checked for accuracy against the published papers. The review findings, and their translation into recommendations for practice in the framework, was subject to peer review by the multidisciplinary PhD supervisor panel.

9.4 RECOMMENDATIONS FOR PRACTICE AND FUTURE RESEARCH

9.4.1 Recommendations for policy and practice regarding health technology assessment of treatment sequences

The recommendations for practice, in terms of implementing quantitative evidence synthesis methods for evaluating treatment sequences are incorporated in the framework presented in Chapter 8, which also includes recommendations for future development of the framework (Section 8.4).

Health technology assessment is time and resource intensive, and there is increased pressure on both researchers and policy makers to expedite the process. The single technology appraisal, based on industry submission, is often used to this end, especially for making decisions on the use of new treatments (Chapter 4, Section 4.2). However, it is important that the clinical context within which these treatments will be used in practice is considered, including the availability of multiple treatments, with all treatments for chronic conditions being part of a treatment sequence. The evaluation of treatment sequences is likely to be more appropriate within a multiple technology appraisal and may also require the modelling of disease pathways. However, it may be unrealistic to consider extensive treatment sequences for the assessment of all new treatments or clinical conditions where multiple treatments are available. More research is needed to establish when it is

necessary to evaluate treatment sequences, and how best to make this decision. This is likely to be a condition-specific endeavour, but the methods will be relevant across different clinical scenarios.

It is important that any proposed solutions, and the guidance provided in the framework for evaluating treatment sequences, are both practical and feasible within the proposed timeframe of a health technology assessment. However, it is also important that this does not result in the omission of important methods required for evaluating treatment sequences that are likely to be burdensome. For example, in a systematic review of multiple complex, multi-dimensional interventions, which are context-dependent, essential methods for identifying, accounting for, and communicating the added complexity in the interventions will be an added burden.^{402 404} Examples of such methods include the development of logic or conceptual models involving the entire research team and key stakeholders¹⁸³
^{400 402 404} (Chapter 7, Sections 7.4.6).

The development of a complex decision model supported by an appropriate evidence base will be resource intensive. The importance of treatment sequences is increasingly recognised for many clinical and policy decisions, for example in the introduction of new targeted therapies for metastatic cancer (Appendix Volume I, Section C4).⁴¹² Therefore, there may be a need to consider moving away from a health technology process that is reliant on developing single use decision models.⁴¹³ However, this may require the commissioned model to be passed to another research team at a later date for further adaptations to address another technology appraisal. It is unclear how the single technology appraisal process, where the model is developed by industry, would fit in here (Chapter 7, Sections 7.4.7). However, some manufacturer evidence submissions to NICE include a model developed by an academic group.

The simplistic assumptions regarding sequencing effects made by modelling studies are likely to result in significant uncertainty around the effectiveness and cost-effectiveness estimates, the impact of which is generally unknown (Chapter 7). This needs to be recognised in decision making, and further evaluated. As discussed in Section 9.3, economic evaluations of single treatments that also account for the downstream cost of subsequent treatments, which represent current practice, are likely to model treatment sequences in an overly simplified way. Furthermore, the most common simplifying assumption made by included modelling studies was that the efficacy of individual treatments were independent of positioning in the sequence. In other words, treatment sequencing effects are ignored. The use of the coding scheme for simplifying assumptions developed as part of the thesis (Chapter 6, Section 6.5.1) can help make this more explicit. This also highlights the need to establish the importance of accounting for treatment-sequencing effects, which is discussed further in the next section (9.4.2). Recommendations for practice relating to the simplifying assumptions made regarding treatment-sequencing effects within decision analytic modelling are presented in Chapter 6 (Section 6.7.4). I also come back to this in Section 9.4.3.

There was little reference made within the existing research on the potential, or actual role, of incorporating patient perspectives into the evolution of treatment sequences. Further work is needed to develop the optimal approach for involving members of the public in health technology assessment of treatment sequences, especially beyond that of helping to define the research question during the scoping stage (Chapter 8, Section 8.4).

9.4.2 Assessing the need to account for treatment sequencing effects and implications for future research

There is a need for more research to assess the importance of accounting for treatment sequences, and the full impact of not doing so. However, the limited evidence base makes this very difficult. A closely related issue is the need to test the validity of making the simplifying assumption of treatment independence within a decision model, which I come back to in Section 9.4.3. The comparison of existing RCTs of treatment sequences with RCTs of the matching single treatments may provide some useful information, but the limited available sequencing studies will be problematic.

Mathematical simulation methods may aid the exploration of the presence and size of a sequencing effect, as they could be used to develop the evidence base for a potential 'ideal' treatment sequence for comparison. Datasets such as Archimedes or patient registries could similarly be used to simulate a treatment sequencing RCT.^{95 356} The extent of the interaction between individual treatments, when used as part of a sequence, could also be explored using a model-based network meta-analysis (discussed in Chapter 5, Section 5.5.3), which is generally used for modelling treatments used at different dosages.²⁰⁶ The interaction between individual treatments could be explored using RCTs comparing the simultaneous administration of combination therapy with sequential administration of the same treatments.^{206 414} However, an important consideration here is that treatment sequences, unlike combination therapy, are not fixed, with the duration of each treatment and the choice of subsequent treatments depending on various factors. This was discussed in Chapter 1 (Sections 1.2 and 1.5.1) and Chapter 3 (Section 3.4.3). The interaction between individual treatments can also be explored using a component-based network meta-analysis, which is increasingly being used for synthesising complex interventions^{61 336} (discussed in Chapter 5, Section 5.9.10.3). This approach is used to assess whether the individual components of the complex intervention interact or not. However, this approach does not account for the time course of treatment sequences, or the fact that the choice of the subsequent treatment is dependent on the impact of the previous one.

9.4.3 Recommendations for practice and research for decision modelling and the use of simplifying assumptions

Decision modelling studies generally considered a limited number of predefined sequences, usually selected based on clinical guidelines or expert opinion. The extent of the treatment sequences being investigated was sometimes chosen based on the available data. Further work is needed to ascertain how best to identify relevant sequences and the number of treatment lines to be considered in the model.

Due to the scarcity of data, many modelling studies applied simplifying assumptions to the available data on discrete treatments used at a single point in the pathway. A coding scheme for all possible assumptions was developed, providing a unique resource. The use of simplifying assumptions and the presence of important gaps in the evidence base may mean that some sequencing models will not be correct, but they may still be more accurate than not incorporating sequences within the model. More research is needed to test this.

The application of simplifying assumptions to the available evidence in order to represent treatment sequencing effects within a decision model results in structural uncertainty, which are generally explored using scenario analysis³⁴⁷ (Chapter 6, Section 6.2.2). Further research is needed to develop methods for selecting the most credible scenario for evaluating treatment sequences. Further research is also needed to identify the most applicable method for assessing the robustness and validity of the simplifying assumptions used, especially when the relevant data are not available. The reasons for selecting certain data sources in preference to others need to be justified. This should ideally be planned from the outset of the health technology assessment and be part of the clinical evaluation. There is also a need for more research to determine how best to test the validity of making the simplifying assumption of treatment independence (introduced in Section 9.4.2). This is especially true when the available evidence is limited to a single RCT for each treatment or RCTs of first-line treatments as, for example, in the modelling studies of new anti-epileptic drugs.^{18 276} The issue and methods of assessing the validity of simplifying assumptions made regarding treatment sequences is discussed further in Chapter 6 (Section 6.7.3), and the external validity of a treatment sequencing model in Chapter 7 (Section 7.4.4).

The coding scheme for simplifying assumptions has the potential to be an important tool for clarifying the extent to which treatment sequencing effects have been accounted for within a decision model. Further research is needed to test whether this scheme is comprehensive. The coding scheme was developed with the intention for it to be applied to a model of any clinical condition. More research is needed to identify any additional simplifying assumptions that are only applicable to specific clinical scenarios. One approach to further develop the coding scheme and test its comprehensiveness would be to apply the coding scheme to all NICE single technology appraisals conducted in the last decade. This would also provide useful information on the extent to which industry submissions have accounted for treatment sequencing effects, and whether this has changed over time or differs across clinical conditions.

9.4.4 Recommendations for practice and future research for meta-analytic methods

There have been some great advances in multi-parameter evidence synthesis methods in recent years, especially in terms of incorporating multiple treatments.^{59 164} However, although these methods are continually evolving, none have been adapted to develop sequence-specific effect estimates. Network meta-analysis has evolved in response to the need for decision making to account for the multiple treatments that are available in practice. There appears to be slow but increasing recognition,

based on the findings of the review of modelling studies, that decision making also needs to account for the use of these treatments used as part of a sequence. However, the advances in evidence synthesis methods have yet to reflect this, with more focus currently being placed on identifying the reasons for heterogeneity within the evaluation of discrete treatments. Further research work is needed to identify how best to develop a summary treatment effect of an intervention that is conditional on the previous treatment being ineffective or partially effective, or the evaluation of whole sequences. The lack of advances in meta-analytic methods for evaluating treatment sequences is also likely to be a reflection of the fact that the primary research is still mainly focused on the evaluation of single treatments. This is discussed further in Section 9.4.5.

The review findings demonstrated that, in most cases, treatment sequences represent complex, multifaceted, dynamic interventions, which will require advanced methods of quantitative evidence synthesis, especially if evaluated using a 'one-step-at-a-time' approach. It may not initially be feasible to account for all the complexity involved in either future meta-analytic techniques or decision analytic modelling, even if a perfect evidence base was available. For example, conditional effect estimates may need to account for the impact of both the immediate prior and other previous treatments, the reason for discontinuation including non-adherence, time on treatment, the duration of benefit or loss of effectiveness over time, evolving disease, and patient-specific prognostic factors that may impact clinical effectiveness. It may be possible to develop a novel method, based on the adaptation of current methods, which accounts for some of this complexity in the first instance. The methods that could potentially be further developed are discussed in Chapters 2 (Section 2.8) and 5 (Section 5.9.10). It is unlikely that a single method would suffice, and a range of solutions could evolve as better data, computing ability, and researcher skills and experience emerge over time. My research provides an important first stage in developing the methods to evaluate treatment sequencing. An ongoing review of methods will likely be required to maintain the framework to guide the implementation of these methods. In the meantime, there is also a need to establish which elements are the most impactful and thus need to be accounted for in the present. This is likely to vary depending on the clinical condition and type of decision problem or research question being considered. The conceptual or logic model discussed in Chapter 7, Section 7.4.6 will provide a useful tool here, which can be used to 'think through' the multiple components of the complex intervention in context, enhance the transparency of underlying assumptions, and assist in communication, both within the review team and with a range of stakeholders including patients and public.⁴⁰⁴

An important limitation for performing a meta-analysis of whole sequences was the lack of available trials. Two types of RCT designs were identified for evaluating treatment sequences, depending on whether they aimed to compare pre-defined (static) or adaptive treatment sequences (Appendix Volume I, Section B). This included the sequential multiple assignment randomised trial (SMART) design, which is a relatively new and innovative trial design created to inform the development and optimisation of time-varying adaptive or dynamic treatment regimens (Appendix Volume I, Section B2).⁴¹⁵⁻⁴¹⁷ I come back to the SMART design in Section 9.4.5. The availability of multiple sequencing

studies for evaluating each specific sequence of interest within a health technology assessment is unlikely. A more realistic scenario would be a need to conduct a synthesis of prospective sequencing trials, including both SMART and RCTs of predefined sequences (Appendix Volume I, Section B1), and non-randomised studies of specific sequences. In other words, the evidence synthesis would need to consider the inclusion of diverse study designs. Data from RCTs of single treatments may also be required. Furthermore, the analysis of the data from a single time-varying SMART design is challenging and not straightforward.^{16 418} A synthesis of the data from multiple SMART studies will therefore be even more complicated; although no such meta-analyses were identified.

Further research is needed:

- i. To identify the best way to conduct a meta-analysis of multiple SMART trials
- ii. To decide how best to meta-analyse data from both SMART and RCTs of predefined sequences (Appendix Volume I, Section B), and whether it is feasible to also incorporate studies of single treatments within the same analysis
- iii. To evaluate how recent adaptations of network meta-analyses methods (e.g. multivariate meta-analysis to incorporate multiple outcome measures or follow-up intervals, or bias-adjustment methods to incorporate observational studies) can be used to improve the scope of the available evidence base for analysing treatment sequences. This is likely to differ according to the clinical scenario
- iv. To identify how best to incorporate data from RCTs and non-randomised studies in the same network meta-analysis of treatment sequences
- v. To identify how best to adjust for the biases that are important for the evaluation of treatment sequences
- vi. To ascertain what methods should be used to adjust for the potential biases relevant to using non-randomised studies to inform treatment sequences
- vii. To ascertain whether disease duration and previous treatments are independent predictors of treatment response for clinical scenarios where treatment sequences is considered important, e.g. the treatment of rheumatoid arthritis
- viii. To identify the best method of estimating and testing modifying factors. Any proposed methods will need to take into account the fact that the evidence available for developing them is likely to be limited

9.4.5 Recommendations for future practice and research regarding primary studies

The review findings demonstrated that the main challenge for evaluating treatment sequences was the limitations and poor reporting of primary studies.

