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

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

By Ruth Ann Lewis

APPENDIX

The Appendix for the thesis is presented as two separate volumes, which serve different purposes.

Volume I (pg 314-383) provides supplementary information to the main text, for example, a summary of the hierarchy of evidence, a description of relevant clinical trial designs, and an overview of clinical conditions.

Volume II (pg 384-494) provides supplementary data for individual chapters presenting the findings of separate methodological reviews, and includes information such as codes used for conducting the meta-analyses and data extraction tables.

APPENDIX VOLUME I

TABLE OF CONTENTS FOR APPENDIX VOLUME I

TABLES AND FIGURES FOR APPENDIX VOLUME I

List of Figures:

List of Tables:

A. THE IMPACT OF STUDY DESIGN ON INTERNAL AND EXTERNAL VALIDITY

A1 Hierarchy of evidence

A number of hierarchies of evidence have been developed to enable different research methods to be ranked according to the validity of their findings.[1 2] These are described as a pyramid, which shows the RCT design, or the meta-analysis of RCTs, at the very top.[2] Most versions of the pyramid represent a hierarchy based on internal validity, or the risk of bias.[2] However, it is also recognised that study design alone is insufficient as a surrogate for the risk of bias.[3] Certain methodological limitations of a study, imprecision, inconsistency and indirectness can also affect the quality of evidence derived from any study design.[2 3]

This next section explores in more detail the issues relating to the selection of study design within evidence synthesis, and the assessment of the body of evidence to inform decision making.

A2 Evidence synthesis of clinical effectiveness

The underlying principles of a meta-analysis, which synthesises the findings of multiple studies, is that it provides greater power to detect a statistically significant difference, and will give a more precise estimate of the true treatment effect in a particular population. However, this is based on the assumption that there is no bias (systematic error) in the summary estimates provided by the included studies. It also assumes that the summary estimates are similar enough for it to make sense to pool the data. The importance of assessing the presence and extent of heterogeneity is discussed in Chapter 2. This next section focuses on the impact of study design on internal and external validity.

It is important that the meta-analysis includes all relevant studies, which should ideally be identified using a systematic search of both published and unpublished literature. This is important to mitigate bias, which can occur if there is a systematic difference between the set of studies conducted, or the outcome measures assessed, and those included in the meta-analysis.[4 5]

A3 The optimal study design for internal validity and for inferring causality

The RCT study design provides the least biased estimate of the treatment effect, and is considered to be the gold standard for assessing the causal effect of an intervention.[6]

Internal validity refers to whether differences in the observed effects between the intervention and control can be attributed to differences in the intervention.[7] The observed effect of an intervention can be due to a number factors. These include the following:

i. The intervention itself
ii. Extraneous factors, such as lifestyle, use of other medication, placebo effect etc.
iii. Information errors, such as incorrect assessment or reporting of outcomes
iv. The natural course of disease, incorporating variability in disease status and the influence different prognostic factors

v. Chance

Randomisation ensures that any (un)known or (un)measured prognostic factors will be the same in both the intervention and the control group.[8] This means that, providing the RCT is well conducted and large enough, the resulting difference in treatment effects between the intervention and control groups can be attributable to the intervention itself. However, the casual inferences from RCTs can be undermined by flaws in the design, conduct, analysis, and reporting leading to a biased estimate of effect.[9] [10] For example, empirical evidence suggests that lack of blinding of participants, outcome assessors, or double blinding (where both participants and personnel/assessors are blinded) in RCTs are associated with exaggerated intervention effect estimates.[10]

By contrast, in an observational study, pre-existing groups of patients are compared that have either used the 'intervention' or the alternative 'control' treatment in line with the course of usual treatment decisions.[11] Here the observed effect cannot be attributed solely to the treatment used as groups may differ in various ways, for example, disease duration or severity. The effect of these factors (and the intervention) can also be confounded by other factors, for example, a higher dose may have been used for patients who were sicker or had a more sever condition, and a lower dose for those with a less severe condition. The logic of confounding (mixing different effects) means that the use of a control or comparison group is key to dealing with the attribution of treatment effect, but is still not sufficient due to selection bias (confounding by indication).[12] [13] Where treatment effects are derived from uncontrolled studies, using for example cross-sectional or before and after data, it will not be possible to disentangle the treatment effect from other effects.[12] I re-visit the type of bias in observational studies that is specific to the evaluation of treatment sequences in Chapter 5 and selection bias in Section 5.5.2 and 5.8.10.

The RCT, or meta-analysis of RCTs, is considered to be at the top of the hierarchy of evidence where study designs are ranked based on internal validity.[14] However, this does not account for the risk of bias within the studies, and the ranking of a poor quality RCT compared to a large well conducted cohort study is unclear.[15] The issue regarding the difficulty of distinguishing between poorly conducted RCTs and well-conducted non-randomised studies, and the impact of this on deciding to limit inclusion to RCTs was also identified in the network meta-analysis of sciatica treatments (Chapter 2, Section 2.6.3).

It is important to note here, that a meta-analysis is conducted by first developing a comparative effect estimate for each individual study, and then pooling across these, and not the arm-based data.[16] This means that each study maintains its own control, or more specifically, the benefit of randomisation. However, the protection between the treatment and outcome relationship provided by randomisation does not hold in analyses that go beyond developing an overall treatment effect, for example meta-regression or subgroup analysis.[11] Here differences in treatment effects may be due to variation in study or patient characteristics, such as disease-severity of participants recruited to different

studies.[11] In other words, the relationship cannot imply causality. In a similar way, the protective effect of randomisation will not hold when making inferences about whole sequences based on a series of RCTs that consider treatments used at a single point in the pathway. Of note here also is that the value of randomisation does not hold across RCTs within a meta-analysis, and it is only the within study randomisation that is maintained.[17] It is therefore important that the studies are similar for it to make sense to pool the data, which is disused in more detail in Chapter 2, Section 2.9.1, as important differences may exist in populations across studies.

A4 The optimal study design for external validity

External validity refers to the extent to which the findings of the study are generalisable to all potential recipients.[7] One of the main limitations of RCTs is that they have poor external validity, limiting the applicability of the causal inferences to 'real' world settings.[18] As described in Chapter 1 (Section 1.3), traditional RCTs tend to evaluate single treatments, use a narrow participant eligibility criteria excluding high risk groups, administer the interventions in a prescriptive manner according to a tight protocol, and may use short-term follow-up. When the 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. The impact of this in terms of the analysis and reporting is generally limited to the trial interventions, with outcomes for subsequent treatments being poorly reported. Loss of follow-up can be controlled for using intention to treat analysis, but this usually leads to the dilution of treatment effect and can mask the detection of harm. The analysis will be especially challenging where there is a large number of drop-outs, or the participants have either switched trial interventions, or gone on to try other treatments. A number of methodological studies have been conducted to discover how best to adjust for the cross-ever effect when analysing the data from RCTs of single treatments,[19] 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 are sometimes subject to the same confounding and biases as observational studies due to the non-adherence and loss to follow-up. Another limitation of RCTs is that they are often underpowered for evaluating safety endpoints.[20 21]. Long-term safety issues, as noted in Chapter 1 (Section 1.3.3), are frequently assessed as part of single arm phase IV trials, also known as open label extension studies.[22]

Pragmatic RCTs that are designed to provide greater external validity are better at reflecting the way interventions are used in practice but, for logistic reasons, do not generally include blinding of participants and care givers.[23] Two other RTC designs that reflect the dynamic nature of treatment regimens used in clinical practice are the N-of-1 RCT and the sequential multiple assignment randomised trial (SMART). The N-of-1 trial was introduced in Chapter 1 (Section 1.7) as a study design that can be used to inform personalised decision making. The SMART design, which was developed to construct and compare dynamic treatment regimens (also referred to as adaptive treatment regimens),[24] is described in more detail in Appendix Volume I, Section B, as it was identified as part of the review of methods. However, the findings of the SMART design are

considered exploratory rather than confirmatory in nature (Section B2),[24] which is discussed in more detail in Chapter 5 (Section 5.8.7 and 5.9.9). I also re-visit the N-of-1 trial design in Chapter 5 (Section 5.9.9), which includes a summary of its limitations for informing treatment sequencing effects.

The advantages of observational studies over RCTs are: they can provide evidence that better reflects real-world practice;[25 26] the results are available more rapidly; they incorporate longer follow-up, and provide findings that are generalisable to much wider group of participants;[25 26] they are less expensive to conduct than RCTs and usually include a much larger sample size.[25 26] For some clinical scenarios, non-randomised studies may be the only source of available data to inform the sequencing effects of certain treatments.[27] The validity and quality of the routinely collected data is key here though. The increasing popularity of analysing 'big data', and the use of linked databases, means the availability of this type of data is likely to improve.[28 29] However, selection bias or confounding caused by an imbalance in prognostic factors between the intervention and control group, is always a concern even in the most rigorously conducted observational studies.[25 28]

A5 Assessment of the risk of bias within individual studies

The assessment of the risk of bias in included studies is a central component of systematic reviews and a key feature of a credible evidence synthesis.[30 31] The risk of bias reflects the likelihood of inaccuracy in the estimate of causal effect in a particular study.[30] Recent recommendations for assessing the risk of bias in systematic reviews of health care interventions are provided by Viswanathan *et al*. (2018).[30]

Non-randomised studies, which include observational studies (e.g. based on patient registries) or non-controlled trials (e.g. single arm phase IV trials) are likely to provide a critical source of evidence for treatment sequencing. However, as previously discussed, their findings are subject to confounding and a range of other biases. The evaluation of the risk of bias in these studies is essential. A variety of methods, such as matching, stratification, regression, and propensity score, have been developed to mitigate the risk of bias at the study design, analysis and interpretation stages of non-randomised studies.[13] The extent to which these have been implemented also needs to be assessed.

Two recently developed tools for assessing the risk of bias are the revised Cochrane Collaboration's Risk of Bias in RCTs (RoB 2.0),[32] and the Risk Of Bias In Non-randomised Studies - of Interventions (ROBINS-I).[33] The ROBINS-I is specifically concerned with evaluating the risk of bias in estimates of the effectiveness or safety (benefit or harm) of an intervention from studies that did not use randomisation to allocate interventions.[33] The tool views each study as an attempt to emulate (mimic) a hypothetical pragmatic randomised trial, and covers seven distinct domains through which bias might be introduced. The domains covered by the tool include:

i. Bias due to confounding

ii. Bias in selection of participants into the study
iii. Bias in classification of interventions
iv. Bias due to departures from intended interventions
v. Bias due to missing data
vi. Bias in measurement of outcomes
vii. Bias in selection of reported results

The ROBINS-I also includes an optional judgment about the direction of the bias for each domain.

There are several unresolved issues in assessing the risk of bias in primary studies for inclusion in meta-analysis. There is currently very little guidance on how to make judgments on the direction of the bias (ie, which of the interventions being compared is the bias predicted to favour).[34] The issue of whether and how to take account of the risk of bias in the meta-analysis is also an issue of ongoing research.[34]

A6 Assessment of the credibility of a body of evidence

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group has developed a widely accepted approach to rating the certainty of a body of evidence (also known as 'quality of evidence' or 'confidence in evidence') in the contexts of systematic reviews, health technology assessments, and clinical guidelines.[3] The GRADE system has been adopted and used by NICE, the Cochrane Collaboration, and the World Health Organization.[35]

According to current GRADE guidance for interventions, the process of rating a body of evidence begins by classifying the design of the relevant studies.[36 37] If the relevant studies are RCTs, the body of evidence begins as a 'high' certainty rating, but if they are non-randomised studies, the body of evidence begins as 'low' certainty. This initial rating is followed by consideration of eight domains, five of which may result in rating down certainty, and three in rating up.[36] The three domains by which non-randomised studies can be rated up include: large effects, dose-effect relations, and when plausible residual confounders or other biases increase certainty.[37] Since the development of ROBINS-I the GRADE working group now consider that the initial assessment should start with assessing the body of evidence using ROBINS-I. This means that the initial GRADE certainty rating from a body of non-randomised studies is based on the assessment of selection bias and confounding, which is undertaken as an integral part of the ROBINS-I tool, rather than starting with an initial description of the underlying study design.[37]

B RANDOMISED CONTROLLED TRIALS OF TREATMENT SEQUENCES

The review of quantitative evidence synthesis methods identified a small number of reviews that included RCTs of treatment sequences. The type of study design used in these studies is explored in more detail here.

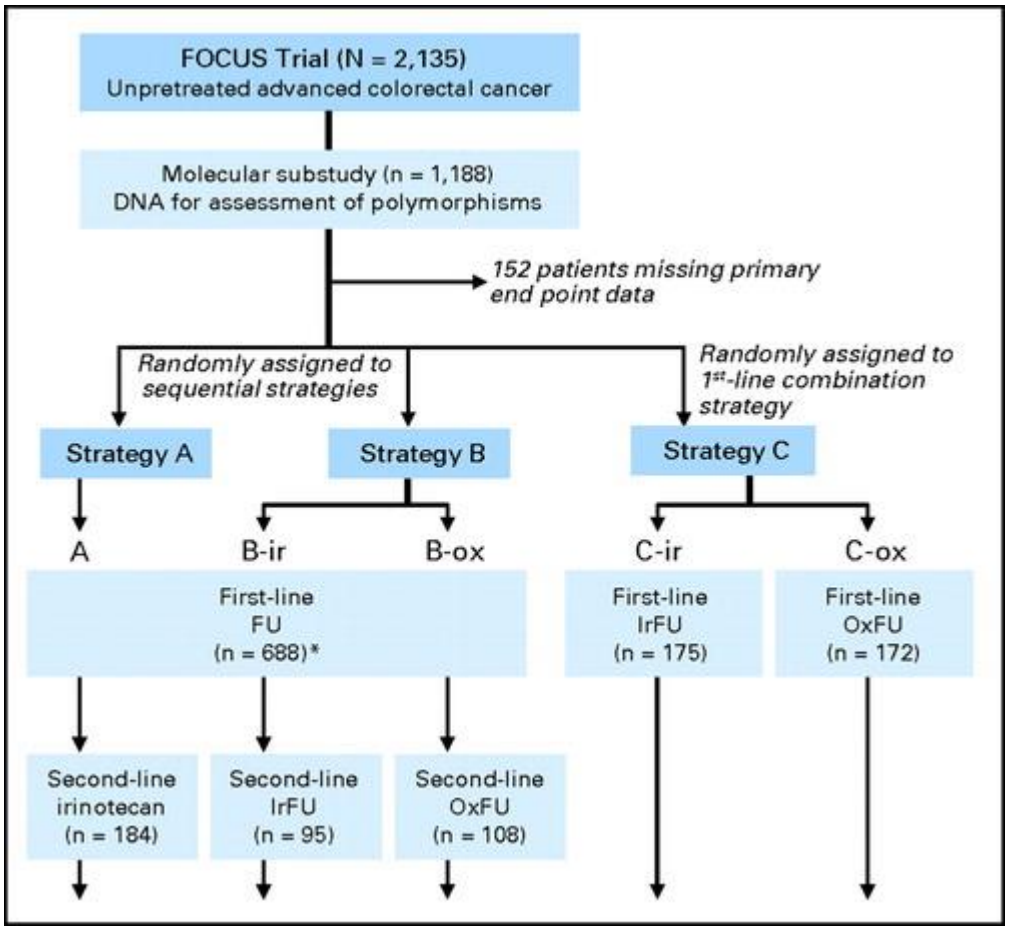
Two types of RCTs were identified for evaluating treatment sequences, depending on whether they aimed to compare the effectiveness of established treatment sequences, or develop the optimal sequence. In other words, the treatment sequences are either predefined (static) or dynamic.