9.4.5.1 Recommendations relating the design of primary research

There is a need for more good quality primary studies of treatments sequences. However, this will require the development and use of innovative, efficient, and adaptive study designs. The SMART design provides a good example of this but does not represent a definitive trial. In many instances a

large number of treatment sequences are likely to be feasible for evaluation, which can soon become outdated as new treatments become available. The SMART design provides a robust approach for developing and selecting treatment sequences for evaluation in a definitive trial.

There is also ongoing research on developing adaptive trial designs to make RCTs more efficient,¹⁰⁷⁴¹⁹ which could also potentially inform the evaluation of treatment sequences. A description of the adaptive trial designs is provided in Appendix Volume I (Section B). Common adaptations made during the implementation of adaptive designs, based on interim analysis include:¹⁰⁷⁴¹⁹

- i. Adding or dropping treatment arms
- ii. Changes to the required sample size to ensure sufficient power
- iii. Changes to the allocation ratio to ensure more patients receive the superior treatment
- iv. Refinement of the existing study population according to their predictive biomarkers (enrichment)
- v. Transition directly from one trial phase to another

There may be scope for applying some of these methods to expedite the SMART design, for example the capability of adding new treatments or dropping inefficient treatment arms. The adaptive trial designs can also potentially be further developed for evaluating treatment sequences, for example by making adaptations based on the interim analysis to identify whether participants are responding to an initial treatment. The adaptive trial design has become especially appealing in the development and evaluation of biological-directed therapies, also referred to as personalised medicine⁴²⁰ (Chapter 1, Section 1.7). They are able to simultaneously evaluate both multiple treatments and biomarkers in heterogeneous patient populations.¹⁰⁷⁴²¹ They could potentially be further adapted to evaluate multiple treatments in a patient population that is heterogeneous due to previous treatments, rather than due to the heterogenous nature of tumours of the same site and stage. Further research is needed to explore these issues in more detail. The N-of-1 trial is also used to inform personalised medicine.¹¹⁰ Further research is also needed to evaluate the potential of using the N-of-1 trial design for evaluating treatment sequences.

The study design and the available evidence base for evaluating new treatments are often driven by the requirements of the regulatory authorities for licencing purposes rather than the need to develop evidence to facilitate the selection of the most effective treatment³¹⁴²² (Chapter 1, Section 1.3.2). The evidence requirements and optimal study design for regulatory approval does not always translate into suitable data for reimbursement decisions. The findings of my research demonstrated that primary research tends to focus on the evaluation of single treatments at a defined point in the pathway (Chapter 6, Section 6.6). The lack of data on the effectiveness of these treatments when used at another point in the pathway is a barrier to making policy decisions about the optimal positioning of new treatments or treatment sequences. The focus of primary research on single treatment is unlikely to change unless the regulatory authorities specify the importance of treatment sequencing or optimal positioning of new treatments. Technology appraisal plays a significant role in opening or closing market access for many new treatments, and the relevancy of the of the data

submitted by industry to health technology assessment agencies is consistently increasing.^{32 423} The reimbursement agencies and health technology assessment bodies are also becoming increasingly well placed to make recommendations on the nature of the clinical evidence required to inform treatment sequences. This in turn is relates to the recommendations for research in Section 9.4.2.

9.4.5.2 Recommendations relating the reporting of primary research

There is a need for better reporting on previous and subsequent treatments by primary studies of single treatments. An increased recognition of the importance of treatment sequencing for many policy decisions may provide more impetus for the demand for improved reporting. This needs to include:

- i. The reporting of specific previous treatments and their duration, as well as the reasons for treatment discontinuation and the timing
- ii. The reporting of specific subsequent treatments in studies that evaluate long-term outcomes or survival associated with single treatments

The availability of individual patient level data from primary studies, including both randomised and non-randomised studies, would greatly enhance both the development and implementation of future methods. This was discussed in Chapter 1 (Section 1.4.1) and Chapter 5 (Section 5.5.3). There is already some impetus for improved access to this type of individual level data for quantitative evidence synthesis methods in general.^{74 75} However, better reporting of previous (and subsequent) treatments is still needed to make this type of data useful for informing treatment sequences.

9.4.6 Recommendations for future practice and research regarding patient registries

Real world patient registries of high quality and validity are likely to provide an important source of data on treatment sequencing effects, although methods are still required to adjust these effects in order to account for any potential bias.^{45 59 333 424} The use of linked databases, or 'big data', are likely to further enhance their use.^{43 44}

Patient registries provided an important data source for treatment sequencing effects used in economic modelling studies included in the review of methods (Chapter 6-7). However, important gaps and limitations for informing treatment sequencing effects still existed in terms of matching the treatments used in practice, with previous treatments often reported as a class rather than individual treatments (Chapter 7). This was true even when the modelling studies had access to individual patient level data.^{225 248 251 273 274} Patient registries are often set up to monitor the potential long-term effects and adverse effects of new treatments, such as biological agents for rheumatoid arthritis⁴²⁵ (Appendix Volume I, Section C3.3). It would be beneficial to consider the type of data that is required for evaluating treatment sequences during the development and planning stages of these registries. An increased recognition of the importance of treatment sequences would likely improve this. It is also important that the data from National patient registries (such as those listed in the Appendix Volume I, Section C3.3) are made available to researchers, including individual-patient level data.

More research is also needed to inform the optimum methods for evaluating registry, or big data, to inform treatment sequencing effects. The methods developed to evaluate SMART trials, such as Q-learning (Appendix Volume I, Section B2),^{16 418} may be useful here. The thesis focused on the use of secondary analysis and did not consider methods for evaluating primary data. The continual improvement of the processing power of computers, ongoing developments in statistics, artificial intelligence and machine learning methods, and related research initiatives such as Archimedes (Chapter 1, Section 1.4.2 and Chapter 7, Section 7.4.8) may be useful here. The methods used for developing computational prediction models of cancer patients' response to therapies based on the analysis of multiple types of genome-wide molecular data in order to support personalised medicine⁴²⁶ (Chapter 1, Section 1.7) may also be useful. A future review of these methods may be required to inform practice.

Further work may also be needed to ascertain whether a better summary estimate of the clinical effectiveness to inform the economic model may be obtained from the evaluation of treatment sequences based on a single linked 'big' data source rather than a meta-analysis of multiple studies (Section 9.4.4). The specific limitations and biases inherent within observational studies for evaluating treatment sequences, including missing and inaccurate data are listed in Chapter 5 (Section 5.9.9). The generic limitations associated with this type of data are also discussed in the Appendix Volume I (Section A).

An alternative approach to obtaining a summary estimate of the clinical effectiveness when no relevant RCTs of treatment sequences exist is to use the observational data to emulate an RCT (Chapter 6, Section 6.7.3).³⁵⁶ Target trial emulation involves the application of design principles from randomised trials to the analysis of data from large observational studies (big data) in order to make causal inferences.⁴²⁷ The targeted trial design emulated using observational data is typically a pragmatic RCT, as it is not usually possible to emulate blinding.^{327 427} If the emulation is successful the analysis of the observational data would yield the same effect estimate, except for random variability, as the target trial if it had been conducted.³²⁷ However, it is acknowledged that it is not possible to emulate the ideal trial and a number of compromises will have to be made, for example there may be a need to choose alternative inclusion criteria due to the type of data capture, the intervention may not be sufficiently defined, or it is not possible to measure enough baseline confounding to emulate random assignment.^{327 427} Hernan and Robins outline a framework for comparative effectiveness research using big data that makes the target trial explicit, provides a structured process for the criticism of observational studies, and helps avoid common methodologic pitfalls.³²⁷ The authors also acknowledge the need for the observational databases to have passed through many high quality validation studies.³²⁷ Hernan and Robins also argue that using an explicit target trial approach has the advantages of improving the quality of the big data, such as patient registries, and can be used to articulate a compelling rationale to modify data structuring and recording practices.³²⁷ They also note that in order to maximise the benefits of big data for making causal

inferences, this explicit target trial approach needs to be used in combination with subject matter expertise, epidemiological and methodological proficiency, and innovative computer science tools.³²⁷

9.5 CONCLUSIONS

The use of appropriate quantitative evidence synthesis of treatment sequencing is essential for informing policy and clinical decision making. The thesis provides a state of the art overview of current practices in conducting quantitative evidence synthesis of treatment sequences. It included a series of integrated literature reviews and meta-analyses that contributed to the development of a novel framework that provides guidance for commissioners, producers, and users of health technology assessment (or similar process), for the evaluation of treatment sequences to inform policy and clinical decision making. The findings of the integrated literature reviews provide important information on when and how to account for treatment sequences, describe the main challenges of doing so, and identify key gaps in the evidence base. The thesis also provides important ground work for developing future meta-analytic or decision analytic methods for evaluating treatment sequences. Further research work involving the wider community of stakeholders is now required to further develop the framework, and make recommendations that are condition specific. Further work is also needed to ascertain the importance of considering treatment sequencing within the evaluation of the evidence and decision making.

REFERENCES

1. Higgins JPT, Green S, (editors). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org*, 2011.
2. Dixon-Woods M, Bonas S, Booth A, Jones DR, Miller T, Shaw RL, et al. How can systematic reviews incorporate qualitative research? A critical perspective *Qualitative Research* 2006;6(1):27-44.
3. Weinstein MC, Fineberg HV, Elstein AS, Frazier HS, Neuhauser D, Neutra RR, et al. *Clinical Decision Analysis*. Philadelphia: WB Saunders, 1980.
4. Petitti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine* New York, NY: Oxford University Press 1994.
5. Ioannidis JPA, Patsopoulos NA, Rothstein HR. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413
6. Kamerow D. How can we treat multiple chronic conditions? *BMJ* 2012;344:e1487.
7. Lewis R, Williams N, Matar HE, Din N, Fitzsimmons D, Phillips C, et al. The effectiveness and cost effectiveness of management strategies for sciatica: systematic review and economic model. *Health Technology Assessment* 2011;15(39):1-578.
8. Scottish Intercollegiate Guidelines Network (SIGN). Diagnosis and management of psoriasis and psoriatic arthritis in adults. 2010(SIGN publication no. 121). [cited 12 Oct 2010]. Available from URL: <http://www.sign.ac.uk>.
9. Murphy SA, Oslin DW, Rush AJ, Zhu J, for MCATS5. Methodological Challenges in Constructing Effective Treatment Sequences for Chronic Psychiatric Disorders. *Neuropsychopharmacology* 2007;32:257–62.
10. Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine* 2000;19(8):1011-28.
11. Fries JF, Williams CA, Singh G, Ramey DR. Response to therapy in rheumatoid arthritis is influenced by immediately prior therapy. *J Rheumatol* 1997;24(5):838-44.
12. Rendas-Baum R, Wallenstein GV, Koncz T, Kosinski M, Yang M, Bradley J, et al. Evaluating the efficacy of sequential biologic therapies for rheumatoid arthritis patients with an inadequate response to tumor necrosis factor- α inhibitors. <http://arthritis-research.com/content/13/1/R25> 2011;13:R25.
13. Kalden JR, Schulze-Koops H. Immunogenicity and loss of response to TNF inhibitors: implications for rheumatoid arthritis treatment *Nature Reviews Rheumatology* 2017;13(12):707-18.
14. Albiges L, Choueiri T, Escudier B, Galsky M, George D, Hofmann F, et al. A systematic review of sequencing and combinations of systemic therapy in metastatic renal cancer. *European Urology* 2015;67(1):100-10.
15. Kidwell KM. SMART Designs in Cancer Research: Past, Present and Future. *Clinical Trials* 2014;11(4):445–56.
16. Chakraborty B. Dynamic treatment regimes for managing chronic health conditions: a statistical perspective. *American Journal of Public Health* 2011;101(1):40-5.
17. Rodgers M, Epstein D, Bojke L, Yang H, Craig D, Fonseca T, et al. Etanercept, infliximab and adalimumab for the treatment of psoriatic arthritis: a systematic review and economic evaluation. *Health Technol Assess*. 2011;15(10):1-329.
18. National Institute for Health and Care Excellence. *Epilepsy: NICE clinical guideline 137*. London: NICE, 2012.
19. Eichelberg C, Vervenne WL, De Santis M, Fischer von Weikersthal L, Goebell PJ, Lerchenmüller C, et al. SWITCH: A Randomised, Sequential, Open-label Study to Evaluate the Efficacy and Safety of Sorafenib-sunitinib Versus Sunitinib-sorafenib in the Treatment of Metastatic Renal Cell Cancer. *European Urology* 2015;68(5):837-47.

20. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott DA. The effectiveness of anti-TNF- α therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. *Rheumatology* 2010;49:2313-21.
21. Anderson JJ, Wells G, Verhoeven AC, Felson DT. Factors predicting response to treatment in rheumatoid arthritis: the importance of disease duration. *Arthritis & Rheumatology* 2000;43(1):22-9.
22. Dimopoulos G, Siempos II, Korbila IP, Manta KG, Falagas M. Comparison of first-line with second-line antibiotics for acute exacerbations of chronic bronchitis: a metaanalysis of randomized controlled trials. *Chest* 2007;132(2):447-55.
23. Jamnitski A, Bartelds GM, Nurmohamed MT, van Schouwenburg PA, van Schaardenburg D, Stapel SO, et al. The presence or absence of antibodies to infliximab or adalimumab determines the outcome of switching to etanercept. *Ann Rheum Dis* 2011;70(2):284-8.
24. Oronsky B, Carter CA, Reid TR, Scicinski J, Oronsky A, Lybeck M, et al. Confirmatory Trials in the Evaluation of Anticancer Medicinal Products in Man--PFS2: A Measure of Therapeutic Action-At-A-Distance. *Neoplasia* 2015;17(9):716-22.
25. Murphy SA, Lynch KG, Oslin D, McKay JR, TenHave T. Developing Adaptive Treatment Strategies In Substance Abuse Research. *Drug and alcohol dependence* 2007;88(Suppl 2):S24-S30.
26. Almirall S, Collins LM, Murphy S. Introduction to Adaptive Interventions and SMART Designs. *Getting SMART About Developing Individualized Sequences of Health Interventions*. Hollywood, Florida: CPDD, 2011.
27. Hawkins N, Sculpher M, Epstein D. Cost-effectiveness analysis of treatments for chronic disease: using R to incorporate time dependency of treatment response. *Med Decis Making* 2005;25(5):511-9.
28. Tosh J, Stevenson M, Akehurst R. Health economic modelling of treatment sequences for rheumatoid arthritis: a systematic review. *Current Rheumatology Reports* 2014;16(10):447.
29. Jameson JL, Longo DL. Precision medicine--personalized, problematic, and promising. *New England Journal of Medicine* 2015;372(23):2229-34.
30. Latimer NR, Abrams KR. NICE DSU Technical Support Document 16: Adjusting survival time estimates in the presence of treatment switching. In: Decision Support Unit S, editor. *NICE DSU Technical support*. University of Sheffield, 2014.
31. Miller JD, Foley KA, Russell MW, MAPE. Current Challenges in Health Economic Modeling of Cancer Therapies: A Research Inquiry. *American Health & Drug Benefits* 2014;7(3):153-62.
32. Ciani O, Jommi C. The role of health technology assessment bodies in shaping drug development. *Drug design, development and therapy* 2014;8:2273-81.
33. Crizotinib for previously treated anaplastic lymphoma kinase-positive advanced non-small-cell lung cancer. Technology appraisal guidance [TA422] Published date: 21 December 2016. Available from: <https://www.nice.org.uk/guidance/ta422> [Accessed 19-08-18].
34. Ceritinib for previously treated anaplastic lymphoma kinase positive non-small-cell lung cancer. Technology appraisal guidance [TA395] Published date: 22 June 2016. Available from: <https://www.nice.org.uk/Guidance/TA395> [Accessed 19-08-18].
35. Crizotinib for untreated anaplastic lymphoma kinase-positive advanced non-small-cell lung cancer. Technology appraisal guidance [TA406] Published date: 28 September 2016. Available from: <https://www.nice.org.uk/guidance/ta406> [Accessed 19-08-18].
36. Ceritinib for untreated ALK-positive non-small-cell lung cancer. Technology appraisal guidance [TA500] Published date: 24 January 2018. Available from: <https://www.nice.org.uk/guidance/ta500> [Accessed 19-08-18].
37. Tikhonova IA, Huxley N, Snowsill T, Crathorne L, Varley-Campbell J, Napier M, et al. Economic Analysis of First-Line Treatment with Cetuximab or Panitumumab for RAS Wild-Type Metastatic Colorectal Cancer in England *PharmacoEconomics - Italian Research Articles* 2018;36(7):837-51.

38. Turner RM, Spiegelhalter DJ, Smith GCS, Thompson SG. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society Series A* 2009;172(Part 1):21-47.
39. Eddy DM, Hasselblad V, Shachter R. An Introduction to Bayesian Methods for Meta-analysis: The Confidence Profile Method. *Med Decis Making* 1990;10(1):15-23.
40. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
41. Hyrich KL, Lunt M, Watson KD, Symmons DP, Silman AJ, British Society for Rheumatology Biologics Register. Outcomes after switching from one anti-tumor necrosis factor alpha agent to a second anti-tumor necrosis factor alpha agent in patients with rheumatoid arthritis: results from a large UK national cohort study. *Arthritis & Rheumatology* 2007;56(1):13-20.
42. Smolen JS, Burmester GR, Combe B, Curtis JR, Hall S, Haraoui B, et al. Head-to-head comparison of certolizumab pegol versus adalimumab in rheumatoid arthritis: 2-year efficacy and safety results from the randomised EXXELERATE study. *Lancet* 2016;388(10061):2763-74.
43. Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data *Clinical Pharmacology & Therapeutics* 2016;99:262-5.
44. Schneeweiss S. Learning from big health care data. *New England Journal of Medicine* 2014;370:2161-3.
45. Bell H, Wailoo AJ, Hernandez M, Grieve R, Faria R, Gibson L, et al. *The use of real world data for the estimation of treatment effects in NICE decision making (updated 12th December). Report by the decision support unit.* Sheffield: NICE Decision Support Unit, SchARR, 2016.
46. Sheldon TA. Making evidence synthesis more useful for management and policy-making [Editorial]. *Journal of Health Services Research & Policy* 2005;10 (1):S1:5.
47. Chalmers I. Trying to do more Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-Date Evaluations. *The Annals of the American Academy of Political and Social Science* 2003;589:22-40.
48. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technology Assessment* 1998;2(19):1-276.
49. Dixon-Woods M, Agarwal S, Jones DR, Young B, Sutton A. Synthesising qualitative and quantitative evidence: a review of possible methods. *Journal of Health Services Research & Policy* 2005;10(1):45–53.
50. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing Methodological Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: Effectiveness of Interventions to Promote Smoke Alarm Ownership and Function. *Evaluation* 2009;15(1):47–71.
51. Lalonde RL, Willke RJ. Comparative Efficacy and Effectiveness: An Opportunity for Clinical Pharmacology *Clinical pharmacology & Therapeutics* 2011;90(6):761-63.
52. Hennessy S. When Should We Believe Nonrandomized Studies of Comparative Effectiveness? *Clinical pharmacology & Therapeutics* 2011;90(6):764-66.
53. Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *The Lancet* 2008;372(9656):2152 - 61.
54. Ades A, Sutton AJ. Multiple parameter evidence synthesis in epidemiology and medical decision-making: current approaches. *Journal of the Royal Statistical Society Series A* 2006;169(1):5-35.
55. Ades AE, Welton NJ, Caldwell DM, Price M, Goubar A, Lu G. Multiparameter evidence synthesis in epidemiology and medical decision-making. *Journal of Health Services Research Policy* 2008;13(3):12-22.
56. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897-900.

57. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. *Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis*. Ottawa: Canadian Agency for Drugs and Technologies in Health, 2009.
58. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine* 2008;27:625-50.
59. Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, et al. GetReal in network meta-analysis: a review of the methodology. *Research Synthesis Methods* 2016;7(3):236-63.
60. Welton NJ, Willis SR, Ades AE. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Research Synthesis Methods* 2010;1(3-4):239-57.
61. Welton N, Caldwell D, Adamopoulos E, Vedhara K. Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease. *American Journal of Epidemiology* 2009;169(9):1158-65.
62. Welton NJ. Estimation of Markov Chain Transition Probabilities and Rates from Fully and Partially Observed Data: Uncertainty Propagation, Evidence Synthesis, and Model Calibration. *Med Decis Making* 2005;25:633-45.
63. Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times *Statistics in Medicine* 2007;26 3681-99.
64. Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Statistics in Medicine* 2003;22:2995-3016.
65. Spiegelhalter D, Best N. Bayesian approaches to multiple sources of evidence and uncertainty in complex cost effectiveness modelling. *Statistics in Medicine* 2003;22:3687.
66. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirical based priors. *Journal of the Royal Statistical Society* 2009;172(1):119-36.
67. Riley RD, Simmonds MC, Look MP. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. *Journal of Clinical Epidemiology* 2007;60:431-39.
68. Riley RD, Steyerberg EW. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods* 2010;1:2-19.
69. Riley RD, Lambert PC, Staessen JA, Wang J, Gueyffier F, Thijs L, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine* 2008;27(11):1870-93.
70. Jansen JP. Network meta-analysis of individual and aggregate level data. *Research Synthesis Methods* 2012;3(2):177-90.
71. Saramago P, Sutton AJ, Cooper NJ, Manca A. Mixed treatment comparisons using aggregate and individual participant level data. *Statistics in Medicine* 2012;31(28):3516-36.
72. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for Population-Adjusted Indirect Comparisons in Health Technology Appraisal. *Med Decis Making* 2018;38(2):200-11.
73. Signorovitch JE, Sikirica V, Erder MH, Xie J, Lu M, Hodgkins PS, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. *Value in Health* 2012;15(6):940-7.
74. Tudur Smith C, Nevitt S, Appelbe D, Appleton R, Dixon P, Harrison J, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials* 2017;18:319.
75. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open* 2017;7(12):e018647.
76. Doshi P, Groves T. Editorial: Clinical trial data: get them while you can. *BMJ* 2014;348:g63

77. Mintzes B, Lexchin J, Quintano AS. Clinical trial transparency: many gains but access to evidence for new medicines remains imperfect. *British Medical Bulletin* 2015;116(1):43–53.
78. Packer M. Data sharing in medical research. *BMJ* 2018;360:k510.
79. Shokraneh F, Adams CE, Clarke M, Amato L, Bastian H, Beller E, et al. Why Cochrane should prioritise sharing data. *2018* 2018;362 k3229
80. Pisani E, Aaby P, Breugelmans JG, Carr D, Groves T, Helinski M, et al. Beyond open data: realising the health benefits of sharing data. *BMJ* 2016;355:i5295.
81. McCabe C, Dixon S. Testing the validity of cost-effectiveness models. *Pharmacoeconomics* 2000;17(5):501-13.
82. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Economics* 2006;15(4):373-81.
83. Saramago P, Manca A, Sutton AJ. Deriving input parameters for cost-effectiveness modeling: taxonomy of data types and approaches to their statistical synthesis. *Value in Health* 2012;15(5):639-49.
84. Cooper NJ, Sutton AJ, Ades AE, Paisley S, Jones DR, Working Group on the Use of Evidence in Economic Decision Models. Use of evidence in economic decision models: practical issues and methodological challenges. *Health Economics* 2007;16(12):1277-86.
85. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD, et al. Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group-6. *Med Decis Making* 2012;32(5):722-32.
86. Briggs A, Sculpher M, K. C. Chapter 2: Key aspects of decision modelling for economic evaluation. *Decision Modelling for Health Economic Evaluation*. Oxford: University Press, 2006.
87. Alemayehu D. Perspectives on Pooled Data Analysis: the Case for an Integrated Approach. *Journal of Data Science* 2011;9:389-97.
88. Cooper NJ, Sutton AJ, Abrams KR, Turner D, Wailoo A. Comprehensive decision analytical modelling in economic evaluation: A Bayesian approach *Health Economics* 2004;13:203-26.
89. Samsa G, Reutter R, Parmigiani G, Ancukiewicz M, Abrahamse P, Lipscomb J, et al. Performing cost effectiveness analysis by integrating randomized trial data with a comprehensive decision model: application to treatment of acute ischemic stroke. *Journal of Clinical Epidemiology* 1999;52(3):259-71.
90. Briggs A, Sculpher M, K. C. *Decision Modelling for Health Economic Evaluation*. Oxford: University Press, 2006.
91. Rutter CM, Zaslavsky AM, Feuer EJ. Dynamic microsimulation models for health outcomes: a review. *Med Decis Making* 2011;31(1):10-8.
92. Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole Disease Modeling to Inform Resource Allocation Decisions in Cancer: A Methodological Framework. *Value in Health* 2012;15:1127-36.
93. Tappenden P, Chilcott J, Brennan A, Squires H, Glynne-Jones R, Tappenden J. Using whole disease modeling to inform resource allocation decisions: economic evaluation of a clinical guideline for colorectal cancer using a single model. *Value in Health* 2013;16(4):542-53.
94. Lord J, Willis S, Eatock J, Tappenden P, Trapero-Bertran M, Miners A, et al. Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technology Assessment* 2013;17(58):1-192.
95. Eddy DM, Schlessinger L. Archimedes: a trial-validated model of diabetes. *Diabetes Care* 2003;26(11):3093-101.
96. Schlessinger L, David M, Eddy DM. Archimedes: a new model for simulating health care systems - the mathematical formulation. *Journal of Biomedical Informatics* 2002;35(1):37-50.
97. Tordrup D, Chouaid C, Cuijpers P, William D, van Dongen JM, Espin J, et al. Priorities for Health Economic Methodological Research: Results of an Expert Consultation. *International Journal of Technology Assessment in Health Care* 2017;33(6):609-19.