B1 Randomised controlled trials of predefined sequences.

The RCTs comparing predefined, or fully formed treatment sequences tended to be pragmatic open label trials.[38 39] Here, the intervention, which is a specific sequence of treatments, is considered as static. It is assumed that variables that modify the timing and choice of subsequent treatments are considered separately to the 'individual treatment' components within a sequence.[40] However, the findings from this type of trial are considered as confirmatory, in that they can provide evidence that a particular treatment sequence is better than control.

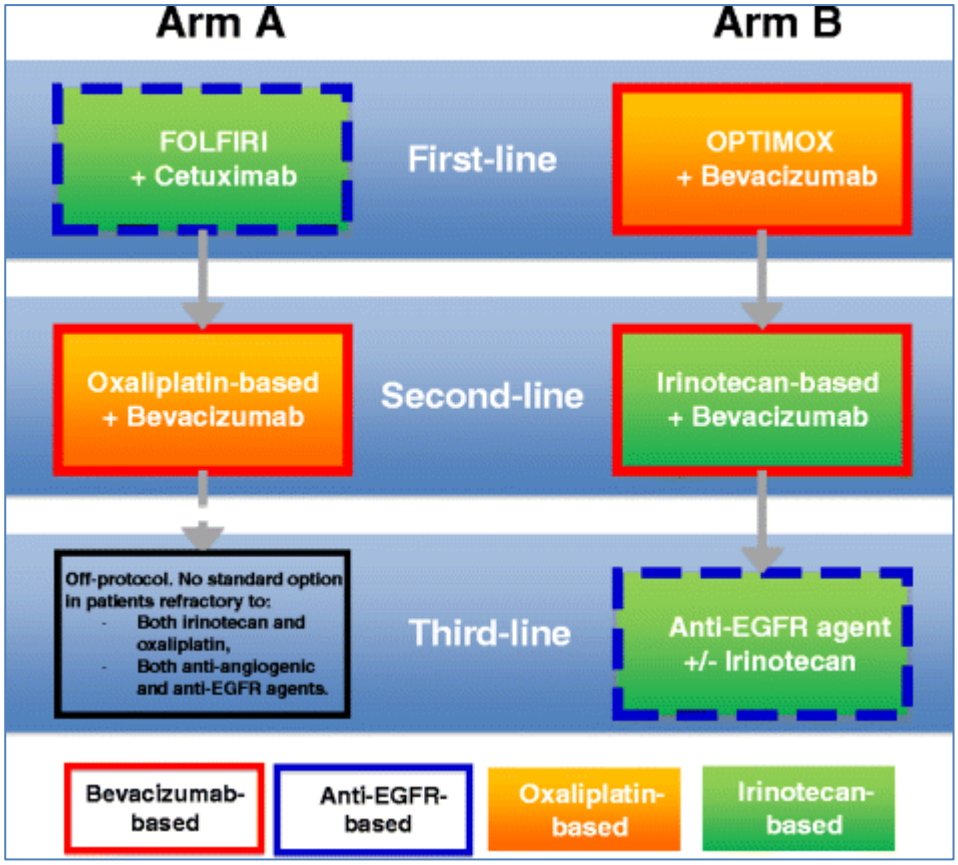
An example of an RCT of predefined sequences is the FOCUS (Fluorouracil, Oxaliplatin, and CPT11-Use and Sequencing) trial, which compared sequential and combination chemotherapy strategies in patients with unpretreated advanced or metastatic colorectal cancer.[38] Patients were randomised to one of three treatment strategies, which is shown in Figure B1. Strategies A and B involved planned sequenced therapy in which first-line therapy was fluorouracil/leucovorin (FU) alone, with irinotecan and oxaliplatin reserved for later, and Strategy C included first-line combination chemotherapy with FU plus either irinotecan or oxaliplatin.[38 41] Other examples of RCTs of predefined sequences for metastatic or colorectal cancer include the CAIRO (CApecitabine, IRinotecan, and Oxaliplatin in advanced colorectal cancer) trial,[42] the randomised GRECOR trial (C97-1 trial conducted by the Groupe Coopérateur Multidisciplinaire en Oncologie)[39] and the multi-line therapy trial in unresectable metastatic colorectal cancer, STRATEGIC-1. Figure B2 provides a schematic of STRATEGIC-1, which was also undertaken by the GRECOR group, and shows how that there is no randomisation beyond the allocation to initial treatment in this type of multiple-line RCT.

Figure B1: Schematic a randomised controlled trial of predefined sequences: FOCUS Trial



Taken from: Braun, M.S., Richman, S.D. et al. (2009). Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. Journal of Clinical Oncology, 27(33), 5519-28.
Note: Random assignments were equal among the three strategies (A, B, and C) with B and C equally subdivided to irinotecan (Ir) and oxaliplatin (Ox) groups, giving five arms in a 2:1:1:1:1 ratio.
Abbreviations: FU Fluorouracil; Ir irinotecan; Ox oxaliplatin (Ox)

Figure B2: Schematic a randomised controlled trial of predefined sequences: STRATEGIC-1 Trial



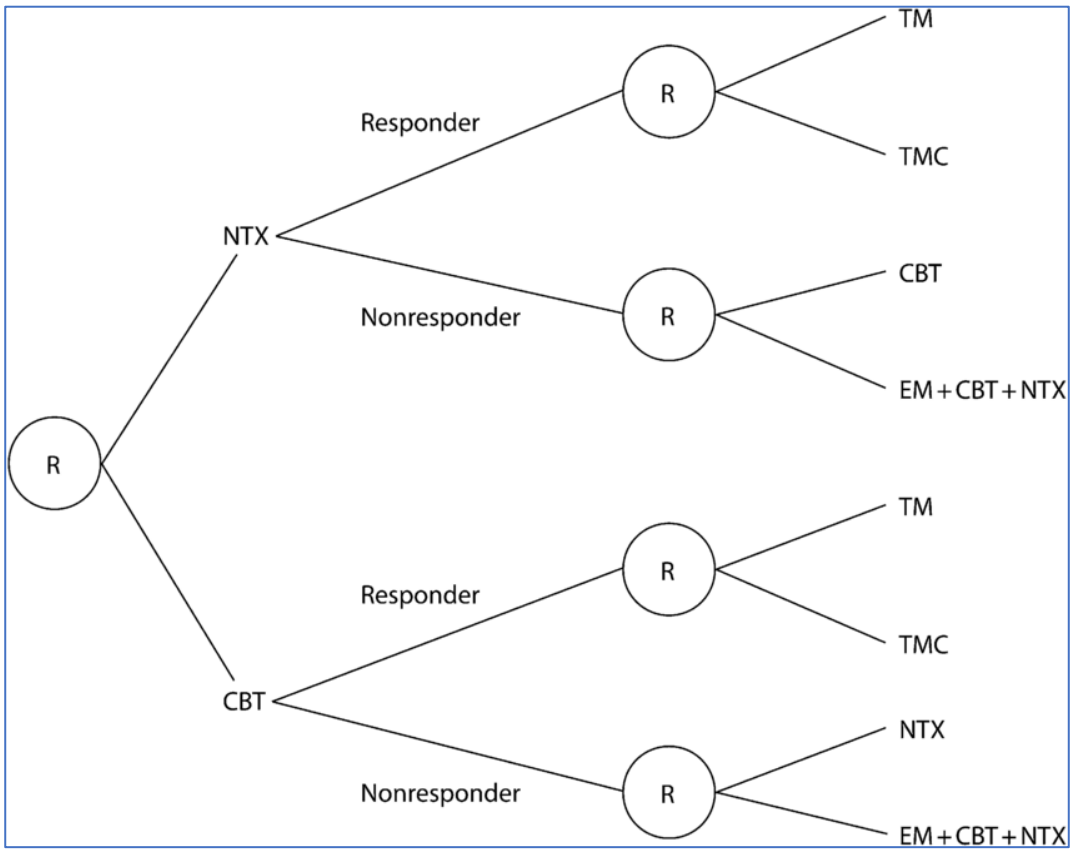
Taken from: Chibaudel, B., Bonnetain, F. et al. (2015). STRATEGIC-1: A multiple-lines, randomized, open-label GERCOR phase III study in patients with unresectable wild-type RAS metastatic colorectal cancer. BMC Cancer, 15: 496.
Abbreviations: EGFR Epidermal Growth Factor Receptor

B2 Randomised controlled trials of adaptive treatments

An alternative type of RCT, which can account for more complicated treatment sequencing is a sequential multiple assignment randomised trial (SMART). This is a relatively new and innovative trial design created to inform the development and optimisation of time-varying adaptive or dynamic treatment regimens.[43-45] Adaptive treatment regimens are individually tailored sequences of treatments based on a series of decision rules that specify how the intensity or type of treatment should change depending on the patient's needs.[46] SMART designs are factorial experimental design used to empirically identify the best tailoring variables and decision rules for an adaptive intervention.[44] They involve a number of intervention stages which correspond to the critical decisions within the adaptive intervention. At the end of each stage participants are randomised to different intervention options.[45] Data resulting from SMART designs can be analysed to obtain information on which treatment is most effective at each stage (including the initial stage), the interactive effects, and the optimal sequence.[47] However, SMART still needs to be followed by a randomised confirmatory trial, as it is designed to develop adaptive interventions rather than confirming that a particular adaptive treatment is better than control.[45]

A schematic of the SMART design is shown in Figure B3 for addiction management. This follows on from the example provided in Chapter 1, Section 1.2.2, on the use of naltrexone or cognitive-behavioural therapy as the initial treatment for alcohol dependency, followed by telephone monitoring or telephone monitoring plus counselling.

Figure B3: Schematic of a sequential multiple assignment randomised trial (SMART) design of addiction management



Taken from: Chakraborty, B. (2011) Dynamic treatment regimes for managing chronic health conditions: a statistical perspective. American Journal of Public Health, 101(1), 40-45.
Abbreviations: CBT cognitive–behavioural therapy; EM enhanced motivation; NTX naltrexone; R randomisation; TM telephone monitoring; TMC telephone counselling and monitoring.

The analysis of the data from a time-varying SMART design is also challenging and not straightforward.[40] They are generally analysed using Q-learning, which is a multi-stage regression approach, but the methods used are continually being developed.[40 48] There is also growing interest in developing methods that go beyond Q-learning to identify new ways to tailor treatments.[48] The synthesis of data from multiple SMART studies will likely be even more complicated.

Examples of the use of the SMART design (or its precursors) include CATIE (Clinical Antipsychotic Trials of Intervention Effectiveness) for treatment of Alzheimer's disease,[49] the STAR*D (Sequenced Treatment Alternatives to Relieve Depression) trial,[50 51] 2-stage cancer trials,[52-54] and the ExTENd (Extending Treatment Effectiveness of Naltrexone) trial to reduce alcohol consumption in alcohol dependent individuals.[45 48]

The SMART design used in the cancer 2-stage treatment strategy trials[55] has also been referred to as two-stage randomisation design (TSRD).[54] This type of two-stage randomisation design is becoming increasingly common in the evaluation of maintenance treatment, where patients are initially randomised to an induction treatment, followed by randomisation to a maintenance treatment conditional on their induction response and consent to further study treatment.[56] However, there is also another type of two-staged trial design commonly used in oncology.[57] This is where a small group of patients are enrolled in a first stage, and then, depending on the outcome of the first stage, another group of patients enrolled in a second stage.[58] This design is usually used within single arm Phase II trials, which are generally used to screen out treatments that are ineffective and select active treatments for future studies. [58] This type of design allows the trial to be stopped early for efficacy, futility, toxicity or poor accrual, and represents an example of an adaptive trial design.[59]

B3 Adaptive clinical trial designs

The term 'adaptive' as used in a SMART design refers to the dynamic nature of the intervention which not only allows for the use of multiple treatments over time, but also accounts for the fact that they are inextricably coupled with the entire system for assigning the treatments.[45 60] This should not however, be confused with the adaptive trial design, which aims to improve the efficacy and flexibility of clinical trials by allowing or even enforcing continual modifications in the design while data are collected.[61] The 'adaptive trial design', unlike the SMART that aims to evaluate 'adaptive treatment strategies', is generally used to provide the best estimate of the effectiveness of discrete treatments and concerned with single staged decision making.[45 62] A brief summary of adaptive trial designs is provided here in order to clarify and explore this further. However, the development of adaptive trial designs were born from the recognition that the exclusive use of 'one indication at a time' approach will not be sustainable,[63] and as such could also incorporate the evaluation of multiple treatment lines (similar to the 2-staged treatment strategies discussed above in Section B2).

The adaptive trial design is attractive as it has the potential to reduce resource use, decrease time to trial completion, limit allocation of participants to inferior interventions, and improve the likelihood that trial results will be scientifically or clinically relevant.[61] Common adaptations made during the implementation of adaptive designs, based on interim analysis include:[61 64]

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

The adaptive trial design has become especially appealing in the development and evaluation of biologically directed therapies, also referred to as personalised medicine, which was described in Chapter 1, Section 1.7. They have great potential for efficiently identifying patients who will be helped

the most by specific treatments.[65] Their appeal also lies in the fact that they can address a number of research questions or hypotheses within the same study.[66] They are able to simultaneously evaluate both multiple experimental treatments and biomarkers, potentially within a diverse patient population with more than one patient type or disease.[64 67]

The biomarker adaptive trial is increasingly used for the evaluation of new targeted therapies for cancer.[68 69] The successful development of new drugs with a companion diagnostic-based genomic alteration of an oncogene has also led to re-thinking of all phases in clinical trial development of cancer drugs.[70] The heterogenetic nature of tumours of the same site and stage means that under a traditional phase III trial design, only a small treatment effect is identified in large-scale trials and many patients are actually treated with non-effective yet toxic and expensive therapies.[68] The use of biologically targeted therapies for cancer is described in more detail in Appendix Volume I, Section C, whilst this section focuses on the evolution of trial designs for these treatments.

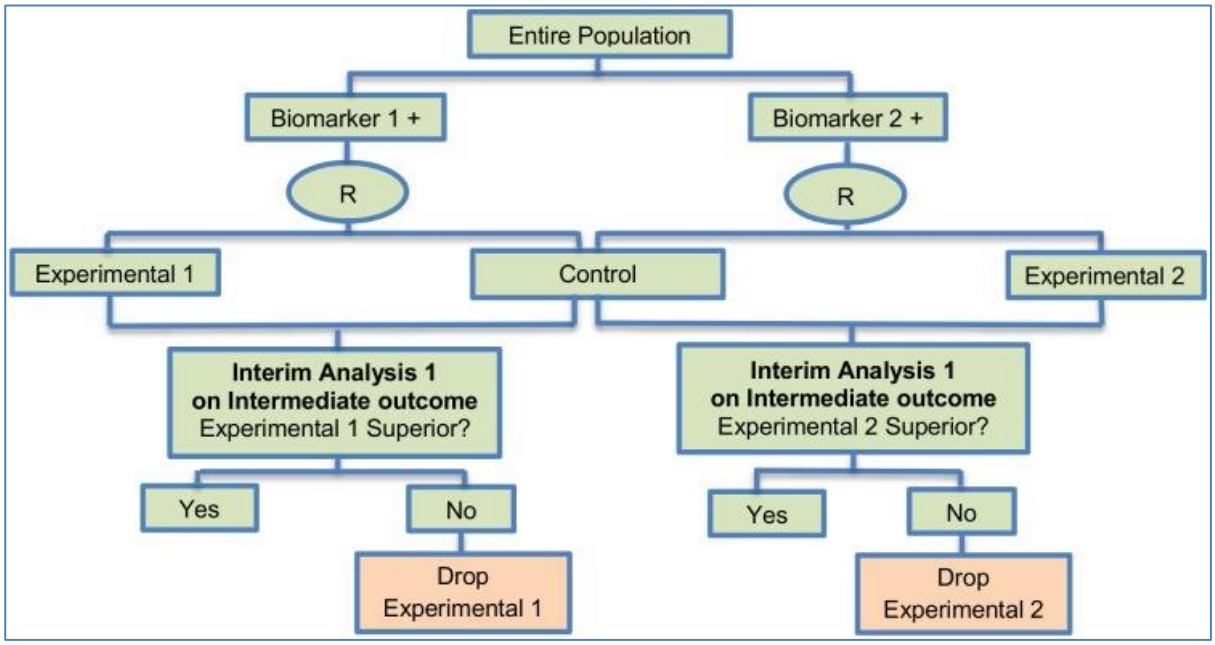
A recent methodological review of biomarker-guided adaptive trial designs in phase II and III identified eight distinct designs, four of which had variations. The eight designs and their adaptations included:[64]

i. Adaptive signature design (phase III) - incorporates the identification of a biomarker-positive subpopulation
ii. Outcome-based adaptive randomisation design (phase II) - randomisation ratio is changed.
iii. Adaptive threshold sample-enrichment design (phase III) - the inclusion criteria of the study population is changed after the initial stage of the study in order to broaden the targeted patient population
iv. Adaptive patient enrichment design (phase III) - information obtained from interim stage is used to broaden the targeted patient population
v. Adaptive parallel Simon two-stage design (phase II; non-randomised design) - the design starts with two parallel studies and, according to the results of the initial stage, selected or unselected patients are enrolled during the second stage
vi. Multi-arm multi-stage (MAMS) designs (phase II/III) - experimental arms can be dropped for futility from the study
vii. Stratified adaptive design (phase II) - the number of patients and decision rules are based on the observed response rates during the first stage of the study
viii. Tandem two stage design (phase II; non-randomised design) - assessment of treatment effectiveness in the entire population at the first stage of the study to make a decision about enriching the targeted patient population

An illustration of the multi-arm multi-stage (MAMS) design is provided in Figure B4 as an example of a biomarker-adaptive trial design. There are two variants of the MAMS approach referred to as the two-staged (phase II-III) adaptive seamless design, and the group sequential design.[64] Examples of clinical trials which use the MAMS approach includes the FOCUS4 trial,[71] which included patients

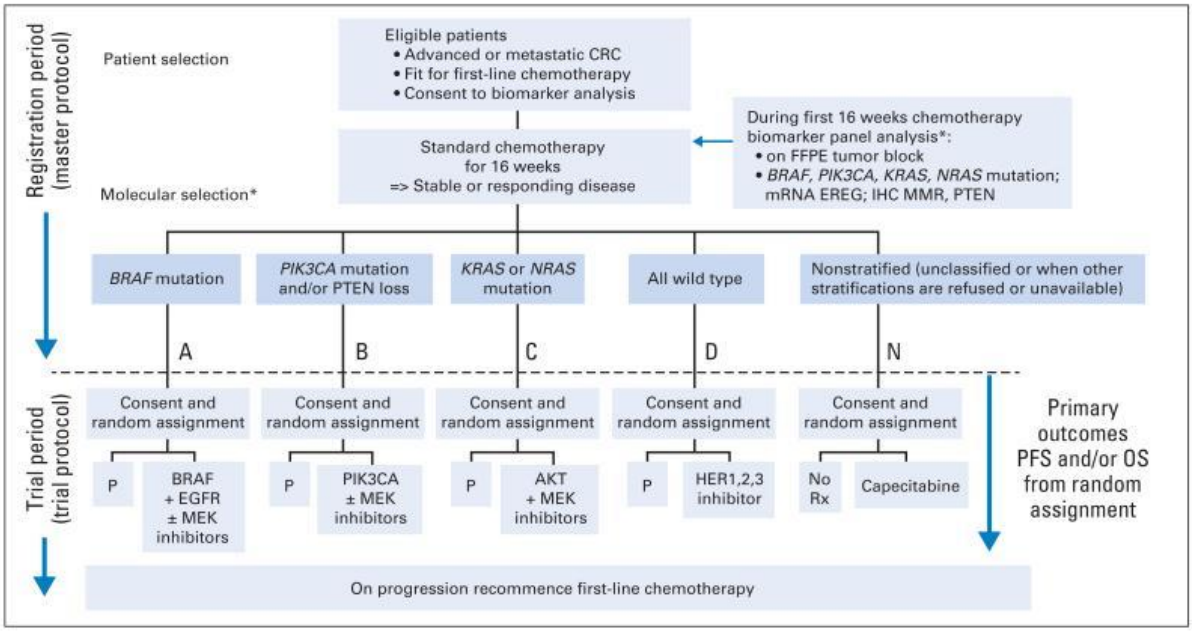
with advanced or metastatic colorectal cancer with stable or responding disease after first-line chemotherapy, and STAMPEDE,[72] which included men with high-risk prostate cancer starting long-term hormone therapy. These two trials, similar to other trials using adaptive designs, were not set up to evaluate treatment sequencing or designed to incorporate randomisation of participants to subsequent treatments. The STAMPEDE trial included the introduction of new treatments as well as dropping ineffective treatment arms, but essentially all the included treatments represented initial treatment. Figure B5 provides an illustration of the FOCUS4 trial, which aimed to simultaneously tests multiple targeted agents after induction therapy, during the maintenance phase, on the basis of the molecular aberration present in the patients' tumours.[73] The first induction therapy in the FOCUS4 trial was not part of the protocol and differed among included participants; there was also heterogeneity of post-progression treatments used.[73]

Figure B4: Illustration of the multi-arm multi-stage (MAMS) clinical trial design



Taken from: Antoniou, M., Jorgensen, A.L., Kolamunnage-Dona, R. (2016) Biomarker-guided adaptive trial designs in phase II and phase III: A methodological review. PLoS One, 11(2), e0149803.
Abbreviations: R randomisation of patients.

Figure B5: Schematic of a multi-arm, multi-stage platform clinical trial: FOCUS4



Taken from: Kaplan R, Maughan T, Crook A, Fisher D, Wilson R, Brown L, Parmar M. (2013) Evaluating many treatments and biomarkers in oncology: a new design. J Clin Oncol, 31(36). 4562–4568.

(*) The molecular cohorts are arranged in a hierarchy from left to right. For example, a patient with both a PIK3CA mutation and a KRAS mutation will be classified into the PIK3CA mutation cohort.

Abbreviations: CRC, colorectal cancer; EGFR, epidermal growth factor receptor; EREG, epiregulin; FFPE, formalin fixed, paraffin embedded; HER, human epidermal growth factor receptor; IHC, immunohistochemistry; MMR, mismatch repair; OS, overall survival; P, placebo; PFS, progression-free survival; Rx, treatment.

The new trial designs developed for evaluating biomarker-guided personalised therapies in oncology are also referred to as platform trials, umbrella trials, and basket (or bucket) trials.[69 74] 'Platform trial' is a descriptor for adaptive studies designed to evaluate multiple treatments in heterogeneous patient populations, with the capability of adding new treatments to be tested or dropping inefficient ones.[66] Both the STAMPEDE and FOCUS4 trials are multi-arm, multi-stage platform trials.[71 72] The term 'umbrella trial' design is used for studies that focus on a single tumour type or histology and incorporate the evaluation of many drugs in multiple subgroups (identified by the biomarkers) under the 'umbrella' of one study.[69 74] The FOCUS4 trial, in which patients are assigned to one of five parallel population-enriched, biomarker-stratified randomised trials (Figure B5),[71] is sometimes referred to as an umbrella trial.[69 73] The term 'basket trial' design is used for studies designed to allow the inclusion of multiple molecular subpopulations of different tumours and histologic types within one study (the basket).[69 74] These trials represent early phase, non-randomised, discovery trials,[63 75 76] and are generally used for the evaluation of drugs that treat rare cancers where it can be difficult to enrol sufficiently large cohorts for a confirmatory trial.[63 77 78] The changing paradigm of disease classification from one of organ and stage of disease to one of patient- and tumour-specific biology means that mutation-specific subtypes for common cancers, such as breast and colorectal cancer, now represent 'rare' subgroups.[63 69] An example of a basket trial design is provided by the MATCH (Molecular Analysis for Therapy Choice) trial, which aims to determine whether targeted therapies for

people whose tumours have certain gene mutations are effective regardless of their cancer type.[78] The trial includes patients with refractory solid tumours or lymphoma for whom no standard treatment for prolonging survival exists. The patients are assigned to histology-independent subgroups and receive the corresponding treatment that march their tumour's identified molecular abnormality.

C. CLINICAL SCENARIOS WHERE TREATMENT SEQUENCING WAS AN IMPORTANT CONSIDERATION FOR NICE DECISION MAKING

C1 Introduction

This next section provides an introduction and summary to three clinical indications within which decision making regarding sequential treatments were considered important. It also explores some of the issues and challenges involved regarding the available evidence base to inform these decisions. The indications include:

i. The introduction and sequential use of new 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

These three indications were identified in Chapter 4 as particularly relevant to NICE decision making, with a member of corresponding NICE technology appraisals or clinical guidelines included in the review of methods. In fact, several studies included in the review of methods considered these three decision problems. Therefore, rather than providing a detailed background explanation for each condition in the methodology review, an overall summary of the clinical context and the potential limitations of the available evidence base for these three indications are presented here. Definitions and brief descriptions of biological therapies are also provided, as a potential class effect of these drugs appears to be an important consideration for treatment sequencing.

The limitations of the available evidence base within these clinical examples are fairly typical of that available for the decision problems in a number of studies included in the review of methods. The thesis focuses on the methods used to evaluate treatment sequences, and as such does not aim to evaluate the effectiveness of treatments within these clinical conditions. This next section, therefore, does not represent a comprehensive review of the literature on effectiveness within these scenarios. Rather it aims to identify potential evidence gaps and challenges faced by reviewers evaluating treatment sequences. The evidence ‘gaps’ discussed here are those identified through the reviews conducted for this thesis and may in some instances no longer exist today. The overall approach used to treat some disease conditions will also have evolved since some of the studies included in the review of methods were conducted. For example, some modelling studies in rheumatology were conducted before the wide spread use of the “treat to target” approach. However, this does not mean that the methods used to evaluate treatment sequencing is not useful, therefore they were included in the review of methods.

C2 Biological therapies

Biological therapies are used for many disease conditions including inflammatory bowel disease, ulcerative colitis, Crohn's disease, multiple sclerosis, Alzheimer's disease, severe allergic disorders, hepatitis B, human immunodeficiency virus (HIV) infection, influenza A, osteoporosis, asthma, sciatic pain, and psoriasis, as well as cancer and inflammatory arthritis. The same underlying issues discussed for the next two clinical scenarios are also likely to be pertinent for most of these conditions.

Biological therapies are modified or man-made versions of substances that occur naturally in the body. They act at a molecular level, targeting specific processes in the cells. They are designed to repair, stimulate, or enhance the immune system's responses.[79] Different biological therapies have different mechanisms of action. As treatment for inflammatory arthritis, which is an autoimmune disease, they are designed to antagonise or modulate the activity of overactive immune cells. As cancer treatments, they may act by stopping cancer cells from dividing and growing, seeking out cancer cells and killing them, or encouraging the immune system to attack cancer cells.[80] Some biological therapies interfere with specific molecules involved in tumour growth and progression.[79]

There is no simple way of grouping biological therapies into different types as most groups overlap, with some drugs belonging to more than one group. They can be classified according to the type of chemical or active ingredient used, or by the way the biologic is used to treat a particular condition. For example, bevacizumab is monoclonal antibody, which targets specific proteins on cancer cells. It is also an anti-angiogenic drug, which inhibits cancers from developing new blood vessels. It works as an anti-angiogenic agent by blocking blood vessel growth factors from attaching to receptors on vascular endothelial cells, and is therefore also an example of an anti-vascular endothelial growth factor agent. Biological agents used to treat inflammatory arthritis target inflammatory cytokines, such as tumour necrosis factor (TNF), interleukin-1 (IL-1), or interleukin-6 (IL-6), and cells such as B and T lymphocytes. Examples include the TNF-inhibitors etanercept, adalimumab, infliximab; the IL-1 inhibitor anakinra; the IL-6 receptor-blocking monoclonal antibody tocilizumab; the anti-B-cell agent rituximab; and the T-cell co-stimulation inhibitor abatacept.[81]

C3 Scenario 1: Biological agents for inflammatory arthritis

C3.1 Inflammatory arthritis

'Chronic inflammatory joint disease' encompasses a large group of arthritic conditions of various aetiologies and clinical presentations, but with the common feature of persistent joint inflammation, or synovitis, and systematic inflammation.[82 83] The most common form of chronic inflammatory arthritis is rheumatoid arthritis; other examples include psoriatic arthritis, and ankylosing spondylitis. These are auto-immune diseases, where the body produces antibodies to its own tissues. The ongoing inflammation or uncontrolled disease activity leads to joint damage and disability, decreased quality of life, and increased co-morbidities such as cardiovascular and lung disease.[81 82] There is currently no cure for these conditions and the main focus of treatment is to maximise the reduction in disease

activity with the ultimate goal of achieving remission or, for patients where the state of remission may not be attainable, low disease activity.