98. Hamilton W, Walter FM, Rubin G, Neal RD. Improving early diagnosis of symptomatic cancer. . *Nature Reviews Clinical Oncology* 2016;13(12):740-49.
99. Guise JM, Chang C, Butler M, Viswanathan M, Tugwell P. AHRQ series on complex intervention systematic reviews-paper 1: an introduction to a series of articles that provide guidance and tools for reviews of complex interventions. *Journal of Clinical Epidemiology* 2017;90(6-10).
100. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* 2008;337:a1655.
101. EUnetHTA. EUnetHTA Project: Overview of Results 2006-2008 *EUnetHTA Secretariat*, 2009.
102. Eddy D. Health Technology Assessment and Evidence-Based Medicine: What Are We Talking About? *Value in Health* 2009;12(2).
103. Taylor M, Righetti C, Conway P, Lebmeier M. Economic evaluation of therapeutic approaches for rheumatoid arthritis: The case for sequential anti-TNF use. *Rheumatology* 2009;48:i112.
104. Zheng Y, Pan F, Sorensen S. Modeling Treatment Sequences in Pharmacoeconomic Models. *Pharmacoeconomics* 2017;35(1):15-24.
105. Abraham I. More research is needed—but what type? *BMJ* 2010;341:c4662.
106. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet]*. Rockville (MD): Agency for Healthcare Research and Quality (US). Available from: <https://www.ncbi.nlm.nih.gov/books/NBK47095/> [Accessed 06/12/2018], 2008.
107. Antoniou M, Jorgensen AL, Kolamunnage-Dona R. Biomarker-Guided Adaptive Trial Designs in Phase II and Phase III: A Methodological Review. *PLoS One* 2016;11(2):e0149803.
108. Renfro LA, An MW, Mandrekar SJ. Precision oncology: A new era of cancer clinical trials. . *Cancer Letters* 2017;387:121-26.
109. Isaacs JD, Ferraccioli G. The need for personalised medicine for rheumatoid arthritis. *Ann Rheum Dis* 2011;70(1):4-7.
110. Duan N, Kravitz RL, Schmid CH. Single-patient (n-of-1) trials: a pragmatic clinical decision methodology for patient-centered comparative effectiveness research. *Journal of Clinical Epidemiology* 2013;66(8 Suppl):S21-8.
111. Senn S. Chapter 6: Within patient studies: Cross-over trials and n-of-1 Studies. In: Max MB, Lynn J, editors. *Interactive Textbook on Clinical Symptom Research: Methods and Opportunities*. Watertown, United States: New England Research Institute, 2003:Internet [accessed 10 September 2011]. Available from: http://painconsortium.nih.gov/symptomresearch/chapter_6/index.htm.
112. Peloso PM. Are individual patient trials (n-of-1 trials) in rheumatology worth the extra effort? *The Journal of Rheumatology* 2004;31(1):8-11.
113. Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. *The Journal of Rheumatology* 1991;18:447-51.
114. Lewis RA, Williams NH, Sutton AJ, Burton K, Din NU, Matar HE, et al. Comparative clinical effectiveness of management strategies for sciatica: systematic review and network meta-analyses. *The Spine Journal* 2015;15(6):1461-77.
115. Fitzsimmons D, Phillips CJ, Bennett H, Jones M, Williams N, Lewis R, et al. Cost effectiveness of different management strategies for sciatica. *Pain* 2014;155(2014):1318-27.
116. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *Journal of the American Medical Association* 1992;268(6):760-65.
117. Bush K, Cowan N, Katz DE, Gishen P. The natural history of sciatica associated with disc pathology. A prospective study with clinical and independent radiologic follow-up. *Spine* 1992;17(10):1205-12.
118. Tubach F, Beaute J, Leclerc A. Natural history and prognostic indicators of sciatica. *J Clin Epidemiol* 2004;57(2):174-79.
119. Koes BW, van Tulder MW, Peul WC. Clinical Review: Diagnosis and treatment of sciatica *BMJ* 2007;334(7607):1313-17.

120. Weber H, Holme I, Amlie E. The natural course of acute sciatica with nerve root symptoms in a double-blind placebo-controlled trial evaluating the effect of piroxicam. *Spine* 1993;18(11):1433-8.
121. Mixer WJ, Barr JS. Rupture of the intervertebral disc with involvement of the spinal canal. *N Engl J Med* 1934;211:210-15.
122. Samanta A, Beardsley J. Evidence based case report: Sciatica: which intervention? . *BMJ* 1999;319:302-3.
123. Konstantinou K, Dunn KM. Sciatica: Review of epidemiological studies and prevalence estimates. *Spine* 2008;33:2464-72.
124. Vroomen PC, de Krom MC, Slofstra PD, Knottnerus JA. Conservative treatment of sciatica: a systematic review. *J Spinal Disord* 2000;13(6):463-9.
125. Hagen KB, Hilde G, Jamtvedt G, Winnem M. Bed rest for acute low back pain and sciatica. *Cochrane Database of Systematic Reviews* 2010;June(6):CD001254.
126. Luijsterburg PAJ, Verhagen AP, Ostelo RWJG, van Os TAG, Peul WC, Koes BW. Effectiveness of conservative treatments for the lumbosacral radicular syndrome: a systematic review. *Eur Spine J* 2007;16(7):881-99.
127. Clarke JA, van Tulder MW, Blomberg SEI, de Vet HCW, van der Heijden GJMG, Bronfort G, et al. Traction for low-back pain with or without sciatica. *Cochrane Database of Systematic Reviews* 2007;Apr 18(2):CD003010.
128. Boswell MV, Hansen HC, Trescot AM, Hirsch JA. Epidural steroids in the management of chronic spinal pain and radiculopathy. *Pain physician* 2003;6(3):319-34.
129. Abdi S, Datta S, Trescot AM, Schultz DM, Adlaka R, Atluri SL, et al. Epidural steroids in the management of chronic spinal pain: a systematic review. *Pain physician* 2007;10(1):185-212.
130. Gibson JNA, Waddell G. Surgical interventions for lumbar disc prolapse. *Cochrane Database of Systematic Reviews* 2007;Jan 24(2):CD001350.
131. Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, et al. Health Outcomes Associated With Various Antihypertensive Therapies Used as First-Line Agents: A Network Meta-analysis *JAMA* 2003;289(19):2534-44
132. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine* 2004;23:3105-24.
133. Cooper NJ, Sutton AJ, Lu G, Khunti K. Mixed Comparison of Stroke Prevention Treatments in Individuals With Nonrheumatic Atrial Fibrillation *Archives of Internal Medicine* 2006;166:1269-75.
134. Anothaisintawee T, Attia J, Nickel JC, Thammakraisorn S, Numthavaj P, McEvoy M, et al. Management of Chronic Prostatitis/ Chronic Pelvic Pain Syndrome: A Systematic Review and Network Meta-analysis *JAMA* 2011;305 (1):78-86.
135. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine* 1996;15:2733-49.
136. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001;10:277-303.
137. Bland MJ, Altman DG. Statistics notes: Bayesians and frequentists. *BMJ* 1998;317(166):1151-60.
138. Ashby D, Smith AFM. Evidence-based medicine as Bayesian decision-making. *Statistics in Medicine* 2000;19:3291-305.
139. Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *Journal of Clinical Epidemiology* 2011;64:163-71.
140. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Healthcare* York, UK: CRD, University of York, 2009.
141. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration *BMJ* 2009;339(b2700).

142. van Tulder MW, Furlan A, Bombardier C, Bouter LM, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 2003;28(12):1290-99.
143. Jackson N, Waters E, For the Guidelines for Systematic Reviews in Health Promotion and Public health Taskforce. Criteria for the systematic review of health promotion and public health interventions. *Health Promotion International* 2005;20(4):367-74.
144. Effective Public Health Practice Project (EPHPP). Quality Assessment tool for Quantitative Studies, 2009.
145. Higgins JPT, Deeks JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester (UK): John Wiley & Sons Ltd, 2008:151-85.
146. Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol*. 2006;59(1):7-10.
147. Spiegelhalter DJ, Thomas A, Best NG, Lunn DJ, editors. *WinBUGS Version v1.4. User manual*. Cambridge: MRC Biostatistics Unit, 2003.
148. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled clinical trials* 1986;7:177-88.
149. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials. *J Clin Epidemiol* 1997;50(6):683-91.
150. Song F, Altman DG, Glenny A-M, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472-76.
151. Stettler C, Allemann S, Wandel S, Kastrati A, Morice MC, Schömig A, et al. Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis *BMJ* 2008;337:a1331.
152. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester (UK): John Wiley & Sons, 2008:243-96.
153. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-60
154. Armon C, Argoff CE, Samuels J, Backonja MM. Assessment: Use of epidural steroid injections to treat radicular lumbosacral pain: Report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology. *Neurology* 2007;68(10):723-29.
155. Chou R, Atlas SJ, Stanos SP, Rosenquist RW. Nonsurgical interventional therapies for low back pain: a review of the evidence for an American Pain Society clinical practice guideline. *Spine* 2009;34(10):1078-93.
156. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003;7(27).
157. Egger M, Smith GD, Schneider M. Chapter 12: Systematic reviews of observational studies. In: Egger M, Smith GD, Altman D, editors. *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books, 2001:211-27.
158. Wiebe N, Vandermeer B, Platt RW, Klassen TP, Moher D, Barrowman NJ. A systematic review identifies a lack of standardization in methods for handling missing variance data. *Journal of Clinical Epidemiology* 2006;59(4):342-53.
159. Thompson S, Ekelund U, Jebb S, Lindroos AK, Mander A, Sharp S, et al. A proposed method of bias adjustment for meta-analyses of published observational studies. *International Journal of Epidemiology* 2011;40:765–77
160. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Medicine* 2013;11:159.
161. Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention

- treatments in individuals with non-rheumatic atrial fibrillation. *Statistics in Medicine* 2009;28(14):1861-81.
162. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: Many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods* 2012;3(2):80-97.
 163. Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network meta-analysis using R: a review of currently available automated packages. *PLoS One* 2014;9(12):e115065.
 164. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ* 2017;358:j3932.
 165. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 2: A Generalised Linear Modelling Framework for Pairwise and Network Meta-Analysis of Randomised Controlled Trials (updated August 2011) In: Decision Support Unit, editor. Sheffield: SchARR, University of Sheffield, 2011.
 166. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Chapter 13: Fixed-Effect Versus Random-Effects Models. In: Borenstein M, Hedges LV, Higgins JP, Rothstein HR, editors. *Introduction to Meta-Analysis* (eds). Chichester, UK: Wiley, 2009.
 167. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006;1(3):515–33.
 168. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemporary Clinical Trials* 2015;45(Pt A):130-8.
 169. Henmi M, Copas JB. Confidence intervals for random effects meta-analysis and robustness to publication bias. *Statistics in Medicine* 2010;29(29):92969-83.
 170. Moreno SG, Sutton AJ, Thompson JR, Ades AE, Abrams KR, Cooper NJ. A generalized weighting regression-derived meta-analysis estimator robust to small-study effects and heterogeneity. *Statistics in Medicine* 2012;31(14):1407-17.
 171. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Chapter 17: Prediction Intervals. In: Borenstein M, Hedges LV, Higgins JP, Rothstein HR, editors. *Introduction to Meta-Analysis* (eds). Chichester, UK: Wiley, 2009.
 172. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 2016;6:e010247.
 173. Graham PL, Moran JL. Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *Journal of Clinical Epidemiology* 2012;65(5):503-10.
 174. Verde PE, Ohmann C. Combining randomized and non-randomized evidence in clinical research: a review of methods and applications. *Research Synthesis Methods* 2015;6(1):45-62.
 175. Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RH, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research Synthesis Methods* 2015;6(4):293-309.
 176. Jackson D, Riley RD. A refined method for multivariate meta-analysis and meta-regression. *Statistics in Medicine* 2014;33(4):541-54.
 177. Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. *Journal of Clinical Epidemiology* 2017;82:20-28.
 178. Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *Journal of Clinical Epidemiology* 2014;67(2):138-43.
 179. Zarin W, Veroniki AA, Nincic V, Vafaei A, Reynen E, Motiwala SS, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. *BMC Medicine* 2017;15(1):3.

180. National Institute for Health and Care Excellence. 5. Identifying the evidence: literature searching and evidence submission. *Developing NICE guidelines: the manual (PMG20)*. London, UK: NICE, 2014.
181. Day D, Furlan A, Irvin E, Bombardier C. Simplified search strategies were effective in identifying clinical trials of pharmaceuticals and physical modalities. *Journal of Clinical Epidemiology* 2005;58(9):874-81.
182. Royle P, Milne R. Literature searching for randomized controlled trials used in Cochrane reviews: rapid versus exhaustive searches. *International Journal of Technology Assessment in Health Care* 2003;19(4):591-603.
183. Kaltenthaler E, Tappenden P, Paisley S, Squires H. *NICE DSU Technical Support Document 13: Identifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-Effectiveness Models*. Sheffield: NICE Decision Support Unit, ScHARR, 2011.
184. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *British Medical Journal* 2005;331:1064-65.
185. Mauskopf J, Samuel M, McBride D, Mallya UG, Feldman SR. Treatment sequencing after failure of the first biologic in cost-effectiveness models of psoriasis: a systematic review of published models and clinical practice guidelines *Pharmacoeconomics* 2014;32(4):395-409.
186. Sullivan SD, Alfonso-Cristancho R, Carlson J, Mallya U, Ringold S. Economic consequences of sequencing biologics in rheumatoid arthritis: a systematic review. *J Med Econ* 2013;16(3):391-6.
187. Reynolds A, Koenig AS, Bananis E, Singh A. When is switching warranted among biologic therapies in rheumatoid arthritis? *Expert Rev Pharmacoeconomics & Outcomes Research* 2012;12(3):319-33.
188. Barbieri M, Drummond MF, Puig Junoy J, Casado Gomez MA, Garcia FJB, Segura PB, et al. Critical appraisal of pharmacoeconomic studies comparing TNF- α antagonists for rheumatoid arthritis treatment. *Expert Reviews. Pharmacoeconomics Outcomes Research* 2007;7(6):613-26.
189. Madan J, Ades AE, Welton NJ. An overview of models used in economic analyses of biologic therapies for arthritis-from current diversity to future consensus. *Rheumatology* 2011;50(suppl 4):iv10-iv18.
190. Marra CA, Bansback N, Anis AH, Shojania K. Introduction to economic modeling for clinical rheumatologists: application to biologic agents in rheumatoid arthritis. *Clinical Rheumatology* 2011;30 (Suppl 1):S9-S18.
191. Van Der Velde G, Pham B, MacHado M, Ieraci L, Witteman W, Bombardier C, et al. Cost-effectiveness of biologic response modifiers compared to disease-modifying antirheumatic drugs for rheumatoid arthritis: a systematic review. *Arthritis Care and Research* 2011;63(1):65-78.
192. Fleurence R, Spackman E. Cost-effectiveness of biologic agents for treatment of autoimmune disorders: structured review of the literature. *The Journal of Rheumatology* 2006;33(11):2124-31.
193. Scholz S, Mittendorf T. Modeling rheumatoid arthritis using different techniques - a review of model construction and results. *Health Economics Review* 2014;4(18).
194. Jafri NS, Hornung CA, Howden CW. Meta-analysis: Sequential Therapy Appears Superior to Standard Therapy for Helicobacter pylori Infection in Patients Naive to Treatment. *Annals of Internal Medicine* 2008;148(12):923-31.
195. Briggs A, Sidhu M, Baker T. A framework for analysing treatment sequences: Incorporating time dependent transitions through partitioned survival analysis. *Value in Health* 2014;17(3):A204-A05.
196. Heeg B, Van Agthoven M, Liwing J, Mellqvist UH, Plesner T, Logman F, et al. Optimal treatment sequencing in multiple myeloma: An exploratory modeling approach. *Blood* 2010;116 (21).
197. Heeg BMS. Developing health economic models of chronic diseases for reimbursement purposes. Groningen: University of Groningen 2015.

198. Christensen AW, Tarp S, Furst DE, Døssing A, Amris K, Bliddal H, et al. Most Trial Eligibility Criteria and Patient Baseline Characteristics Do Not Modify Treatment Effect in Trials Using Targeted Therapies for Rheumatoid Arthritis: A Meta-Epidemiological Study. *PLoS One* 2015;10(9):e0136982.
199. Connock M, Frew E, Evans B-W, Bryan S, Cummins C, Fry-Smith A, et al. The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review. *Health Technology Assessment* 2006;10(7).
200. Cooper C, Katona C, Lyketsos K, Blazer D, Brodaty H, Rabins P, et al. A systematic review of treatments for refractory depression in older people. *American Journal of Psychiatry* 2011;168(7):681-8.
201. Finnerup NB, Otto M, McQuay HJ, Jensen TS, Sindrup SH. Algorithm for neuropathic pain treatment: An evidence based proposal. *Pain* 2005;118(3):289-305.
202. Grothey A, Sargent D, Goldberg RM, Schmoll HJ. Survival of patients with advanced colorectal cancer improves with the availability of fluorouracil-leucovorin, irinotecan, and oxaliplatin in the course of treatment. *Journal of Clinical Oncology* 2004;22(7):1209-14.
203. Heng DY, Signorovitch J, Swallow E, Li N, Zhong Y, Qin P, et al. Comparative Effectiveness of Second-Line Targeted Therapies for Metastatic Renal Cell Carcinoma: A Systematic Review and Meta-Analysis of Real-World Observational Studies. *PLoS One* 2014;9(12):e114264.
204. Hind D, Tappenden P, Tumor I, Eggington S, Sutcliffe P, Ryan A. The use of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer: systematic review and economic evaluation. *Health Technol Assess* 2008;12(15):iii-ix, xi-162.
205. Kanters S, Druyts, E., Mills EJ, Thorlund K. What drives the comparative effectiveness of biologics vs. methotrexate in rheumatoid arthritis? Meta-regression and graphical inspection of suspected clinical factors. *Rheumatology* 2014;53(7):1264-73.
206. Mandema JW, Salinger DH, Baumgartner SW, Gibbs MA. A Dose-Response Meta-Analysis for Quantifying Relative Efficacy of Biologics in Rheumatoid Arthritis. *Clinical Pharmacology & Therapeutics* 2011;90:828-35.
207. National Institute for Health and Care Excellence. *Colorectal cancer: NICE clinical guideline 131*. London: NICE, 2011.
208. Nixon RA, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statistics in Medicine* 2007;26:1237–54.
209. Ruhé HG, Huyser J, Swinkels JA, Schene AH. Switching antidepressants after a first selective serotonin reuptake inhibitor in major depressive disorder: a systematic review. *J Clin Psychiatry* 2006;67(12):1836-55.
210. Salliot C, Finckh A, Katchamart W, Lu Y, Sun Y, Bombardier C, et al. Indirect comparisons of the efficacy of biological antirheumatic agents in rheumatoid arthritis in patients with an inadequate response to conventional disease-modifying antirheumatic drugs or to an anti-tumour necrosis factor agent: a meta-analysis. *Ann Rheum Dis* 2011;70(2):266-71.
211. Schmitz S, Adams R, Walsh CD, Barry M, FitzGerald O. A mixed treatment comparison of the efficacy of anti-TNF agents in rheumatoid arthritis for methotrexate non-responders demonstrates differences between treatments: a Bayesian approach. *Ann Rheum Dis* 2012;71(2):225-30.
212. Schoels M, Aletaha D, Smolen JS, Wong JB. Comparative effectiveness and safety of biological treatment options after tumour necrosis factor α inhibitor failure in rheumatoid arthritis: systematic review and indirect pairwise meta-analysis. *Ann Rheum Dis* 2012;71:1303–08.
213. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA, et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *Canadian Medical Association Journal* 2009;181(11):787-96.
214. Stenner F, Chastonay R, Liewen H, Haile SR, Cathomas R, Rothermundt C, et al. A pooled analysis of sequential therapies with sorafenib and sunitinib in metastatic renal cell carcinoma. *Oncology* 2012(6):333-40.

215. Suarez-Almazor M, Ortiz Z, Lopez-Olivo M, Pak C, Skidmore B, Kimmel B, et al. Infliximab and etanercept in rheumatoid arthritis: timing, dose escalation, and switching [Technology report no 86] Ottawa: Canadian Agency for Drugs and Technologies in Health 2007.
216. Albert DA, Aksentijevich S, Hurst S, Fries JF, Wolfe F. Modeling therapeutic strategies in rheumatoid arthritis: use of decision analysis and Markov models. *J Rheumatol* 2000;27(3):644-52.
217. Anis AH, Bansback N, Sizto S, Gupta SR, Willian MK, Feldman SR. Economic evaluation of biologic therapies for the treatment of moderate to severe psoriasis in the United States. *J Dermatolog Treat* 2011;22(2):65-74.
218. Bansback NJ, Brennan A, Ghatnekar O. Cost effectiveness of adalimumab in the treatment of patients with moderate to severe rheumatoid arthritis in Sweden. *Ann Rheum Dis* 2005;64:995-1002.
219. Barton P, Jobanputra P, Wilson J, Bryan S, Burls A. The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis. *Health Technol Assess* 2004;8(11):iii, 1-91.
220. Beard SM, Roskell N, Le TK, Zhao Y, Coleman A, Ang D, et al. Cost effectiveness of duloxetine in the treatment of fibromyalgia in the United States. *J Med Econ* 2011;14(4):463-76.
221. Bensmail D, Ward AB, Wissel J, Motta F, Saltuari L, Lissens J, et al. Cost-effectiveness modeling of intrathecal baclofen therapy versus other interventions for disabling spasticity. *Neurorehabil Neural Repair* 2009;23(6):546-52.
222. Beresniak A, Ariza-ariza R, Garcia-Llorente JF, Ramirez-Arellano A, Dupont D. Modelling cost-effectiveness of biologic treatments based on disease activity scores for management of rheumatoid arthritis in Spain. *International Journal of Inflammation* 2011.
223. Beresniak A, Baerwald C, Zeidler H, Krüger K, Neubauer AS, Dupont D, et al. Cost-effectiveness simulation model of biologic strategies for treating to target rheumatoid arthritis in Germany. *Clinical and Experimental Rheumatology*. 2013 31(3):400-8.
224. Brennan A, Bansback N, Reynolds A, Conway P. Modelling the cost-effectiveness of etanercept in adults with rheumatoid arthritis in the UK. *Rheumatology* 2004;43(1):62–72.
225. Brennan A, Bansback N, Nixon R, Madan J, Harrison M, Watson K, et al. Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology* 2007;46(8):1345-54.
226. Cameron DA, Camidge DR, Oyee J, Hirsch M. Economic evaluation of fulvestrant as an extra step in the treatment sequence for ER-positive advanced breast cancer. *Br J Cancer* 2008;99(12):1984-90.
227. Chen Y-F, Jobanputra P, Barton PM, Bryan S, Clark W, Fry-Smith A, et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technology Assessment* 2006;10(42):1-248.
228. Cimmino MA, Leardini G, Salaffi F, Intorcchia M, Bellatreccia A, Dupont D, et al. Assessing the cost-effectiveness of biologic agents for the management of moderate-to-severe rheumatoid arthritis in anti-TNF inadequate responders in Italy: a modelling approach. *Clin Exp Rheumatol* 2011;29(4):633-41.
229. Clark W, Jobanputra P, Barton P, Burls A. The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis. *Health Technol Assess* 2004;8(18):iii-iv, ix-x, 1-105.
230. Coyle D, Judd M, Blumenauer B, Cranney A, Maetzel A, Tugwell P, et al. *Infliximab and etanercept in patients with rheumatoid arthritis: a systematic review and economic evaluation*. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 2006.
231. Davies A, Vardeva K, Loze J-Y, L'Italien GJ, Sennfalt K, van Baardewijk M. Cost-effectiveness of atypical antipsychotics for the management of schizophrenia in the UK. *Curr Med Res Opin* 2008;24(11):3275-85.