C3.2 Biological agents for rheumatoid arthritis

More recent understanding of the pathophysiology and the course of rheumatoid arthritis has led to a move away from the traditional pyramid treatment approach to the current treat-to-target strategies, which aim to reach the target of remission, or low disease activity, in as short a time as possible in order to prevent the accumulation of irreversible damage.[84-86] Where these targets are not achieved, or the patient experiences a serious adverse effect, treatment is adjusted or switched as required. The state of remission, once reached, must then be maintained during the 10-25 year course of the disease. Patients who have achieved sustained remission are considered to have less disability, less erosive joint damage, and better quality of life.[81] However, about half of those who achieve remission will experience a flare-up or relapse within six to 12 months.[87] The treatment goals of remission or low disease activity are defined using specific criteria rather than global impression, and are measured using various scales. The most commonly used scale is the 28 joint count Disease Activity Score (DAS28), with remission defined as a score of less than 2.6.

The mainstay treatment for rheumatoid arthritis are disease modifying anti-rheumatic drugs (DMARDs), which are used to reduce synovitis, systemic inflammation and disability. There are two main types of DMARDs; conventional synthetic DMARDs, and newer biological DMARDs.[85 86] However, more recently the first synthetic targeted DMARD has been approved, with more in development.[86] Clinical guidelines recommend the use of the conventional DMARD, methotrexate, as first-line treatment.[85 86] Biological DMARDs are generally only given to people who have failed to respond to conventional DMARDs or who have experienced side-effects. However, they are also indicated as the initial treatment for patients with a poor prognosis.[85] At the time of the thesis there were five classes of biological DMARDs available for the treatment of inflammatory arthritis (listed in Section C2, Appendix Volume I), each inhibiting a different aspect of the immune-driven inflammatory pathway.[81] The first biological DMARDs to be approved for the treatment of inflammatory arthritis were TNF-inhibitors, and are recommended as first-line biological treatment (TA375, 2016). Biological agents can be used as monotherapy, but they are often used as an add-on to the previous conventional DMARD, generally methotrexate. The clinical effectiveness of TNF-inhibitors is often treated as interchangeable, due to the limitation of the evidence base. However, meaningful differences have been observed in their efficacy and safety profiles.[88]

Given the chronic nature of rheumatoid arthritis and that treatment failure is common, patients generally receive a sequence of treatments over time. Studies have shown that only 30% of patients achieve low disease activity with methotrexate,[89] and up to a third of patients do not respond to TNF-inhibitors;[90-93] approximately 50% of patients receiving TNF-inhibitors do not achieve a substantial clinical response (i,e, ACR50 or ACR70).[84 93] The main reasons for treatment switching include the treatment failing to provide sufficient benefit, the occurrence of side effects or intolerance, and patient

choice. The gradual reduction in drug efficacy over time, also referred to as secondary loss of efficacy, is a common phenomenon associated with chronic conventional DMARD treatment.[94] This phenomenon, which can be indicative of the onset of acquired drug resistance (immunogenicity), is also recognised with TNF-inhibitors.[94 95] This is caused by the patient developing antibodies that block the action of the drug, especially biological agents. This can occur within a few months of starting treatment or years later. When this occurs patients can be switched to a biological agent with a different mode of action. Inadequate response to one TNF-inhibitor does not preclude a response to another.[96] A second agent from the same class may still be effective due to individual differences in bioavailability, pharmacokinetic properties, immunogenicity, and mechanism of action.[93 97] The combination of a biological agent with a conventional DMARD such as methotrexate is also used to reduce the risk of developing anti-drug antibodies.[98]

A number of factors have been identified that predict poor prognosis in rheumatoid arthritis including severe arthritis with multiple joint involvement, late presentation, greater disability and presence of radiographic joint erosions at baseline, the presence of inflammatory markers associated with the disease such as rheumatoid factor (RF), and specific blood tests that indicate whole-body inflammation such as high levels of C-reactive protein (CRP). Some of these factors have also been identified as predictors of treatment failure with biological DMARDs, for example severe disease, poor functional status represented by higher Health Assessment Questionnaire (HAQ) score associated with the presence of rheumatoid factor, antibodies to citrullinated protein, and increased serum IgG concentration.[89] Economic evaluations and meta-analyses of treatment sequences may need to account for different patient subgroups. For example, one modelling study (Tran-Duy, 2014) included in the review of methods (Chapters 6-7) accounted for a differential treatment selection for those with positive or negative rheumatoid factor.[99] There are currently no validated genetic biomarkers for predicting response to individual biological agents.

For rheumatoid arthritis, two specific patient groups have attracted clinical and research interest, newly diagnosed patients with active disease who have not yet received DMARD, and patients with long-standing DMARD-refractory disease who have failed to respond to initial TNF-inhibitor treatment.[81] Early rheumatoid arthritis is now recognised as a distinct entity, along with the importance of early diagnosis and a ‘therapeutic window’.[100] A number of modelling studies in the review of methods evaluated the use of biological therapy for early onset rheumatoid arthritis, as outlined in Chapter 6, Section 6.6.

C3.3 The available evidence base to inform treatment sequencing of biologics for rheumatoid arthritis

This clinical scenario provides an example of where the licensing process has also left important gaps in the evidence base, including the sequential use of TNF-inhibitors, conventional DMARDs after the failure of biological therapy, and head-to-head comparison of TNF-inhibitors with a novel biological agent after the failure of first-line TNF-inhibitor.

The available evidence from RCTs

First generation TNF-inhibitors were evaluated by RCTs in two different situations, representing their use in early and late stage disease.

i. The pivotal trials of first-line TNF inhibitors in early stage disease generally included patients who had not previously received methotrexate. The TNF-inhibitor was usually compared to placebo, after a period of washout from any previous DMARDs. (For example ERA[101] and ASPIRE[102])
ii. Studies of late stage disease included patients who had failed a number of previous conventional DMARDs. In this situation, the new TNF-inhibitor was frequently added to an existing failed conventional DMARD, usually methotrexate, whilst patients in the control group received placebo in addition to the background methotrexate. Hence, the use of TNF-inhibitors represented an additional step in the treatment pathway. (For example ATTRACT[103] and the etanercept trial presented by Weinblatt *et al*.[104])

Later TNF inhibitors or other biologicals were also studied in a third situation for patients with late stage disease who had failed to respond to both conventional DMARDs and their first TNF-inhibitor. The novel biological agents were compared to either placebo (for example GO-AFTER)[105] or methotrexate (for example ATTAIN[106] and REFLEX[107]). Depending on the outcomes of the trials the new biological agents were licensed for any of these indications, with more recent drug development programmes having adopted all three approaches in order to establish a broad range of indicators for their new drug.[81]

One sequencing RCT aimed to answer the question of whether all newly diagnosed patients should be treated with initial combination therapy or start with the new and expensive TNF-inhibitor, infliximab: the Dutch Behandel Strategieen (BeSt) study.[108] Patients with recent onset rheumatoid arthritis (n=508) were randomised to one of four treatment groups:

i. Group 1 had sequential monotherapy: they were treated with one drug at a time, starting with methotrexate and switching to other drugs if there was no improvement (sulfasalazine, leflunomide, and then methotrexate with infliximab, if necessary).
ii. Group 2 followed a step-up regimen, beginning on methotrexate, but with more drugs added on as necessary (sulfasalazine, then hydroxychloroquine, and then prednisone, then switching to methotrexate with infliximab).
iii. Group 3 was started immediately on a combination of methotrexate, sulfasalazine, and a tapered high-dose prednisone (switching sulfasalazine for cyclosporin if necessary and then to methotrexate with infliximab).
iv. Group 4 also had combination therapy from the beginning, but with methotrexate and infliximab (and then if necessary leflunomide, sulfasalazine, cyclosporin, and prednisone).

The findings supported the use of early intensive treatment to suppress disease activity in order to minimise joint damage and disability.

Non-randomised evidence on sequential biologics

Open label extension studies

An alternative source of evidence for sequential TNF-inhibitors are prospective open label phase IV extension studies, where patients in placebo-controlled trials could switch to an alternative TNF-inhibitor on treatment failure. For example:

- OPPOSITE,[96] in which patients were switched to infliximab after an incomplete response to etanercept
- ReACT,[109] in which patients could switch to etanercept or infliximab after inadequate response to adalimumab
- Bingham, 2009,[110] in which patients could switch to etanercept after inadequate response to infliximab

These studies were used to inform a number of included modelling studies and are described in more detail in Chapter 6, Section 6.6.

This type of study, as noted in Section A (Appendix Volume I), is generally undertaken to assess the long-term safety and efficacy of interventions in patients previously enrolled in an RCT. They usually only include a proportion of the patients recruited to the original trial, do not generally include a comparative group, and no longer include blinding to the treatment allocation.

Open label observational studies

The comparison of the effectiveness of a second TNF-inhibitor with an alternative biological agent is also available from prospective, open label, observational real-world studies. For example:

- SWITCH-RA,[111] where patients who were non-responsive to one previous TNF-inhibitor were enrolled ≤4 weeks after starting rituximab or a second TNF-inhibitor.

Patient registries

A number of patient registers, which tend to be country specific, have been set up to monitor the potential long-term effects and adverse effects of biological agents. Studies based on patient registries have been used to assess the clinical effectiveness of sequential TNF-inhibitors.[27 112]

Examples of patient registers include:

- British Society of Rheumatology Biologics Registry (BSRBR),[113] which included all patients starting on a TNF antagonist agent within the UK NHS. The register was launched in October 2001 with the aim of recruiting a sample size of 4000 patients with a follow-up of 5 years each, for each of the three NICE-approved TNF-inhibitors (etanercept, infliximab and adalimumab) and a similarly sized conventional DMARD-treated control group.
- Registry of the Radbound University Nijmegen Centre,[99] also known as the Nijmegen Inception Cohort.[99 114] This is an open longitudinal study of patients with early rheumatoid arthritis and no prior DMARD use. It has been underway since 1985 at the department of Rheumatology of the University Medical Centre, Nijmegen in the Netherlands.

- Dutch RhEumatology Arthritis Monitoring (DREAM) biologic registry.[99] This started in 2003 to monitor and evaluate the use of biologics in patients who had not responded to methotrexate and at least one other conventional DMARD in the Netherlands.
- National Databank for Rheumatic Diseases. A patient-based multi-disease, multi-purpose rheumatic disease data bank which includes patients from the US and Canada.
- Swiss Clinical Quality Management in Rheumatologic Disease (SCQM-RA).[115] A national registry developed in collaboration with the Swiss Rheumatology Association (SGR)
- South Swedish Arthritis Treatment Group (SSATG) register is a large, prospective, observational study cohort, involving 11 rheumatology units.[116]
- Spanish Registry for Adverse Events with Biologic Therapies in Rheumatic Diseases (BIOBADASER). A National registry of patients on biological therapies establishes by the Spanish Society of Rheumatology (SSR).[117]
- German Rheumatoid Arthritis Observation of Biological Therapy (RABBIT) registry. An ongoing long-term prospective cohort study of patients with rheumatoid arthritis treated with biological or conventional DMARDs.
- Danish nationwide rheumatological database (DANBIO). Includes patients receiving biological therapy in Denmark since 2000.[118]
- Norwegian DMARD (NOR-DMARD) register. A 5-center register, established in December 2000, of all DMARD prescriptions to patients with inflammatory arthropathies.[119]

Evidence on conventional DMARDs after the failure of biological therapy

Randomised controlled trials evaluating the clinical effectiveness of conventional DMARDs were conducted prior to the introduction of biological therapies. Consequently it is not surprising that there is little evidence of the effect of conventional DMARDs after inadequate response to biological therapy.

Evidence to support the withdrawal of biological agents

Another question relevant to a treatment sequence in rheumatoid arthritis is whether a biological agent could be withdrawn once an acceptable disease state had been achieved. It is possible that a milder conventional DMARD could be used to maintain remission, if either the disease goes into natural remission or the drug results in a remission that would withstand drug cessation or dose reduction.[120]

C3.4 NICE technology appraisals of biological agents for rheumatoid arthritis

Four technology appraisals conducted to inform NICE guidance on the use of biological agents for rheumatoid arthritis (TA130, TA72, TA36, TA195) were included in the methodology review. The NICE guidance for these appraisals have since been replaced by an update appraisal published in January 2016 (TA375). This was identified as an ongoing appraisal (ID537) during the literature search. The update included a review of all previous appraisals evaluating the first-line biological agents: adalimumab, etanercept, infliximab, certolizumab pegol, golimumab, tocilizumab and

abatacept. This mega-review was undertaken by a review group at the School of Health and Related Research (ScHARR), University of Sheffield. The earlier technology appraisals were all conducted by a review group at the West Midlands Health Technology Assessment Collaboration at the University of Birmingham. The Birmingham team developed an economic model that allowed for the comparison of a sequence of conventional DMARDs with and without the addition of a TNF-inhibitor. This was then further adopted and updated to inform subsequent technology appraisals and NICE guidance on the use of biological agents for rheumatoid arthritis. The initial Birmingham Preliminary model (BPM) was based on fixed sequences of conventional DMARDs, whilst the subsequent improved version, the Birmingham model (BRAM), used a random generated sequence of conventional DMARDs. The modelling approach used is explored in more detail in Chapter 6. Both models start at the point where treatment with conventional DMARD are first initiated, requiring effect estimates for a sequence of treatments used across the patient's lifetime.

C4 Scenario 2: Biological therapies for advanced or metastatic cancer

C4.1 Advanced or metastatic cancer

A number of studies included in the review of methods evaluated treatment sequences for advanced and metastatic cancer, where the treatment did not have a curative intent. This scenario provides background information for these studies. This section does not, however, account treatments for advanced cancer with curative intent.