232. Davies A, Cifaldi MA, Segurado OG, Weisman MH. Cost-effectiveness of sequential therapy with tumor necrosis factor antagonists in early rheumatoid arthritis. *J Rheumatol* 2009;36(1):16-26.
233. Denis P, Le Pen C, Umuhire D, Berdeaux G. Treatment carryover impacts on effectiveness of intraocular pressure lowering agents, estimated by a discrete event simulation model. *Eur J Ophthalmol* 2008;18(1):44-51.
234. Diamantopoulos A, Benucci M, Capri S, Berger W, Wintfeld N, Giuliani G, et al. Economic evaluation of tocilizumab combination in the treatment of moderate-to-severe rheumatoid arthritis in Italy. *J Med Econ* 2012;15(3):576-85.
235. Diamantopoulos A, Finckh A, Huizinga T, Sungher DK, Sawyer L, Neto D, et al. Tocilizumab in the treatment of rheumatoid arthritis: a cost-effectiveness analysis in the UK. *Pharmacoeconomics* 2014;32(8):775-87.
236. Dranitsaris G, Truter I, Lubbe MS, Sriramanakoppa NN, Mendonca VM, Mahagaonkar SB. Using Pharmacoeconomic Modelling to Determine Value-Based Pricing for New Pharmaceuticals in Malaysia *Malaysian Journal of Medical Sciences* 2011;18(4):32-43.
237. Erhardt W, Bergenheim K, Duprat-Lomon I, McEwan P. Cost effectiveness of saxagliptin and metformin versus sulfonylurea and metformin in the treatment of type 2 diabetes mellitus in Germany: a Cardiff diabetes model analysis. *Clin Drug Invest* 2012;32(3):189-202.
238. Finckh A, Bansback N, Marra CA, Anis AH, Michaud K, Lubin S, et al. Treatment of very early rheumatoid arthritis with symptomatic therapy, disease-modifying antirheumatic drugs, or biologic agents: a cost-effectiveness analysis. *Annals of Internal Medicine* 2009;151(9):612-21.
239. Frankum LE, Nightengale B, Russo CL, Sarnes M. Pharmacoeconomic analysis of sequential treatment pathways in the treatment of onychomycosis. *Manag Care Interface* 2005;18(1):55-63.
240. Greenhalgh J, Knight C, Hind D, Beverley C, Walters S. Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies. *Health Technology Assessment* 2005;9(9):1-156, iii-iv.
241. Hallinen TA, Soini EJ, Eklund K, Puolakka K. Cost-utility of different treatment strategies after the failure of tumour necrosis factor inhibitor in rheumatoid arthritis in the Finnish setting. *Rheumatology* 2010;49(4):767-77.
242. Heeg B, Buskens E, Botteman M, Caleo S, Ingham M, Damen J, et al. The cost-effectiveness of atypicals in the UK. *Value in Health* 2008;11(7):1007-21.
243. Heeg B, van Agthoven M, van Beurden-Tan C, Liwing J, Mellqvist U-H, Plesner T, et al. Chapter 5: Treatment Sequencing Survival Model for patients with Multiple Myeloma ineligible for stem cell transplantation (SCT). In: Heeg B, editor. *Developing health economic models of chronic diseases for reimbursement purposes*. Groningen: University of Groningen, 2015.
244. Holmes CB, Zheng H, Martinson NA, Freedberg KA, Walensky RP. Optimizing treatment for HIV-infected South African women exposed to single-dose nevirapine: balancing efficacy and cost. *Clin Infect Dis* 2006;42(12):1772-80.
245. Jobanputra P, Barton P, Bryan S, Burls A. The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation *Health Technology Assessment* 2002;6(21):1-110.
246. Kielhorn A, Porter D, Diamantopoulos A, Lewis G. UK cost-utility analysis of rituximab in patients with rheumatoid arthritis that failed to respond adequately to a biologic disease-modifying antirheumatic drug. *Current Medical Research and Opinion* 2008;24(9):2639-50.
247. Knoester PD, Deckers CL, Termeer EH, Boendermaker AJ, Kotsopoulos IA, de Krom MC, et al. A cost-effectiveness decision model for antiepileptic drug treatment in newly diagnosed epilepsy patients. *Value in Health* 2007;10(3):173-82.
248. Kobelt G, Lekander I, Lang A, Raffener B, Botsios C, Geborek P. Cost-effectiveness of etanercept treatment in early active rheumatoid arthritis followed by dose adjustment. *International Journal of Technology Assessment in Health Care* 2011;27(3):193-200.

249. Launois R, Payet S, Saidenberg-Kermanac'h N, Francesconi C, Riou Franca L, Boissier MC. Budget impact model of rituximab after failure of one or more TNFalpha inhibitor therapies in the treatment of rheumatoid arthritis. *Joint, Bone, Spine* 2008;75(6):688-95.
250. Lee HY, Yang BM, Hong JM, Lee TJ, Kim BG, Kim JW, et al. Cost-utility analysis for platinum-sensitive recurrent ovarian cancer therapy in South Korea: Results of the polyethylene glycolated liposomal doxorubicin/carboplatin sequencing model. *ClinicoEconomics and Outcomes Research* 2013;5(1):297-307.
251. Lindgren P, Geborek P, Kobelt G. Modeling the cost-effectiveness of treatment of rheumatoid arthritis with rituximab using registry data from Southern Sweden. *International Journal of Technology Assessment in Health Care* 2009;25(2):181-89.
252. Lux MP, Hartmann M, Jackisch C, Raab G, Schneeweiss A, Possinger K, et al. Cost-utility analysis for advanced breast cancer therapy in Germany: results of the fulvestrant sequencing model. *Breast Cancer Res Treat* 2009;117(2):305-17.
253. Maetzel A, Strand V, Tugwell P, Wells G, Bombardier C. Cost Effectiveness of Adding Leflunomide to a 5-Year Strategy of Conventional Disease-Modifying Antirheumatic Drugs in Patients with Rheumatoid Arthritis. *Arthritis & Rheumatism* 2002;47(6):655-61.
254. Malotki K, Barton P, Tsourapas A, Uthman AO, Liu Z, Routh K, et al. Adalimumab, etanercept, infliximab, rituximab and abatacept for the treatment of rheumatoid arthritis after the failure of a tumour necrosis factor inhibitor: a systematic review and economic evaluation. *Health Technology Assessment* 2011;15(14):1-278.
255. McEwan P, Evans M, Bergenheim K. A population model evaluating the costs and benefits associated with different oral treatment strategies in people with type 2 diabetes. *Diabetes Obes Metab* 2010;12(7):623-30.
256. Merkesdal S, Kirchhoff T, Wolka D, Ladinek G, Kielhorn A, Rubbert-Roth A. Cost-effectiveness analysis of rituximab treatment in patients in Germany with rheumatoid arthritis after etanercept-failure. *European Journal of Health Economics* 2010;11(1):95-104.
257. National Institute for Health and Care Excellence. *Breast cancer (advanced): NICE clinical guideline 81*. London: NICE, 2009.
258. National Institute for Health and Care Excellence. *Crohn's disease: management: NICE clinical guideline 152*. London: NICE, 2012.
259. Orme M, Collins S, Loftus J. Long-term medical management of primary open-angle glaucoma and ocular hypertension in the UK: optimizing cost-effectiveness and clinic resources by minimizing therapy switches. *J Glaucoma* 2012;21(7):433-49.
260. Puolakka K, Blafield H, Kauppi M, Luosujarvi R, Peltomaa R, Leikola-Pelho T, et al. Cost-effectiveness modelling of sequential biologic strategies for the treatment of moderate to severe rheumatoid arthritis in Finland. *Open Rheumatology Journal* 2012(1):38-43.
261. Russell A, Beresniak A, Bessette L, Haraoui B, Rahman P, Thorne C, et al. Cost-effectiveness modeling of abatacept versus other biologic agents in DMARDS and anti-TNF inadequate responders for the management of moderate to severe rheumatoid arthritis. *Clinical Rheumatology* 2009;28:403-12.
262. Saraux A, Gossec L, Goupille P, Bregman B, Boccard E, Dupont D, et al. Cost-effectiveness modelling of biological treatment sequences in moderate to severe rheumatoid arthritis in France. *Rheumatology* 2010;49(4):733-40.
263. Sawyer L, Samarasekera EJ, Wonderling D, Smith CH. Topical therapies for the treatment of localized plaque psoriasis in primary care: a cost-effectiveness analysis. *Br J Dermatol* 2013;168(5):1095-105.
264. Schadlich PK, Zeidler H, Zink A, Gromnica-Ihle E, Schneider M, Straub C, et al. Modelling cost effectiveness and cost utility of sequential DMARD therapy including leflunomide for rheumatoid arthritis in Germany: II. The contribution of leflunomide to efficiency. *Pharmacoeconomics* 2005;23(4):395-420.

265. Schipper LG, Wietske K, den Broeder AA, van der Laar MA, Adang EMM, Fransen J, et al. Treatment strategies aiming at remission in early rheumatoid arthritis patients: starting with methotrexate monotherapy is cost-effective. *Rheumatology* 2011;50:1320-30.
266. Shepherd J, Jones J, Takeda A, Davidson P, Price A. Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation. *Health Technol Assess* 2006;10(28):iii-iv, xi-xiv, 1-183.
267. Sizto S, Bansback N, Feldman SR, Willian MK, Anis AH. Economic evaluation of systemic therapies for moderate to severe psoriasis. *Br J Dermatol* 2009;160(6):1264-72.
268. Smith KJ, Roberts MS. Sequential medication strategies for postherpetic neuralgia: a cost-effectiveness analysis. *J Pain* 2007;8(5):396-404.
269. Soini EJ, Martikainen JA, Vihervaara V, Mustonen K, Nousiainen T. Economic evaluation of sequential treatments for follicular non-hodgkin lymphoma. *Clin Ther* 2012;34(4):915-25.e2.
270. Tanno M, Nakamura I, Ito K, Tanaka H, Ohta H, Kobayashi M, et al. Modeling and cost-effectiveness analysis of etanercept in adults with rheumatoid arthritis in Japan: a preliminary analysis. *Modern Rheumatology* 2006;16(2):77-84.
271. Tebas P, Henry K, Nease R, Murphy R, Phair J, Powderly WG. Timing of antiretroviral therapy. Use of Markov modeling and decision analysis to evaluate the long-term implications of therapy. *Aids* 2001;15(5):591-9.
272. Tran-Duy A, Boonen A, van de Laar MAFJ, Franke AC, Severens JL. A discrete event modelling framework for simulation of long-term outcomes of sequential treatment strategies for ankylosing spondylitis. *Ann Rheum Dis* 2011;70(12):2111-8.
273. Tran-Duy A, Boonen A, Kievit W, van Riel PL, van de Laar MA, Severens JL. Modelling outcomes of complex treatment strategies following a clinical guideline for treatment decisions in patients with rheumatoid arthritis *Pharmacoeconomics* 2014;32(10):1015 - 28.
274. Wailoo A, Bansback N, Brennan A, Nixon R, Michaud K, Wolfe F. Modeling the Cost Effectiveness of Etanercept, Adalimumab and Anakinra Compared to Infliximab in the Treatment of Patients with Rheumatoid Arthritis in the Medicare Program. *Agency for Healthcare Research and Quality publications. Technology Assessment Program: AHRQ, U.S. Department of Health and Human Services, 2006.*
275. Welsing PMJ, Severens JL, Hartman M, van Riel PLCM, Laan RFJM. Modeling the 5-year cost effectiveness of treatment strategies including tumor necrosis factor-blocking agents and leflunomide for treating rheumatoid arthritis in the Netherlands. *Arthritis & Rheumatology* 2005;51(6):964-73.
276. Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, et al. Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation. *Health Technol Assess* 2005;9(15):1-157, iii-iv.
277. Wong Y-N, Meropol NJ, Speier W, Sargent D, Goldberg RM, Beck JR. Cost implications of new treatments for advanced colorectal cancer. *Cancer* 2009;115(10):2081-91.
278. Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Vergel YB, et al. Etanercept and efalizumab for the treatment of psoriasis: a systematic review. *Health Technology Assessment* 2006;10(46):1-233, i-iv.
279. Wu B, Wilson A, Wang Ff, Wang Sl, Wallace DJ, Weisman MH, et al. Cost Effectiveness of Different Treatment Strategies in the Treatment of Patients with Moderate to Severe Rheumatoid Arthritis in China. *PLoS ONE [Electronic Resource]* 2012;7(10).
280. Chalkidou K. Comparative effectiveness review within the U.K.'s National Institute for Health and Clinical Excellence. *Issue Brief (Commonw Fund)* 2009;59:1-12.
281. Wailoo A, Pickstone C. A Review of the NICE Single Technology Appraisal Process. Report by the Decision Support Unit. Sheffield: NICE Decision Support Unit, SchARR, University of Sheffield, 2008.
282. Adler AI, Shaw EJ, Stokes T, Ruiz F. Newer agents for blood glucose control in type 2 diabetes: summary of NICE guidance. *British Medical Journal* 2009;338:1328-29.

283. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal 2013*. London: NICE, 2013.
284. Tosh J, Trowman R. Briefing paper for methods review working party on treatment sequences and downstream costs. *Briefing paper for the update of the NICE methods guide*: NICE Decision Support Unit, 2011:http://nicedsu.org.uk/wp-content/uploads/2016/03/DSU_TAMethodsGuideReviewSupportingDocuments.pdf.
285. Seymour MT, Maughan TS, Ledermann JA, Topham C, James R, Gwyther SJ, et al. Different strategies of sequential and combination chemotherapy for patients with poor prognosis advanced colorectal cancer (MRC FOCUS): a randomised controlled trial. *Lancet* 2007;370(9582):143-52.
286. Tournigand C, André T, Achille E, Lledo G, Flesh M, Mery-Mignard D, et al. FOLFIRI followed by FOLFOX6 or the reverse sequence in advanced colorectal cancer: a randomized GERCOR study. *Journal of Clinical Oncology* 2004;22(2):229-37.
287. Koopman M, Antonini NF, Douma J, Wals J, Honkoop AH, Erdkamp FL, et al. Sequential versus combination chemotherapy with capecitabine, irinotecan, and oxaliplatin in advanced colorectal cancer (CAIRO): a phase III randomised controlled trial. *Lancet* 2007;370(9582):135-42.
288. Suissa S. Immortal time bias in pharmaco-epidemiology. *American Journal of Epidemiology* 2007;67(4):492-9.
289. Chibaudel B, Bonnetain F, Shi Q, Buyse M, Tournigand C, Sargent DJ, et al. Alternative end points to evaluate a therapeutic strategy in advanced colorectal cancer: evaluation of progression-free survival, duration of disease control, and time to failure of strategy--an Aide et Recherche en Cancerologie Digestive Group Study. *Journal of Clinical Oncology* 2011;29(31):4199-204.
290. Hudson M, Suissa S. Avoiding common pitfalls in the analysis of observational studies of new treatments for rheumatoid arthritis. *Arthritis Care & Research* 2010;62(6):805-10.
291. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, et al. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled Clinical Trials* 2004;25(1):119-42.
292. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *American Journal of Psychiatry* 2006;163(1):28-40.
293. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology* 1989;18(1):269-74.
294. Altman DG, Matthews JN. Statistics notes. Interaction 1: Heterogeneity of effects. *British Medical Journal* 1996;313(7055):486.
295. National Institute for Health and Care Excellence. Single technology appraisal: User guide for company evidence submission template. Process and methods [PMG24] Published date: January 2015 Last updated: April 2017 2017.
296. Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 2002;21:1559-73.
297. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365(9454):176–86
298. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;326(7382):219.
299. Rothman KJ. Curbing type I and type II errors. *European Journal of Epidemiology* 2010;25(4):223-4.
300. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015;351:h5651.
301. Thorlund K, Druyts E, Aviña-Zubieta JA, Wu P, J. ME. Why the findings of published multiple treatment comparison meta-analyses of biologic treatments for rheumatoid arthritis are

- different: an overview of recurrent methodological shortcomings. *Ann Rheum Dis* 2013;72:1524-35
302. Hider SL, Buckley C, Silman AJ, Symmons DP, Bruce IN. Factors influencing response to disease modifying antirheumatic drugs in patients with rheumatoid arthritis. *The Journal of Rheumatology* 2005;32(1):11-16.
 303. Rubbert-Roth A, Finckh A. Treatment options in patients with rheumatoid arthritis failing initial TNF inhibitor therapy: a critical review. *Arthritis Research & Therapy* 2009;11(Suppl 1):S1.
 304. Hyrich KL, Watson KD, Silman AJ, Symmons DP, British Society for Rheumatology Biologics Register. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. *Rheumatology (Oxford)* 2006;45(12):1558-65.
 305. Hetland ML, Christensen IJ, Tarp U, Dreyer L, Hansen A, Hansen IT, et al. Direct comparison of treatment responses, remission rates, and drug adherence in patients with rheumatoid arthritis treated with adalimumab, etanercept, or infliximab: results from eight years of surveillance of clinical practice in the nationwide Danish DANBIO registry. *Arthritis & Rheumatology* 2010;62(1):22-32.
 306. van Vollenhoven RF. Switching between anti-tumour necrosis factors: trying to get a handle on a complex issue. *Ann Rheum Dis* 2007;66(7):849-51.
 307. Emery P. Optimizing outcomes in patients with rheumatoid arthritis and an inadequate response to anti-TNF treatment. *Rheumatology* 2012;51(Suppl 5):v22-v30.
 308. Schmitz S, Adams R, Walsh C. The use of continuous data versus binary data in MTC models: A case study in rheumatoid arthritis. *BMC Medical Research Methodology* 2012;12:167.
 309. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002;55(1):86-94.
 310. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.
 311. Kwan P, Brodie MJ. Early identification of refractory epilepsy. *New England Journal of Medicine* 2000;342(5):314-19.
 312. Tudur Smith C, Marson AG, Chadwick DW, Williamson PR. Multiple treatment comparisons in epilepsy monotherapy trials. *Trials* 2007;8:34.
 313. Efthimiou O, Mavridis D, Debray TP, Samara M, Belger M, Siontis GC, et al. Combining randomized and non-randomized evidence in network meta-analysis. *Statistics in Medicine* 2017;36(8):1210-26.
 314. Panayidou K, Gsteiger S, Egger M, Kilcher G, Carreras M, Efthimiou O, et al. GetReal in mathematical modelling: A review of studies predicting drug effectiveness in the real world. *Research Synthesis Methods* 2016;7(3):264-77.
 315. Ades AE, Madan J, Welton NJ. Indirect and mixed treatment comparisons in arthritis research. *Rheumatology (Oxford)* 2011;50 (Suppl 4):iv5-9.
 316. Dias S, Sutton AJ, Welton NJ, Ades AE. NICE DSU Technical Support Document 3: Heterogeneity: subgroups, meta-regression, bias and bias-adjustment (updated April 2012). In: Decision Support Unit, editor. Sheffield: SchARR, University of Sheffield, 2011.
 317. Govan L, Ades AE, Weir CJ, Welton NJ, Langhorne P. Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. *Statistics in Medicine* 2010;29(12):1340-56.
 318. Mirza RD, Punja S, Vohra S, Guyatt G. The history and development of N-of-1 trials. *Journal of the Royal Society of Medicine* 2017;110(8):330-40.
 319. Pigott T, Noyes J, Umscheid CA, Myers E, Morton SC, Fu R, et al. AHRQ series on complex intervention systematic reviews-paper 5: advanced analytic methods. *Journal of Clinical Epidemiology* 2017;90:37-42.

320. Demeyin WA, Frost J, Ukoumunne OC, Briscoe S, Britten N. N of 1 trials and the optimal individualisation of drug treatments: a systematic review protocol. *Systematic Reviews* 2017;6(1):90.
321. Shadish WR. Analysis and meta-analysis of single-case designs: an introduction. *Journal of School Psychology* 2014;52(2):109-22.
322. Guyatt G. N of 1 randomized trials: a commentary. *Journal of Clinical Epidemiology* 2016;76:4-5.
323. Sibbald B, Roberts R. Understanding controlled trials Crossover trials. *BMJ* 1998;316:1719
324. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Statistics in Medicine* 2002;21(3):371-87.
325. Fisher DJ, Copas AJ, Tierney JF, Parmar MKB. A critical review of methods for the assessment of patient-level interactions in individual participant data meta - analysis of randomized trials, and guidance for practitioners. *Journal of Clinical Epidemiology* 2011;64(9):949-67.
326. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. *PLoS Medicine* 2015;12(7):e1001855.
327. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American Journal of Epidemiology* 2016;183(8):758-64.
328. Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clinical Pharmacology & Therapeutics* 2011;90(6):777-90.
329. Kwan P, Sander JW. The natural history of epilepsy: an epidemiological view. *J Neurol Neurosurg Psychiatry. Journal of Neurology, Neurosurgery, and Psychiatry* 2004;75(10):1376-81.
330. Pal SK, Vogelzang NJ. Sequential treatment strategies and combination therapy regimens in metastatic renal cell carcinoma. *Clinical advances in hematology & oncology* 2013;11(3):146-55.
331. McAlister FA, Laupacis A, Wells GA, Sackett DL. Users' Guides to the Medical Literature: XIX. Applying clinical trial results B. Guidelines for determining whether a drug is exerting (more than) a class effect. *JAMA* 1999;282(14):1371-7.
332. Schmitz S, Adams R, Walsh C. Incorporating data from various trial designs into a mixed treatment comparison model. *Statistics in Medicine* 2013;32(17):2935-49.
333. Faria R, Hernandez Alava M, Manca A, Wailoo AJ. NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. In: Decision Support Unit S, editor. *NICE DSU Technical support*. <http://www.nicedsu.org.uk>, 2015.
334. Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. *Med Decis Making* 2013;33(5):679-91.
335. Mercier F, Claret L, Prins K, Bruno R. A Model-Based Meta-analysis to Compare Efficacy and Tolerability of Tramadol and Tapentadol for the Treatment of Chronic Non-Malignant Pain. *Pain and Therapy* 2014;3(1):31-44.
336. Caldwell DM, Welton NJ. Approaches for synthesising complex mental health interventions in meta-analysis. *Evidence-Based Mental Health* 2016;19(1):16-21.
337. Novielli N, Cooper NJ, Abrams KR, Sutton AJ. Meta-analysis of the accuracy of two diagnostic tests used in combination: application to ddimer and wells score for the diagnosis of deep vein thrombosis. *Value in Health* 2013;16(4):619-28.
338. Guise J-M, Chang C, Viswanathan M, Glick S, Treadwell JR, Umscheid CA, et al. *Systematic Reviews of Complex Multicomponent Health Care Interventions*. Rockville (MD): Agency for Healthcare Research and Quality (US), 2014.

339. Noyes J, Gough D, Lewin S, Mayhew A, Michie S, Pantoja T, et al. A research and development agenda for systematic reviews that ask complex questions about complex interventions. *Journal of Clinical Epidemiology* 2013;66(11):1262-70.
340. Beresniak A, Dupont DM, Becker JC, Merkesdal S. Interest of modelling in rheumatoid arthritis. *Clinical and Experimental Rheumatology* 2012;30(4 Suppl 73):S96-101.
341. Kuntz K, Sainfort F, Butler M, Taylor B, Kulasingam S, Gregory S, et al. Decision and Simulation Modeling in Systematic Reviews. Rockville (MD): gency for Healthcare Research and Quality (US), 2013.
342. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;26(9):781-98.
343. Briggs AH. Handling Uncertainty in Cost-Effectiveness Models. *Pharmacoeconomics* 2000;17(5):479-500.
344. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 2006;15(12):1295-310.
345. Griffin S, Claxton K, Hawkins N, Sculpher M. Probabilistic analysis and computationally expensive models: Necessary and required? *Value in Health* 2006;9(4):244-52.
346. Stevenson M, Tappenden P, Squires H. Methods for handling uncertainty within pharmaceutical funding decisions. *International Journal of Systems Science* 2014;45(1):60-68.
347. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. *Med Decis Making* 2011;31(4):662-74.
348. Bojke L, Claxton K, Culpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health* 2009;12(5):739-49.
349. Strong M, Oakley JE, Chilcott J. Managing structural uncertainty in health economic decision models: a discrepancy approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2012;61(1):25-45.
350. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial-based economic evaluation for health care decision making? *Health Economics* 2006;15(7):677-87.
351. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011;342:d1766.
352. Groot Koerkamp B, T. S, C. WM, G HM. The combined analysis of uncertainty and patient heterogeneity in medical decision models. *Med Decis Making* 2011;31(4):650-61.
353. Scott DL, Strand V. The effects of disease modifying anti-rheumatic drugs on the Health Assessment Questionnaire score. Lessons from the leflunomide clinical trials database. *Rheumatology* 2002;41(8):899-909.
354. Bombardieri S, Ruiz AA, Fardellone P, Geusens P, McKenna F, Unnebrink K, et al. Effectiveness of adalimumab for rheumatoid arthritis in patients with a history of TNF-antagonist therapy in clinical practice. *Rheumatology* 2007;46(7):1191-9.
355. Bingham CO, Ince A, Haraoui B, Keystone EC, Chon Y, Baumgartner S. Effectiveness and safety of etanercept in subjects with RA who have failed infliximab therapy: 16-week, open-label, observational study. *Current Medical Research and Opinion* 2009;25(5):1131-42.
356. Cain LE, Saag MS, Petersen M, May MT, Ingle SM, Logan R, et al. Using observational data to emulate a randomized trial of dynamic treatment-switching strategies: an application to antiretroviral therapy. *International Journal of Epidemiology* 2016;45(6):2038-49.
357. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. *Value in Health* 2010;13(5):557-64.
358. Grigore B, Peters J, Hyde C, Stein K. A comparison of two methods for expert elicitation in health technology assessments. *BMC Medical Research Methodology* 2016;16:85.
359. Drummond MF, Iglesias CP, Cooper NJ. Systematic reviews and economic evaluations conducted for the National Institute for Health and Clinical Excellence in the United Kingdom: a game of two halves? *International Journal of Technology Assessment in Health Care* 2008;24(2):146-50.