Treatment for advanced and metastatic disease has changed dramatically over the past 20 years for many cancers, including the advent of numerous biological or targeted therapies. For example, in advanced and or metastatic renal cell carcinoma seven new biological agents came on the market between 2006 and 2012.[121] The goal of treatment, when the cancer is not considered curable, is to prolong progression-free survival, maintain patients' quality of life, and ultimately prolong overall survival.[122] This is usually achieved using a sequence of treatments, generally based on a practice of switching treatments when the patient experiences unacceptable toxicity or disease progression. The increasing armamentarium and the fact that targeted therapies are associated with fewer or milder adverse effects means that maintenance therapy is now commonly used after initial treatment in some cancers.[123 124] Maintenance therapies are based on the introduction of additional treatment, typically lasting until disease progression, for patients who have a response or stable disease after a fixed duration of first-line treatment.[123] As a complete opposite to this, treatment holidays, or planned treatment interruptions are also used for some cancers in order to reduce the burden of treatment,[124 125] or to limit treatment resistance.[122 126]

The advancements and increased number of available systemic treatments means that many patients with advanced or metastatic cancer are treated long-term with multiple therapies. However, the multiple mechanisms of acquired resistance to targeted therapies mean that most patients progress at some point and are likely to require subsequent therapies.[125 127 128] This means that clinicians must make decisions about the best treatment beyond the effectiveness of first-line, and

need to make choices regarding the optimum sequencing approach from the onset. How best to use these agents in sequence, and how to expose patients to as many agents as possible is an ongoing challenge for clinical decision-making.[122]

The consideration of the optimal sequence of therapies, or assessing the place-in-pathway for specific treatments is an important issue for policy decision making.[121] The high costs of these new treatments also makes cost-effectiveness an important consideration.[129-131] There are a number of challenges regarding the available evidence for evaluating treatment sequences for advanced or metastatic cancers. As outlined in Chapter 1 (Section 1.3), many clinical trials focus on treatments used at a single point in the treatment pathway. The marketing considerations for new biological agents have also influenced the available trial designs, with the drugs having been investigated in a treatment-line that was optimum for the pharmaceutical company rather than the most appropriate treatment line.[125] The trials of second-line treatments are also sometimes based on the use of a first-line treatment that is no longer the treatment of choice in clinical practice.[132] As discussed in Section B3 of Appendix Volume I, the development of biotechnology and advancement in the understanding of genomics has led to the recognition of the heterogenetics of tumours of the same site and stage, with biological markers playing an important role in selecting subgroups of patients who are likely to respond to treatment.[68] The development of drug-resistance is also an important consideration.[122] Another challenge relating to the available evidence is that the outcome overall survival associated with first-line, and possibly second-line, treatments will be confounded by the differential use of subsequent-line treatments. Some of these challenges are illustrated in the clinical scenario relating to renal cell carcinoma (Section C4.3-5, Appendix Volume I).

C4.2 Optimum endpoint for evaluating sequencing of cancer treatments

Overall survival is an objective and unambiguous outcome measure, and considered the gold standard in cancer trials.[133 134] However, the drawback of overall survival is that it requires large patient numbers and prolonged follow-up, and can be confounded by the use of effective subsequent-line therapies, crossover within RCTs from the control to the investigational drug, and mortality unrelated to cancer.[124] (The issue of dealing with the impact of trial participants switching from the control to the intervention drug within the analysis of RCTs of 'single' treatments was introduced in Section A4 of Appendix Volume I; and discussed in Chapter 1, Section 1.3.1). Furthermore, the implicit assumption here is that the clinical trial intervention is the only treatment that matters, with overall survival providing a measure of the elapsed time between the date of randomisation (intention to treat) and death, and what happens in-between being irrelevant for measurement, and therefore a 'black box'.[135]

Progression free survival (PFS) and time-to-progression (TTP) are now recognised as valid surrogate end-points by drug regulatory authorities, and are increasingly being used as the primary outcome in cancer drug trials.[133 136] The growing number of new treatments being developed, and increasing demands for rapid evaluation and early availability of efficacious therapies in advance or metastatic

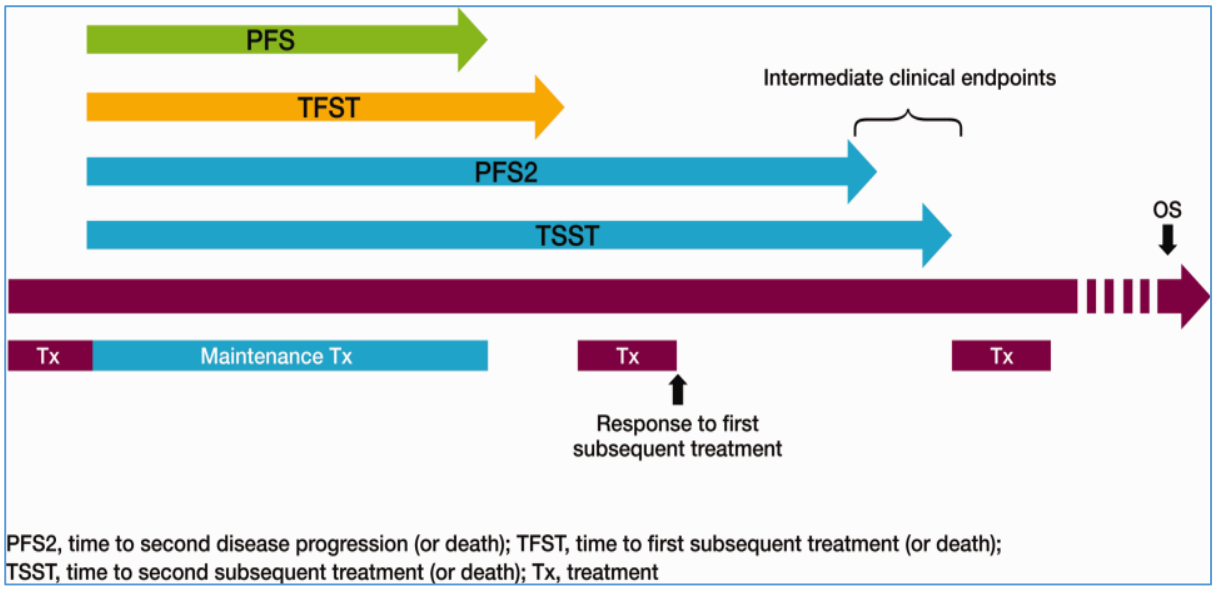
cancer means that these surrogate outcomes have become key for regulatory marketing approval decisions.[124] Progression-free survival, within the RCT setting, is defined as the time from randomisation to objective tumour progression or death from any cause.[136] The main advantage of this endpoint is that it is not confounded by subsequent treatment lines and can be assessed early.[124] [133] However, progression cannot be measured precisely and is therefore subject to measurement bias and error, and the time to progression is dependent on the frequency and timing of these measurements.

A large effect, in terms of progression free survival, is generally expected to be associated with a corresponding beneficial effect on overall survival.[136] However, an experimental treatment found to be advantageous in terms of progression free survival may also be associated with poor survival. This may be due to, for example, long-term toxicity, different resistance profiles to treatments used after progression, or to biological changes leading to increased metastatic potential.[136] Conversely, as noted in Chapter 1, Section 1.2.2, a short progression free survival may not necessarily predict a poor overall survival if exposure to the experimental treatment sensitises the tumour to the effect of subsequent treatment.[135] The European Medicines Agency (EMA) recommends that when further treatments are likely to be used, and in particularly where lack of efficacy of further treatments might be a concern, outcomes of subsequent treatments in terms of objective response rate, and progression free survival after next line of treatment should also be available where practicable.[136] This issue is a particular concern for maintenance treatment, where the prolonged administration of an agent can lead to resistance relapse that may reduce the ability of the patient to benefit from the same or similar agents in the future.[123] To account for this the European Medicines Agency (EMA) recently recommended the use of 'progression free survival 2' as an intermediate clinical endpoint for evaluating the efficacy of maintenance therapy in oncology trials. Progression free survival 2 is defined by the EMA as the time from randomisation to objective tumour progression on next-line treatment or death from any cause.[136] It also recommends that in some cases time to the start of third-line therapy or, for maintenance therapy, time to second subsequent treatment (TSST) can be used as proxy for progression free survival 2.[136] The main difference between the two endpoints, progression free survival 2 and time to start of the next treatment, is that at the time of the analysis, some patients may have experienced a second objective disease progression but not yet received third-line treatment.[137]

The different endpoints used for evaluating the efficacy of maintenance therapy are illustrated in Figure C1. It is important to note that 'progression free survival 2' referred to here is different from 'second progression free survival', which refers to the progression free survival associated with the next line of treatment (i.e. the interval between relapse/start of the next line-treatment and second disease progression or death from any cause).[138] In the example presented in Figure C1, the 'initial' treatment is the maintenance treatment, and progression-free survival 2 refers to the period from treatment initiation (or randomisation) to progression on the first subsequent treatment, which would be the same as second-line treatment in a clinical trial of two treatment lines. Progression free

survival 2 includes the intention-to-treat population, whereas second progression free survival is limited to the subset of patients who have relapsed and received the next treatment, and are therefore, likely to have a more aggressive disease.[138] For optimal results, progression free survival 2 needs to be pre-specified as the clinical trial endpoint, and the initial and subsequent treatments defined in the protocol.[123 138] However, this may not always be appropriate as they may, for example, need to be selected based on performance status.[137] Freidlin *et al*. also note that quantifying the extended disease control using progression free survival 2 may require additional logistic considerations associated with long-term outcomes, and defining this endpoint for patients who never receive subsequent therapy post progression (after randomisation) requires careful consideration.[123]

Figure C1: An overview of clinical endpoints with respect to a disease course involving multiple rounds of subsequent treatment.



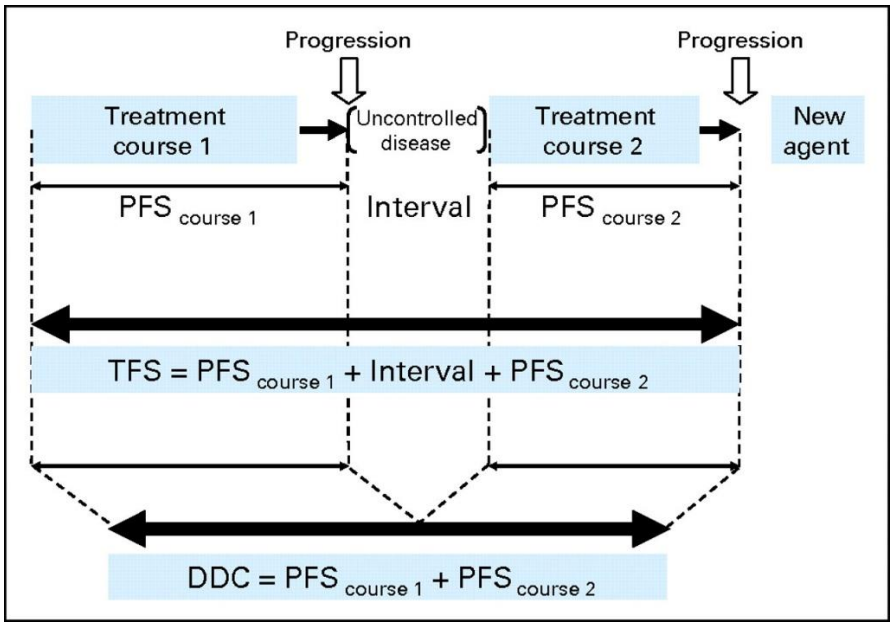
Taken from: Matulonis UA, Oza AM, Ho TW, Ledermann JA. (2015) Intermediate clinical endpoints: a bridge between progression-free survival and overall survival in ovarian cancer trials. Cancer, 121(11), 1737-46.
Abbreviations: OS overall survival; PFS, progression-free survival; PFS2, time to second disease progression or death; TFST, time to first subsequent therapy or death; TSST, time to second subsequent therapy or death; Tx, treatment

An alternative endpoint sometimes reported in oncology trials is time to treatment failure (TTF). This is defined as the time from randomisation to discontinuation of therapy for any reason including death, progression, toxicity or add-on of new anti-cancer therapy.[136] However, this is a composite endpoint influenced by factors unrelated to efficacy, for example, discontinuation may be a result of toxicity, patient preference, or the clinician's reluctance to continue therapy.[133] The usefulness of this endpoint, for informing treatment sequences is therefore dependent on how well it is recorded. Time to change treatment (TCT) is another similar end point, which can also provide useful information about how effective or tolerable an intervention is, but again will yield a more reliable source of information if the reasons for discontinuation or treatment switching is recorded.[121]

Two composite endpoints that have been proposed for evaluating either a fixed sequence of two treatment lines or 'stop-and-go' treatment strategies in advanced colorectal cancer include duration of

disease control and time to failure of strategy.[139 140] These are illustrated in Figure C2. Duration of disease (DDC) control is defined as the sum of the progression free survival of each active line of treatment, except when progressive disease is observed at either reintroduction or second-therapy. In other words, the sum of progression free survival for first-line treatment plus the progression free survival for second-line treatment, if the second treatment achieved stabilisation or response.[139] Time to failure of strategy (TFS) is defined as the total progression free survival from the initiation of the sequence to disease progression while on all the planned agents, or disease progression during a treatment-free interval and no further therapy received within one month, or death. This is similar to the progression free survival 2 descried above. The use of profession free survival associated with each successive treatment line to inform treatment sequencing assumes that all treatment effect from each treatment line stops on progression.

Figure C2: Illustration of two composite endpoints duration of disease control (DDC) and time to failure of strategy (TFS)



Taken from: Chibaudel B, Bonnetain F, Shi Q, Buyse M, Tournigand C, Sargent DJ, et al. (2011) 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. J Clin Oncol, 29(31), 4199-204.
Abbreviations: DDC Duration of disease control; PFS progression free survival; TFS Time to failure of strategy.

A potential challenge of using RCT evidence for new cancer treatments is the lack of maturity of the data for progression-free survival and even more so for overall survival, especially when studies have been terminated early, or are statistically powered according to the outcome progression-free survival.[141]

The importance of adequate reporting and interpretation the endpoint progression free survival within the primary studies for evaluating treatment sequences in advance cancer is discussed in Chapter 5 (Section 5.3.2), using a literature review by Stenner *et al*., as an example. The use of progression

free survival associated with each treatment line for modelling treatment sequences is discussed in Chapter 7, Sections 7.3.2 and 7.3.4.

C4.3 Targeted therapies for metastatic renal cell carcinoma

This scenario represents an example of a cancer that often presents as advanced or metastatic disease, and where the introduction of biological or targeted therapy has revolutionised its treatment.[143] It provides an example of where some sequencing trials exist but important gaps in the evidence base for informing decision making still remain. The scenario also illustrates the limitation of using overall survival for evaluating treatment sequences or multiple lines of treatment.

Metastatic renal cell carcinoma is largely resistant to chemotherapy, radiotherapy and hormonal therapy. Up until 2005, immunotherapy including the cytokines interleukin-2 (IL-2) and interferon alfa (IFNa) was the mainstay treatment.[132] A number of new agents have since been approved based on their ability to improve response rates, progression free survival or overall survival as first- or second-line treatment of metastatic renal cell carcinoma.[125] Targeted therapies have been shown to prolong survival and result in fewer adverse effects than cytokine treatment.[125 144] These targeted therapies fall into two mechanistic categories: vascular endothelial growth factor (VEGF)-based therapies, and mammalian target of rapamycin (mTOR) inhibitors. The first category includes vascular endothelial growth factor receptor-tyrosine kinase inhibitors (VEGFR-TKIs) and an anti-vascular endothelial growth factor monoclonal antibody (bevacizumab).

The most common first-line therapy for advanced or metastatic renal cell carcinoma at the time of the review was VEGFR-TKI.[145 146] However, clinical guidelines continue to evolve regarding the use of first-line therapy in treating naïve patients, and second-line therapy after relapse.[146] Most patients eventually develop resistance to first-line targeted therapy, usually within a year or less.[146 147] On disease progression, the subsequent therapy could potentially include an agent hitting the same target or using a different mechanism of action.[146 147] It has been suggested that changing the mechanism of action may decrease potential for cumulative toxicity.[146] However, there is also evidence suggesting that there is no absolute cross-resistance between VEGFR-TKIs.[122] The degree of cross-resistance that occurs between agents is not known, nor which agents, when used consecutively, are associated with the most favourable outcome.[146 148] According to Fischer, *et al*. around 33-59% of patients receive second-line treatment. Retrospective cohort studies suggest that up to 20% of patients proceed to third-line treatment.[132 149] Recent reviews of on-going and recently completed trials concluded that the optimal sequencing approach of targeted therapies for advanced or metastatic renal cell carcinoma, to maximise long-term clinical benefit, remains unclear.[132 144 146 150]

C4.4 The available evidence base to inform treatment sequencing of biologics for metastatic renal cell carcinoma

Several sequences have been evaluated in RCTs as well as prospective and retrospective studies, but the optimal sequence of treatments remains unclear.[125 142 146 150 151] Existing sequencing trials

have only considered a limited number of treatment lines, and the rapid expansion of available treatments meant that they would quickly become outdated.[132 146] The limitations of the data means that some reviewers use the treatment effects of individual treatments to represent a 'class effect' at a given line of therapy.[132] The NICE technology appraisal process evaluates specific treatments used in line with their respective marketing authorisations, which means that very few RCTs are eligible for the review purpose.

At the time I conducted the literature review for the thesis, the available RCT evidence to inform decisions on the optimal sequence of targeted therapies included four sequencing RCTs (SWITCH, RECORD III, SRART, and SWITCH-2), but two were still ongoing.[132] All four were limited to two lines of therapy, but one ongoing trial (START)[152] included patients who had progressed or were intolerant to at least one prior TKI. The remaining three trials included a treatment naive patient population. Two RCTs (SWITCH[153] and SWITCH-2[154]), one of which is ongoing,[154] evaluated a sequence of named TKIs. One RCT (RECORD III) compared a sequence consisting of a TKI followed by mTOR, with a sequence of the same drugs used in reverse order.[155] The remaining RCT, which was ongoing, included the following sequences TKI-mTOR, TKI-VEGF, mTOR-VEGF, mTOR-TKI, VEGF-TKI, VEGF-mTOR, in a patient population who had previously failed a TKI. Two completed single-treatment RCTs (AXIS;[143] RECORD-1[156]) were used to inform treatment sequencing of a TKI followed by an mTOR. RECORD 1 compared the mTOR everolimus with placebo in patients with TKI refractory metastatic renal cell carcinoma (21% of patients were post-sunitinib, and 53% had received two previous treatments including one TKI and cytokine). The second trial (AXIS) compared two TKIs in patients with metastatic renal cell carcinoma who had relapsed on one prior first-line treatment (sunitinib, bevacizumab, temsirolimus, or cytokine based therapy).[145 146] The study included subgroup analyses of overall survival according to previous treatment. In summary, the RCTs provided little data for informing whether one sequence was better than the other, and provided no direct comparison of TKI followed by TKI with mTOR followed by TKI. Only two RCTs provided data on third-line treatment. In concluding their narrative review finings, Pal *et al*. (2013) noted that a future RCT was needed for comparing the following three lines of treatment: TKI-TKI-mTOR vs TKI-mTOR-TKI.[146] This review was not included in the review methods as it was based on a narrative synthesis.

A number of clinical trials have evaluated the use of second-line treatment with targeted therapies in a cytokine pre-treated population. However, the use of first-line cytokine therapies has been largely replaced by sunitinib, or other VEGF/VEGFR inhibition therapies.[132] Most or the remaining evidence on second-line treatment is based on first-line sunitinib, with results generally being applied to all VEGFR inhibiting agents considered in the first-line setting, without having studied their use in sequence.[132] Sequencing after mTOR inhibition is poorly defined.[132]

The available observational studies included retrospective case series, uncontrolled trials, and registry or database studies.[146] These included much larger numbers of patients than the RCTs and included data on up to three lines of treatment. However, Pal *et al*., reported having problems with

interpreting the data from these studies as they only accounted for patients who had completed the entire sequence, and overlooked patients who were lost after first or second treatment due to lack of efficacy, clinical deterioration, or drug acceptability issues.[146] The studies indicated that 10-34% of patients with metastatic renal cell carcinoma received second- or third-line therapy, and most were lost after first-line treatment.[146] This is an important limitation of using registry data to inform treatment sequencing, which I also discuss in Chapter 5, Section 5.3.5-6 and 5.8.10.

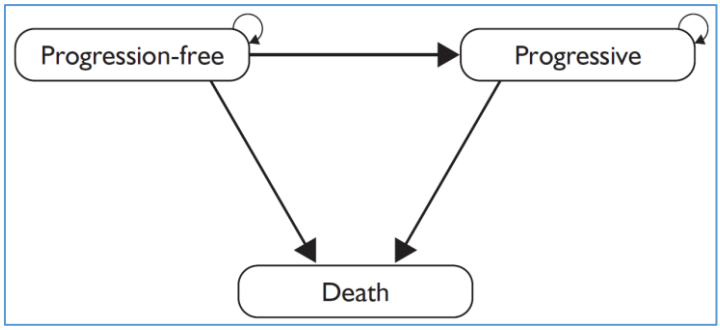
Not all patients benefited from the new targeted therapies, and as yet no predictors have been identified to select patients who might benefit, or those who demonstrate primary resistance to specific drugs.[144] Studies investigating sequencing beyond the first-line treatment setting generally did not use prognostic criteria to define their patient selection.[132]

C4.5 NICE technology appraisals of targeted therapies for metastatic renal cell carcinoma

Three technology appraisals (TA169, TA178, TA219) of targeted therapies for renal cell cancer were identified during the website search, indicating that treatment sequencing may have been a potential issue during the decision-making process. At present only sunitinib, pazopanib and axitinib have been recommended for the treatment of advanced or metastatic renal cell carcinoma by NICE. Sunitinib (TA169, 2009) and pazopanib (TA215 2011) are recommend as first-line treatment, and axitinib (TA333, 2015) after the failure of treatment with a first-line TKI or a cytokine. Sorafenib, temsirolimus, everolimus and bevacizumab are not currently recommended by NICE. However, the appraisal of everolimus (TA219) is currently under review (ID1015) as part of the cancer drugs fund (CDF) rapid reconsideration process.

The multiple technology assessment (TA178) of targeted therapies for advanced renal cell carcinoma was not included in the methodology review as both the evaluation of clinical and cost-effectiveness considered first- and second-line treatments separately.[157] The results of the systematic review were synthesised narratively and a decision-analytic Markov-type model was developed to simulate disease progression over time and estimate the cost-effectiveness of the interventions under consideration. A similar model structure, employing survival analysis, was used for evaluating both first and second-line treatments, but using different data to inform the model parameters. The model, which is illustrated in Figure C3, used three states: progression free survival, progressive disease, and death. The same modelling approach was also proposed as being the most informative for developing the *de* novo economic evaluation for the planned multiple technology assessment in 2016 (ID897). This was supposed to represent an update of previous reviews for selected second-line treatments. However, the review has been suspended, and TA178 has been put on the static appraisals list. The modelling approach is also analogous to that used in other NICE technology appraisals that were identified, for example the appraisal of treatment for metastatic colorectal cancer (TA118)[158] and ovarian cancer (TA91).[159] Alternative model structures that account for both first and subsequent treatments for advanced or metastatic cancers in the same model were identified during the review of methods, which are presented in Chapter 7.

Figure C3: Influence diagram for the economic model in NICE TA187



Taken from: Thompson Coon, J., M. Hoyle, et al. (2010) Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. Health Technology Assessment, 14(2), 1-184, iii-iv.
Note: The boxes represent health states and arrows represent transitions between states. At any moment a patient is assumed to be in one of the states. Patients move between states once during each model cycle. This means that if a patient is in PFS, for example, then during the next cycle they can either die, move to PD or stay in PFS. The health states of a cohort of patients are modelled at each discrete model cycle. All patients enter the model in PFS, having been diagnosed with advanced/metastatic renal cell carcinoma. Patients remain in PFS until they die or the disease progresses. Once patients enter the PD state they remain there until death.
Abbreviations: PD progressive disease; PFS progression free survival

C5 Scenario 3: The use of newer antiepileptic drugs

C5.1 Treatment of epilepsy

Epilepsy is not a uniform condition but comprises many different seizure types and epilepsy syndromes. The severity of the condition and the prognosis vary according to the type of epilepsy. Seizures can be broadly categorised into two types, partial seizures, also categorised as ‘focal’ or ‘localisation-related’ epilepsies, and generalised seizures.

The aim of treatment is to eliminate seizures completely, whilst at the same time keeping the side effects of treatment to a minimum so that the person can lead as normal a life as possible. However, partial seizure freedom, representing a 50-99% reduction in seizures, may also represent treatment success for some patients. Treatment with a single drug is generally preferred, to minimise the risk of adverse effects. Monotherapy is initiated by increasing the dose gradually until seizures are controlled or adverse effects become unacceptable. If treatment fails, it is considered preferable to try alternative monotherapies before moving on to combination treatment. Epilepsy is resistant to drug therapy in a third of patients, which means up to 30% of individuals will continue to have seizures on monotherapy.

Adults with epilepsy are expected to take anti-epileptic drugs (AEDs) for most of their lives, and therefore any economic modelling needs to take into account the longer-term costs and outcomes, including those associated with patients not responding to treatment. In childhood epilepsy, however, treatment withdrawal is likely to be an option for patients who have achieved complete seizure freedom. There are also specific subgroups of patients with epilepsy for whom treatment sequencing decisions will differ for example, due to specific adverse effect profiles of some drug. These include, for example, people with learning disabilities, pregnant women, and women of child bearing age.

The time spent on a specific antiepileptic drug and the cause of treatment failure are both important considerations when modelling treatment sequences for epilepsy. Randomised controlled trials indicate that unacceptable side-effects tend to lead to treatment discontinuation earlier than lack of effectiveness, often occurring within the titration period, whilst the majority of patients who discontinue due to lack of effect will do so within a year.[160-163] An RCT[161] and observational study by the National General Practice Study of Epilepsy[164] showed that patients who are likely to achieve complete remission will do so quickly, usually within the titration period, and that most complete remissions are sustained in the long term.[162 163]

C5.2 The available evidence base underpinning the use of newer antiepileptic drugs

This clinical scenario provides an example of where the evidence base included a large pragmatic trial for which individual patient-level data was available for at least two-year follow-up and previous antiepileptic drugs used. The Standard And New Antiepileptic Drugs (SANAD) trial was funded by the Health Technology Assessment Programme of the UK National Health Service.[160] It set out to compare clinicians' choices of one of the standard drug treatments (carbamazepine or valproate) with new antiepileptic drugs used as first-line monotherapy for newly diagnosed patients with epilepsy. It collected data on each patient's subsequent sequencing of treatments and included a far greater number of patient years of follow-up than any previous study. There is now a SANAD-II, which incorporates newer anti-epilepsy drugs that were not available during SANAD-I.[165]

The majority of drug trials in epilepsy are of monotherapy in either newly diagnosed patients or, at the polar opposite, combination therapy in patients with refractory epilepsy. There is an evidence gap in the middle, which an evaluation of sequential treatments might provide some insight to. In order to protect patients from the potential dangers of a non-effective agent used as monotherapy, a new agent is frequently initially studied as adjunctive (combination) therapy in refractory patients, and when their efficacy and tolerability as adjunctive treatments are established, they are then studied as monotherapy. However, two agents, felbamate and oxcarbazepine, were studied as adjunctive therapy and monotherapy simultaneously. Most of the available trials of newer antiepileptic drugs are supported by Industry, and as such aim to answer restricted licensing questions.[166] They compare the drugs used at different doses, have relatively short-term treatment durations, and often fail to limit recruitment to either partial or generalised onset seizures.[163]

C5.3 NICE technology appraisals and clinical guidelines of newer antiepileptic drugs

This clinical scenario highlights the potential variability in the way treatment sequences are conceived within decision models developed by different review groups for informing NICE decisions. Two technology appraisals and one clinical guideline that evaluated the introduction of 'newer' antiepileptic drugs were included in the review of methods. These are listed in Table 4.2 and 4.3. Both the technology assessments, one that included epilepsy in adults (TA76) and the other children (TA79), included an economic evaluation of the optimum treatment sequence. The updated clinical guideline (CG137) for epilepsy treatment also included *de novo* economic evaluations that

considered treatment sequences. The review of NICE guidance also identified a single technology appraisal (TA232) where the evidence review group criticised the manufacture's model for not considering treatment sequencing (listed in Table 4.1 and 4.2).

The overall model structure developed for the technology appraisal of antiepileptic drugs for adults (TA76), published in 2004, was also used in the clinical guideline (CG137), published in 2006. This was based on three elements of treatment sequencing, which represented different patient populations: monotherapy for newly diagnosed patients, monotherapy for refractory patients, and combination therapy. However, the two economic evaluations differed in the way they implemented the resulting models. The model used in the technology appraisal was based on the dichotomised outcome of treatment response versus non-response, with non-response leading to treatment switching. The clinical guideline model, on the other hand, considered complete and partial response separately. In the technology appraisal it was assumed that treatment effect of each antiepileptic drug was independent of positioning in the sequence. However, in the clinical guideline it was assumed that an antiepileptic drug used as second-line monotherapy would be less effective than when used as first-line treatment if the first drug was discontinued due to inefficacy. If the first drug was discontinued due to toxicity, the response to the second was assumed to be independent of response to the first. The differences between the assumptions and modelling approaches used in these two economic evaluations are explored in more detail in Chapter 6 and 7 respectively. One of the reasons why the clinical guideline model was able to consider more detail was that they used individual patient-level data from the SAND trial to estimate the parameters for the model. The authors noted that clinical opinion and observational data suggest the cause of treatment failure is prone to influence the likelihood of response to subsequent treatment (CG137). However, most primary studies reported treatment response as an aggregate marker for efficacy and tolerability, and did not differentiate between the two when reporting treatment discontinuation. The use of data from the SANAD trial[160] allowed the model to differentiate between patients who withdrew from first-line treatment due to adverse effects and those who withdrew due to inefficacy. The assumptions used to assess the response to second-line treatment were based on an observational study.[167] Patients who had not achieved remission but were not classified as having failed treatment were assumed to persist with first-line monotherapy for two years, at which point the patient was classified as having failed due to inadequate seizure control and moved on to second-line treatment.