360. Kim SY, Goldie SJ. Cost-effectiveness analyses of vaccination programmes: a focused review of modelling approaches. *Pharmacoeconomics* 2008;26(3):191-215.
361. Barton P, Bryan S, Robinson S. Modelling in the economic evaluation of health care: selecting the appropriate approach *Journal of Health Service Research Policy* 2004;9(2):110-8.
362. Bentley TG, Kuntz KM, Ringel JS. Bias associated with failing to incorporate dependence on event history in Markov models. *Med Decis Making* 2010.
363. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health* 2012;15(6):796-803.
364. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? . *Value in Health* 2010;13(8):1056-60.
365. Cooper K, Brailsford SC, Davies R. Choice of modelling technique for evaluating health care interventions. *Journal of the Operational Research Society* 2007;58(2):168-76.
366. Karnon J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics* 2003;12(10):837-48.
367. Karnon J, Haji Ali Afzali H. When to use discrete event simulation (DES) for the economic evaluation of health technologies? A review and critique of the costs and benefits of DES. *Pharmacoeconomics* 2014;32(6):547-58.
368. Karnon J, Stahl J, Brennan A, Caro JJ, Mar J, Möller J, et al. Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Value in Health* 2012;15(6):821-7.
369. O'Mahony JF, Newall AT, van Rosmalen J. Dealing with Time in Health Economic Evaluation: Methodological Issues and Recommendations for Practice. *Pharmacoeconomics* 2015;33(12):1255-68.
370. Roberts M, Russell LB, Paltiel AD, Chambers M, McEwan P, Krahn M, et al. Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-2. *Med Decis Making* 2012;32(5):678-89.
371. Stahl JE. Modelling methods for pharmacoeconomics and health technology assessment: an overview and guide. *Pharmacoeconomics* 2008;26(2):131-48.
372. Siebert U, Alagoz O, Bayoumi AM, Jahn B, Owens DK, Cohen DJ, et al. State-Transition Modeling: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-3. *Value in Health* 2012;15:812-20.
373. Davis S, Stevenson M, Tappenden P, Wailoo AJ. *NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation*. Sheffield: NICE Decision Support Unit, ScHARR, 2014.
374. Standfield L, Comans T, Scuffham P. Markov modeling and discrete event simulation in health care: a systematic comparison. *International Journal of Technology Assessment in Health Care* 2014;30(2):165-72.
375. Cao Q, Buskens E, Feenstra T, Jaarsma T, Hillege H, Postmus D. Continuous-Time Semi-Markov Models in Health Economic Decision Making: An Illustrative Example in Heart Failure Disease Management. *Med Decis Making* 2016;36(1):59-71.
376. Gray A, Clarke P, Wolstenholme J, Wordsworth S. Chapter 8. Decision analytic modelling; decision trees. In: Gray A, Clarke P, Wolstenholme J, Wordsworth S, editors. *Applied methods of cost-effectiveness analysis in health care*. Oxford: University Press, 2010.
377. Harousseau JL, Palumbo A, Richardson PG, Schlag R, Dimopoulos MA, Shpilberg O, et al. Superior outcomes associated with complete response in newly diagnosed multiple myeloma patients treated with nonintensive therapy: analysis of the phase 3 VISTA study of bortezomib plus melphalan-prednisone versus melphalan-prednisone. *Blood* 2010;116(19):3743-50.
378. Niesvizky R, Richardson PG, Rajkumar SV, Coleman M, Rosiñol L, Sonneveld P, et al. The relationship between quality of response and clinical benefit for patients treated on the

- bortezomib arm of the international, randomized, phase 3 APEX trial in relapsed multiple myeloma. *British Journal of Haematology* 2008;143(1):46-53.
379. Richardson PG, Barlogie B, Berenson J, Singhal S, Jagannath S, Irwin DH, et al. Extended follow-up of a phase II trial in relapsed, refractory multiple myeloma: final time-to-event results from the SUMMIT trial. *Cancer Letters* 2006;106(6):1316-9.
380. Hawkins N, Epstein D, Drummond M, Wilby J, Kainth A, Chadwick DW, et al. Assessing the cost-effectiveness of new pharmaceuticals in epilepsy in adults: the results of a probabilistic decision model. *Med Decis Making* 2005;25(5):493-510.
381. Shah D, Khan N, Hawkins N, Briggs A. PRM62 Time-Dependency for Treatment Sequences: A Case-Example in Epilepsy. *Value in Health* 2012;15(7):A471.
382. Tosh J, Brennan A, Wailoo A, Bansback N. The Sheffield rheumatoid arthritis health economic model. *Rheumatology* 2011;50(4):26-31.
383. Wailoo AJ, Bansback N, Brennan A, Michaud K, Nixon RM, Wolfe F. Biologic drugs for rheumatoid arthritis in the Medicare program: a cost-effectiveness analysis. *Arthritis & Rheumatology* 2008;58(4):939-46.
384. Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E, et al. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet* 2008;371(9617):987-97.
385. Genovese MC, McKay JD, Nasonov EL, Mysler EF, da Silva NA, Alecock E, et al. Interleukin-6 receptor inhibition with tocilizumab reduces disease activity in rheumatoid arthritis with inadequate response to disease-modifying antirheumatic drugs: the tocilizumab in combination with traditional disease-modifying antirheumatic drug therapy study. *Arthritis & Rheumatology* 2008;58(10):2968-80.
386. Fleischmann R, Burgos-Vargas R, Ambs P, Alecock E, Kremer J. LITHE: tocilizumab inhibits radiographic progression and improves physical function in rheumatoid arthritis (RA) patients (Pts) at 2 yrs with increasing clinical efficacy over time. *Arthritis & Rheumatology* 2009;60((10 Suppl)):637.
387. Hyrich KL, Deighton C, Watson KD, BSRBR Control Centre Consortium, Symmons DPM, Mark Lunt M. Benefit of anti-TNF therapy in rheumatoid arthritis patients with moderate disease activity. *Rheumatology* 2009;48(10):1323-27.
388. Barton P. Development of the Birmingham Rheumatoid Arthritis Model: past, present and future plans. *Rheumatology (Oxford)* 2011;50 (Suppl 4):iv32-iv38.
389. Barton P. Further cost-effectiveness analysis of sequential TNF inhibitors for rheumatoid arthritis patients. Report to NICE, 2008.
390. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, et al. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis & Rheumatology* 2006;54(9):2793-806.
391. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993;13(4):322-38.
392. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models *Pharmacoeconomics* 2000;17(5):461-77.
393. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the economic evaluation of health care programmes*. Third ed. Oxford, UK, 2005.
394. Briggs A, Sculpher M. An Introduction to Markov Modelling for Economic Evaluation. *Pharmacoeconomics* 1998;13(4):397-409.
395. Woods B, Sideris E, Palmer S, Latimer N, Soares M. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review [Available from <http://www.nicedsu.org.uk>], 2017.

396. Chrosny W, Stevenson M, Munzer A. Comparison of Markov and discrete event simulation modeling techniques with application to cost effectiveness analyses [Abstract PRM70]. *Value in Health* 2013;16(7):A587.
397. Stevenson M, Archer R, Tosh J, Simpson E, Everson-Hock E, Stevens J, et al. Adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and abatacept for the treatment of rheumatoid arthritis not previously treated with disease-modifying antirheumatic drugs and after the failure of conventional disease-modifying antirheumatic drugs only: systematic review and economic evaluation. *Health Technology Assessment* 2016;20(35):1-610.
398. Caro JJ. Discretely Integrated Condition Event (DICE) Simulation for Pharmacoeconomics. *Pharmacoeconomics* 2016;34(7):665-72.
399. Tappenden P. Conceptual modelling for health economic model development In: HEDS Discussion Paper 12/05. <http://eprints.whiterose.ac.uk/74464/>. Accessed 12 Feb 2017. 2012.
400. Squires H, Chilcott J, Akehurst R, Burr J, Kelly MP. A Framework for Developing the Structure of Public Health Economic Models. *Value in Health* 2016;19(5):588-601.
401. Gerhardus A, Oortwijn W, Van der Wilt G. How to avoid giving the right answers to the wrong questions: the need for integrated assessments of complex health technologies. *International Journal of Technology Assessment in Health Care* 2017;33(5):541-43.
402. Wahlster P, Brereton L, Burns J, Hofmann B, Mozygemba K, Oortwijn W, et al. Guidance on the integrated assessment of complex health technologies - The INTEGRATE-HTA Model [Online]. Available from: <http://www.integrate-hta.eu/downloads/>, 2016.
403. Gerhardus A, Oortwijn W, van der Wilt GJ. How to avoid giving the right answers to the wrong questions: the need for integrated assessments of complex health technologies. *International Journal of Technology Assessment in Health Care* 2017;33(5):541-43.
404. Rohwer A, Booth A, Pfadenhauer L, Brereton L, Gerhardus A, Mozygemba K, et al. Guidance On The use of logic models in health technology assessments of complex interventions [Online]. Available from: <http://www.integrate-hta.eu/downloads/> 2016.
405. Chhatwal J, He T. Economic evaluations with agent-based modelling: an introduction. *Pharmacoeconomics* 2015;33(5):423-33.
406. Briggs AD, Wolstenholme J, Blakely T, Scarborough P. Choosing an epidemiological model structure for the economic evaluation of non-communicable disease public health interventions. *Population Health Metrics* 2016;14:17.
407. National Institute for Health and Care Excellence. *Prostate cancer diagnosis and treatment: NICE clinical guideline 58*. London: NICE, 2008.
408. National Institute for Health and Care Excellence. *Atrial fibrillation: the management of atrial fibrillation: NICE clinical guideline 36. Quick reference guideline*. London: NICE, 2006.
409. Owens DK, Shachter RD, Nease R, F. Jr. Representation and analysis of medical decision problems with influence diagrams. *Med Decis Making* 1997;17(3):241-62.
410. Díez FJ, Yebra M, Bermejo I, Palacios-Alonso MA, Calleja MA, Luque M, et al. Markov Influence Diagrams. *Med Decis Making* 2017;37(2):183-95.
411. Collins English Dictionary. *Collins English Dictionary*. Glasgow: HarperCollins Publishers, 2017.
412. Messner DA. Evaluating the comparative effectiveness of treatment sequences in oncology: a new approach. *Journal of Comparative Effectiveness Research* 2015;4(6):537-54.
413. Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB, et al. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Med Decis Making* 2012;32(5):33-43.
414. Dear RF, McGeechan K, Jenkins MC, Barratt A, Tattersall MHN, Wilcken N. Combination versus sequential single agent chemotherapy for metastatic breast cancer. *Cochrane Database of Systematic Reviews* 2012;12(CD008792).
415. Almirall D, Compton SN, Gunlicks-Stoessel M, Duan N, Murphy SA. Designing a pilot sequential multiple assignment randomized trial for developing an adaptive treatment strategy. *Statistics in Medicine* 2012;31(17):1887-902.

416. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. *American Journal of Preventive Medicine* 2007;32(5 Suppl):S112-8.
417. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. *Annual Review of Clinical Psychology* 2012;8 21-48.
418. Nahum-Shani I, Ertefaie A, Lu XL, Lynch KG, McKay JR, Oslin DW, et al. A SMART data analysis method for constructing adaptive treatment strategies for substance use disorders. *Addiction* 2017;112(5):901-09.
419. Thorlund K, Haggstrom J, Park JJH, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians *BMJ* 2018;360:k698.
420. Korn EL, Freidlin B. Adaptive Clinical Trials: Advantages and Disadvantages of Various Adaptive Design Elements. *Journal of the National Cancer Institute* 2017;109(6).
421. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *New England Journal of Medicine* 2017;377(1):62-70.
422. Brixner DI, Watkins JB. Can CER be an effective tool for change in the development and assessment of new drugs and technologies? *Journal of managed care pharmacy* 2012;18(5 Supp A):S06-11.
423. Tafuri G, Pagnini M, Moseley J, Massari M, Petavy F, Behring A, et al. How aligned are the perspectives of EU regulators and HTA bodies? A comparative analysis of regulatory-HTA parallel scientific advice. *British journal of clinical pharmacology* 2016;82(4):965-73.
424. Berger ML, Sox H, Willke RJ, Brixner DL, Eichler HG, Goettsch W, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Value in Health* 2017;20(8):1003-08.
425. Silman A, Symmons D, Scott DG, Griffiths I. British Society for Rheumatology Biologics Register. *Ann Rheum Dis* 2003;62 (Suppl 2):ii28-9.
426. Azuaje F. Computational models for predicting drug responses in cancer research. *Briefings in Bioinformatics* 2017;18(5):820–29.
427. Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. *European Journal of Epidemiology* 2017;32(6):473-75.