Both economic evaluations were based on the same assumption regarding long-term response, and the probability of switching treatment after achieving remission. It was noted in both the technology appraisal (TA76) and clinical guideline (CG137) that observational data and clinical experience indicated that the probability of a patient changing treatment decreases as the time they have been on the given treatment increases. In the technology appraisal of retigabine (TA232) the evidence review group did not accept the validity of the manufacturer's modelled assumption that people whose epilepsy responds to treatment with retigabine do not experience any change in clinical response over time. The Committee heard from the clinical specialists that people with epilepsy are

likely to take anti-epileptic drugs over a lifetime, but that they would switch between drugs during this time.

The modelling technique used in technology appraisal 79 of newer antiepileptic drugs in children (TA79), published in 2004, is described in more detail in Chapter 7. It accounted for different levels of response, the potential for subsequent treatment withdrawal, and the need to consider different patient subgroups. However, the limitations of the evidence base was a big challenge.

D Description of the decision problems evaluated in studies modelling treatment sequences for rheumatoid arthritis and an assessment of how treatments administered before and after the decision point to interest were accounted for

Thirty-three modelling studies that were included in the review of methods (Chapter 3) investigated treatment sequences for rheumatoid arthritis (Table D1). Thirty of these studies investigated treatment sequences that included biological therapies. A brief overview of the treatments for rheumatoid arthritis and the available evidence base is provided in Section C3. This section provides an overview of the decision problems and treatment sequences evaluated by these studies, focusing on those that investigated the use of biological therapies. It also includes an assessment of the different approaches that were adopted by these studies to account for treatments administered before and after the decision point to interest (discussed in Chapter 6, Section 6.2).

D1 Description of the economic decision problems and type of treatment sequences evaluated in rheumatoid arthritis

The included studies of rheumatoid arthritis investigated the cost-effectiveness of different pre-defined treatment sequences. The economic decision problem, relating to treatment sequencing, and the starting point of the model for each included study is outlined in Table D1. The decision problems considered by the studies are divided into in four categories, which relate to the use of:

i. Conventional dmards or TNF-inhibitors in early disease
ii. TNF-inhibitors in established rheumatoid arthritis, or after inadequate response or intolerance to prior conventional dmards
iii. TNF-inhibitor or an alternative biological agent after failure of the first TNF-inhibitor
iv. TNF-inhibitor or alternative biological agent after the failure of two TNF-inhibitors

Each category is represented in Table D1 as a separate column. The decision problem(s) considered in the individual studies is represented by the dark grey shading, and the pale grey illustrates the additional 'sections' of the treatment pathway, or predefined treatment sequences, that were also included in the model. Different symbols are used to differentiate between the use of TNF-inhibitors and alternative biological agents, with the former being represented by a cross, and the latter with a black dot. The letter c is used to represent the use of conventual DMARDs. Essentially Table D1 provides a visual summary of the treatment sequences and corresponding decision being point modelled by each study.

The actual treatment sequences that were evaluated by the included studies, which investigated the use of biological therapies are illustrated in Figures D1-D3. Six studies (Chen, 2006; Davies, 2009; Finckh, 2009; Kobelt, 2011; Schipper, 2011; Tanno, 2006) investigated the introduction of biological agents in early disease in a patient population who were either DMARD-naïve or only failed to respond to one previous conventional DMARD (Figure D1; corresponding to decision problem 1 in Table D1).[95 168-172] Twelve further studies (Bansback 2005; Barton 2004; Brennan, 2004; Coyle, 2006; Diamantpoulus, 2012; Diamantopoulois, 2014; Jobanputra, 2002; Russell, 2009; Tran-Duy, 2014;

Wailoo, 2006; Welsing, 2005; Wu, 2012) 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 (Figure D2; corresponding to decision problem 2 in Table D1).[99 114 115 173-181] Twelve studies (Beresniak, 2011; Beresniak, 2013; Brennan, 2007; Cimmino 2011; Clark, 2004; Hallinen, 2010; Kielhorn, 2008; Lindgren, 2009; Malottki, 2011; Merkesdal, 2010; Puolakka, 2012; Saraux, 2010) investigated the use of biological agents in a patient population who have had an inadequate response to previous TNF-inhibitors (Figure D3; corresponding to decision problem 3-4 in Table D1).[182-193]

Table D1: Number of rheumatology studies considering each decision problem
(*Studies ordered according to the decision problem, highlighted by the dark grey shading, and then alphabetically*)

Author year	Model type	Model starting point / defined population entering model	Decision problem (patient population)			
			1) DMARDs for early RA (DMRAD naive)	2) Biologics for established RA (≥2 cDMARDs)	3) 2nd biologic (IR to 1st TNF)	4) 3rd biologic
Albert, 2000	Markov cohort	Patients with early RA and DMARD naïve.	c			
Maetzel, 2002	Markov cohort	Patients with early RA and eligible for MTX	c			
Schadlich, 2005	Partitioned survival	Patients with early RA and DMARD naïve.	c			
Chen, 2006	IPS	Patients with early RA and DMARD naive; lifetime horizon	+	+	+	c / +
Davies, 2009	IPS	TNF in early RA	+		+	c
Finckh, 2009	IPS	TNF in early RA	+		+	+
Kobelt, 2011	IPS	TNF in early RA	+		+	c
Schipper, 2011	Markov cohort	TNF in early RA	+	+	+	●
Tanno, 2006	Markov cohort	100% IR rate assumed for first cDMARD (bucillamine) in the sequence	+ *(after IR to 1 cDMARD)*		c	c
Bansback 2005	IPS	Patients with IR to ≥2 cDMARDs; lifetime horizon		+	c	
Barton 2004	IPS	Patients with early RA and DMARD naive; lifetime horizon	c	+	c	c
Brennan, 2004	IPS	Patients with IR to ≥2 cDMARDs; lifetime horizon		+	c	c
Coyle, 2006	Markov cohort	All patients entering model received MTX as their first treatment	c	+	c	
Diamantpoulus, 2012	IPS	Patients with IR to cDMARDs; lifetime horizon		+ / ●	+ / ●	+
Diamantopoulois, 2014	IPS	Patients with IR to ≥2 cDMARDs; lifetime horizon		+ / ●	+ / ●	+
Jobanputra, 2002	IPS	Patients with early RA and DMARD naive; lifetime horizon	c	+	c	c
Russell, 2009	Decision tree	Patients with IR to cDMARDs, eligible for biologic therapy		+ / ●	+ / ●	+
Tran-Duy, 2014	IPS	Patients newly diagnosed	c	●	●	+ / ●

Author year	Model type	Model starting point / defined population entering model	Decision problem (patient population)			
Wailoo, 2006*	IPS	Patients with IR to cDMARDs, eligible for biologic therapy		+ / ●	+	+
Welsing, 2005	Markov cohort	Patients with IR to ≥2 cDMARDs including MTX (eligibility for TNF in the Netherlands)		+	c	
Wu, 2012	Markov cohort	Patients with IR to cDMARDs, eligible for biologic therapy		+	●	c
Beresniak, 2011	Decision tree	100% IR rate assumed for first TNF (ETA) in the sequence.		+	+ / ●	+ / ●
Beresniak, 2013	Decision tree	100% IR rate assumed for first TNF (ADA) in the sequence.		+	+ / ●	+ / ●
Brennan, 2007**	IPS	Patients with IR to ≥2 cDMARDs; lifetime horizon		+	+	
Cimmino 2011	Decision tree	100% IR rate assumed for first TNF (ETA) in the sequence		+	+ / ●	+ / ●
Clark, 2004	IPS	Patients with early RA and DMARD naive; lifetime horizon	c	+	+ / ●	●
Hallinen, 2010	IPS	Patients who have had an IR to their first TNF inhibitor (ETA, ADA)			+ / ●	+ / ●
Kielhorn, 2008	IPS	Patients who have had an IR to their first TNF			+ / ●	c / +
Lindgren, 2009	IPS	Patients who have had an IR to their first TNF			+ / ●	●
Malottki, 2011	IPS	Patients who have had an IR to their first TNF			+ / ●	c
Merkesdal, 2010	IPS	All sequences started after IR to the TNF (ETA)			+ / ●	
Puolakka, 2012	Decision tree	100% IR rate assumed for first TNF (ETA, INF, or ADA) in the sequence		+	+ / ●	+
Saraux, 2010	Decision tree	100% IR rate assumed for first TNF (ETA) in the sequence		+	+ / *	+ / ●

Abbreviations: ADA adalimumab; ASM advanced simulation model (developed by Dr Ariel Beresniak); BRAM Birmingham Rheumatoid Arthritis Model; BSRBR British Society of Rheumatology Biologics Register DT decision tree; cDMARDs conventional disease-modifying anti-rheumatic drugs; ETA etanercept; INF infliximab; IR inadequate response; IPS individual patient simulation; MTX methotrexate; NA not applicable; RA rheumatoid arthritis; TNF tumour necrosis factor α inhibitor

Note: Dark grey shading represents the decision problem of interest and the pale grey represents the starting point of the model, i.e. the position in the treatment pathway that patients entering the model are at or the start of the predefined treatment sequences being modelled. The symbols represent the type of treatments being evaluated.

Symbols: **c** conventional DMARDs**; +** TNF-inhibitors; ● alternative biologic (using a different mode of action)

* Main decision problem was the cost-effectiveness of first-line biological agents, the model was subsequently used to assess the cost-effectiveness of a 2nd or 3rd TNF-inhibitor.

** Undertaken as part of a submission to NICE by the British Society of Rheumatology. Main decision problem was whether to prescribe a TNF-inhibitor for patients who have failed at least 2 cDMARDs or continue with cDMARD; a subsidiary question included the use of sequential TNF-inhibitors.

Figure D1: Treatment sequences evaluated by studies investigating the use of biologics in early rheumatoid arthritis

(studies are ordered alphabetical)

Treatments indicated in brackets are those that patients entering the model are assumed to have already received. Treatments listed in square brackets indicate those that were compared as part of the economic evaluation, so only one of these are included in each modelled sequence.

Chen, 2006: investigated **adding** a TNF-inhibitor to a sequence of DMARDs as the 1st, 3rd or last active drug **(early vs late** introduction of the TNF-inhibitor) and 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**
W*hen TNF(s) added as the third drug it replaced SSZ+HCQ+MTX*

Davies, 2009: Compared sequences **starting** with 1 to 2 consecutive TNF-inhibitors vs a conventional DMARD
MTX - MTX+HCQ - LEF – gold
TNF I - MTX+HCQ - LEF – gold
TNF I - **TNF II** - MTX+HCQ - LEF - gold

Finckh, 2009: investigated **early vs late** introduction of TNF-inhibitors
NSAIDs - 3 DMARDs – 3 **TNF**s
3 DMARDs – 3 **TNF**s
3 **TNF**s - 3 DMARDs
Each treatment represents three consecutive treatments

Kobelt, 2011: Compared 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: investigated **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*

Tanno, 2006: compared sequences **with and without** a TNF-inhibitor
(BUC) - MTX - SSZ - MTX+SSZ
(BUC) - **ETA** - MTX - SSZ - MTX+SSZ
Included a patient cohort who had failed a previous course of the DMARD bucillamine

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

TNF-inhibitors: ETA etanercept

Other abbreviations: DMARD disease-modifying anti-rheumatic drugs; TNF-inhibitor tumour necrosis factor α inhibitor; vs versus.

Figure D2: Treatment sequences evaluated by studies investigating the use of biologics in established rheumatoid arthritis, after the failure of previous conventional DMARDs

(*studies are ordered by date reflecting the availability of new treatments over time*)

Treatments indicated in brackets are those that patients entering the model are assumed to have already received. Treatments listed in square brackets indicate those that were compared as part of the economic evaluation, so only one of these are included in each modelled sequence.

Barton, 2004 and Jobanputra, 2002: investigated **adding** TNF-inhibitors to a sequence of DMARDs
(the two studies investigated the same treatment sequences and used identical data sources)
SSZ - MTX - **TNF** - gold - AZA - PEN - HCQ - LEF - CyC – CyC+MTX
SSZ - MTX - gold - AZA - PEN - HCQ - LEF - CyC – CyC+MTX
(*TNF-inhibitor also added as the last drug*)

Brennan, 2004: compared sequences **with and without** a TNF-inhibitor
ETA - gold - LEF - CyC+MTX
- gold - LEF - CyC+MTX

Welsing, 2005: compared sequences starting with a TNF-inhibitor vs a conventional DMARD
LEF - usual treatment;
TNF - usual treatment;
LEF - **TNF** - usual treatment;
TNF - LEF - usual treatment;
Treatments used as part of usual treatment were not specified

Bansback 2005: Comparison of ADA, TNF-inhibitors [ETA, INF, ETA] and non-biologics. Patients assumed to only receive one biologic, after which they could receive up to 3 more DMARDs (same fixed sequence for all comparators), the latter being considered as part of palliative care
- LEF – 2 consecutive conventional DMARDs
[**ADA**, ETA, INF, ETA] - LEF – 2 consecutive conventional DMARDs
Generic estimate used for DMARDs

Coyle, 2006: investigated **adding** TNF-inhibitors to a sequence of DMARDs
MTX - MTX+SSZ - MTX+SSZ+HCQ - gold
MTX - MTX+SSZ - MTX+SSZ+HCQ - **TNF** - gold
MTX - MTX+SSZ - MTX+SSZ+HCQ - gold - **TNF**
(If MTX discontinued due to toxicity, patients moved straight to gold)

Wailoo, 2006: 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

Russell, 2009: compared various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)
ETA - INF - ADA
ABA - ETA - INF
ETA - **ABA** - INF

Diamantpoulus, 2012: investigated **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

Wu, 2012: investigated **adding** a TNF-inhibitor to the start of two sequences, with and without the biological agent RTX
- gold - LEF - CyC - MTX
ETA - gold - LEF - CyC - MTX
INF - gold - LEF - CyC - MTX
ADA - gold - LEF - CyC - MTX
ETA - RTX - gold - LEF - CyC - MTX
INF - RTX - gold - LEF - CyC - MTX
ADA - RTX - gold - LEF - CyC - MTX

Diamantopoulois, 2014: investigated **adding** a new biologic (non-TNF-inhibitor) to two different sequences of biological agents, depending on patient's tolerance to MTX
For MTX contraindicated population:
CZP - ETA - ADA
TOC - CZP - ETA - ADA
CZP - **TOC** - ETA - ADA
For MTX tolerant population:
CZP - RTX - ETA - ADA - ABA - ADA - INF
TOC - CZP - RTX - ETA - ADA - ABA - ADA - INF
CZP - **TOC** - RTX - ETA - ADA - ABA - ADA - INF

Tran-Duy, 2014: compared sequences with and without biological agents. The 2 TNF-inhibitors randomly chosen from ETA, ADA, INF), GOL, CZP; and the 2 non-TNF-inhibitor biologics from RTX, ABA, TOC
MTX - SSZ or LEF - AZA - CyC - CYC - HCQ – gold
MTX - SSZ or LEF – TNF I - TNF II - non-TNF I - non-TNF II - AZA - CyC - CYC - HCQ – gold

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

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

Other biological agents targeting different proteins: ABA abatacept; RTX rituximab; TOC tocilizumab

Other abbreviations: DMARD disease-modifying anti-rheumatic drugs; TNF-inhibitor tumour necrosis factor α inhibitor; vs versus.

Figure D3: Treatment sequences evaluated by studies investigating the use of biologics in patients with an inadequate response to ≥1TNF-inhibitor
(studies are ordered alphabetically)

Treatments indicated in brackets are those that patients entering the model are assumed to have already received. Treatments listed in square brackets indicate those that were compared as part of the economic evaluation, so only one of these are included in each modelled sequence.

Beresniak, 2011: compared 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: compared 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: compared sequence of conventional DAMRDs with and without TNF-inhibitors. 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 conventional DMARDs*) - ongoing conventional DMARDs
(≥*2 conventional DMARDs*) - TNFI - ongoing conventional DMARDs
(≥*2 conventional DMARDs*) - TNF I - **TNF II** - ongoing conventional DMARDs
Rather than specifying particular conventional DMARDs at different positions, a generalised DMARD was used. The number of treatments included in the sequences were not stated, but it was noted that after the sixth treatment in each arm, it was assumed that patients would no longer respond but still receive some maintenance therapy on conventional DMARDs.

Cimmino 2011: compared 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: investigated **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]**
Both baseline sequences implemented with and without ETA, representing 1-2 consecutive TNFs.

Hallinen, 2010: investigated **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
Assumed all patients entering model have had an IR to one TNF inhibitor. The sequence 'gold - CyC - MTX' was described as BSC.

Kielhorn, 2008: investigated **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*)
Assumed all patients entering model have had an IR to one TNF inhibitor. MTX was described as palliative treatment

Lindgren, 2009: investigated **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

Malottki, 2011: investigated **adding** a biological agent to sequences, representing the comparison of second-line biological agents
(TNF) [**ADA, ETA, INF, RTX, or ABT**] - LEF - gold - CyA - AZA
(TNF) - LEF - gold - CyA - AZA

Assumed all patients entering model have had an IR to one TNF inhibitor.

Merkesdal, 2010: investigated **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
Assumed all patients entering model have had an IR to ETA, a TNF inhibitor.

Puolakka, 2012: compared 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

Saraux, 2010: compared various biological treatment sequences (included TNF-inhibitor vs a non-TNF-inhibitor)
ETA - **ABA** - ADA
ETA - **RTX** - ADA
ETA - ADA - **ABA**
ETA - ADA - INF

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

TNF-inhibitors: ADA adalimumab; ETA etanercept; CZP certolizumab pegol; INF infliximab.
Other biological agents targeting different proteins: ABA abatacept; ANA anakinra; RTX rituximab; TOC tocilizumab

Other abbreviations: BSC best supportive care (palliative care); DMARD disease-modifying anti-rheumatic drugs; IR inadequate response; TNF-inhibitor tumour necrosis factor α inhibitor; vs versus.

D2 Assessment of how treatments administered before and after the decision point to interest were accounted for in modelling studies of biological agents for rheumatoid arthritis

The included modelling studies of rheumatoid arthritis used a range of different starting points in the treatment pathway, even when considering the same decision problem (Table D1). The extent to which subsequent treatments, after the decision point, or the point at which the modelled sequences diverged also varied.

This next section evaluates the type of evidence and data sources used to inform treatment effects at two different points in the treatment pathway, mainly:

i. The initial treatments used prior to starting TNF-inhibitors (or prior to the decision point it the model)
ii. Conventional DMARDs used after the failure of biological DMARDs

Twenty-one studies considered a decision problem relating to the sequential use of TNF-inhibitors which, represents an important RCT evidence gap (Chapter 5). The variation in the type of data used to address this evidence gap is explored in Chapter 6 (Section 6.6).

In order to avoid repetition, the phrase ‘inadequate response to previous treatment’ is used here 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 (Section C3).

D2.1 The evidence and data sources used to inform the clinical effects of treatments used prior to the decision point

The extent to which the modelling studies accounted for the effects of treatments used prior to the decision point of interest was generally dependant on the starting point of the model (illustrated in Table D1), as opposed to the decision problem; this in turn was influenced by the choice of modelling technique and the available evidence base. The different modelling techniques used are discussed in more detail in Chapter 7. The potential influence of the limited evidence base is discussed here. Three modelling approaches were used to account for, or ‘ignore’, prior treatments, which had an impact on the type of data required.

Decision models starting at the decision point

The most common approach identified was to include a patient population entering the model with a treatment history matching the decision problem in terms of the number of previous treatments used or relevant line of therapy. For example, studies investigating the use of TNF-inhibitors in established rheumatoid arthritis included a patient population who had previously received at least two conventional DMARDs (Figure D2).[114 173 175 180] The data used to inform the effects of the initial modelled treatments (decision point) were generally obtained from pivotal RCTs. The patient population entering the model was usually chosen to reflect those included in these clinical trials, thus matching the licenced indication for the initial treatments. Using this approach means that the model starts at a defined line of therapy, but the specific previous drug sequence used is not considered. It has the advantage of not having to match the exact previous treatments failed by the included participants. However, it assumes that the distribution of previous treatments and disease duration of participants included in these clinical trials are representative of those whom the clinical decision problem relates to, and real-world practice. One study (Wailoo, 2006) compared the use of two separate data sources to inform the probability of response to the first modelled treatment, which included one of four biological agents or a conventional DMARD (Figure D2).[178] In the first base case analysis, data were taken from an accompanying meta-regression of 13 RCTs (Nixon, 2007), which is described in more detail in Chapter 5 (Section 5.2).[194] Disease duration and baseline HAQ score were included in the meta-regression as covariates, but not the number of previous conventional DMARDs used.[194] In the second base case analysis, the probability was derived directly from a patient registry, the National Data Bank (NDB) for Rheumatic Diseases. (A list of patient registries used for evaluating the effects of biological agents for rheumatoid arthritis is provided in the Appendix Volume I, Section C3) Here, the sampled population was restricted to patients on first-line biological

agent. The response rates were found to be substantially lower in the registry data for etanercept and adalimumab, than in the regression data. This study was based on the Sheffield AHRQ model (Chapter 7, Section 7.3.6.2).

The data source used to inform the treatment effect of the initial modelled treatment in some studies was a single RCT. However, the treatment history of the patient population included in the model did not always match those included in the RCT. For example, one study (Tanno, 2006) specified that patients entering the model had failed bucillamine, in order to reflect clinical practice, but the RCT evidence used to inform the effects of the initial treatments being modelled (etanercept or methotrexate) did not match the failure of this previous treatment (Figure D1).[172] The treatment effect of etanercept was obtained from an RCT that included patients who had already received an average of three previous conventional DMARDs (with about 90% having had methotrexate); the treatment effect of methotrexate was taken from a separate referenced study but no further details were provided.

Decision models starting at a discrete point in the pathway

The second approach identified was to model an initial treatment used prior to the decision point, and apply the assumption that the entire patient population on entering the model had an inadequate response to the first modelled treatment, i.e. prior to the divergent point for the sequences being compared. This allowed the initial treatments to be costed appropriately, as part of a treatment sequence. This approach was used by a series of studies using the same modelling technique, referred to as the advanced simulation model (Chapter 7, Section 7.3.2.3),[188] to compare different biological agents used as first (Figure D2) or subsequent-line (Figure D3) treatment. These studies can be identified in Table D1 as those that include a single pale grey shading prior to the decision problem, and in Figure D2 (Russell, 2009)[177] or Figure D3 (Beresniak, 2011, Beresniak, 2013, Cimmino, 2011, Puolakka, 2012, Saraux, 2010).[182 184 185 187 188] Although the models assumed a 100% inadequate response rate to an explicit initial TNF-inhibitor, the potential sequencing effects of prior conventional DMARDs were not considered. The data used to inform the treatment effects were taken from individual pivotal trials or uncontrolled open label follow-up studies. However, the evidence used to inform the treatment effects of the second TNF-inhibitor (divergent point) did not match the specific prior TNF-inhibitor that failed (first treatment modelled). I come back to this when discussing the evidence used to inform sequencing effects of TNF-inhibitors below.

Decision models starting at the point of diagnosis

In the third approach the whole treatment pathway was modelled, starting from the initial diagnosis. Five studies (Clark, 2004; Barton, 2004; Chen, 2006; Jobanputra, 2002; Tran-Duy, 2014) used individual patient simulation to developed the decision population within the actual model,[99 169 174 181 183] and one study (Coyle, 2006) used Markov cohort modelling.[179] These studies can be identified in Table D1 as those that include pale grey shading starting from the fourth column representing

conventional DMARDs for early rheumatoid arthritis. The modelling methods are described in more detail in Chapter 7.

Four studies (Clark, 2004; Barton, 2004; Chen, 2006; Jobanputra, 2002) were based on the Birmingham Rheumatoid Arthritis Model (BRAM) discussed in Chapter 7, Section 7.3.6.4.[169 174 181 183] Newly diagnosed patients entering the model, with early disease, would proceed to follow one of two predefined sequences of treatments based on a randomly generated draw. However, the model outcomes would only be collected from a point at which the sequences diverged, based on the decision problem of interest, for example the addition of a TNF-inhibitor after inadequate response to the first two conventional DMARDs. Data for the first two conventional DMARDs were obtained from an RCT of each treatment in early rheumatoid arthritis. One study (Tran-Duy, 2014) used an individual sampling model to implement two separate treatment strategies, with and without TNF-inhibitors (Tran-Duy Model, Section 7.3.6.4).[99] All patients entering the model were newly diagnosed, and only became eligible for a TNF-inhibitor after the failure of two consecutive conventional DMARDs. All patients initially received methotrexate followed randomly by either sulphasalazine or leflunomide (Figure D2). The treatment effects for the model were taken from the Nijmegen rheumatoid arthritis inception cohort in the Netherlands (Appendix Volume I, Section C3), which included patients with early disease and no prior DMARD use. The final study, which was based on a Markov cohort model, investigated adding one of two TNF-inhibitors to a predefined sequence of conventional DMARDs.[179] The initial part of the model prior to the decision point, which was not considered to be the focus of the analysis, was based on a published model by Maetzel *et al*., which investigated adding leflunomide to a sequence of conventional DMARDs (Maetzel model, Section 7.3.3.3).[195] Treatment effects used to develop the parameters for this earlier model were obtained from a systematic review of individual treatments, which included RCTs and observational studies for four conventional DMARDs. Treatment sequencing effects were not considered.

D2.2 The evidence and data sources used to inform the clinical effects of conventional DMARDs used at a later stage in the sequence or after biological agents

The main issues regarding the evidence base for conventional DMARDs used in the later part of modelled sequences is that there is limited data on the effects of these drugs used in established arthritis, or after an inadequate response to a biological agent. The data sources and methods used to obtain the treatment effects of conventional DMARDs varied quite considerably in terms of those used as the comparator drug at a specific decision point, and informing the use of these drugs in later stages of included sequences.

The evidence used to inform the effects of a conventional DMARD used as the comparator and the subsequently displaced treatment

Studies that investigated adding a biological agent to the start of an established sequence of conventional DMARDs in early disease (Figure D1) generally modelled methotrexate as both the first-line comparator treatment and the treatment that was subsequently displaced lower down the

sequence. The same effect estimate was used for methotrexate, irrespective of whether it was the comparator or the next treatment, after the failure of a TNF-inhibitor, in three studies (Section 6.5.3.6).[168 169 172]

In the studies of late stage disease, where a biological agent was introduced after the failure of previous DMARDs, the control, and displaced drug, tended to be gold therapy (Figures D2 and D3).[173 174 176 179 181 183 189] Different data sources and approaches were used to inform the treatment effects of gold in established disease. For example, one study (Brennan, 2004) which modelled etanercept or gold as the first modelled treatment (Figure D2) noted that there were no published studies of gold that matched the patient population in the etanercept RCT in terms of disease duration.[176] The treatment response for gold was therefore based on a pooled generic estimate taken from a meta-analysis of numerous conventional DMARDs in patients with a disease duration greater than 10 years (discussed in Section 6.5.3.2). Another study (Coyle, 2006) used data from a published RCT of gold injections versus auranofin, an orally administered gold, but no details were given on the patient population or previous treatments (Figure D2). Another example is provided by a study using the Birmingham Rheumatoid Arthritis Model (Barton, 2004) to evaluate the introduction of a TNF-inhibitor after the failure of conventional DMARDs (Figure D2). This study (Barton, 2004) used a substitution effect for gold taken from sulfasalazine.[174] This was obtained from an RCT of leflunomide versus sulfasalazine in patients with early disease who had not previously received sulfasalazine. In another study that used the Birmingham Rheumatoid Arthritis Model (Clark, 2004) to investigate adding a new biological agent after the failure of both conventional DMARDs and TNF-inhibitors, but before gold (Figure D3),[183] the treatment effect for gold was informed by a prospective cohort study, which included participants with early rheumatoid arthritis treated with gold over a five year period. The number of previous conventional DMARDs used by those who were recruited ranged from zero to three. Finally one study (Hallinen, 2010), which included a patient population with an inadequate response to a previous TNF-inhibitor (Figure D3), assumed that the treatment effect of the subsequent gold was same as methotrexate used as the control in RCTs of biological agents.[189] However, this study also noted that the sequence of three conventional DMARDs used after TNF-inhibitor represented best supportive care.

The evidence used to inform the effects of conventional DMARD used later in the sequence

A range of different approaches were used to select the data sources to inform the treatment effects for conventional DMARD sequences used after the decision point, which generally equated to a patient population with late stage disease. The use of conventional DMARDs after the failure of methotrexate and several biological agents can be conceived as representing palliative or best supportive care, and therefore assumed to have limited effects.[196] Nineteen studies of rheumatoid arthritis investigated treatment sequences where conventional DMARDs were used in the later part of the included sequences, and usually after the failure of biological agents (Table D1). The three main data sources that were identified are described briefly below in order to show the wide variation in the approaches used.

The use of data from studies of conventional DMARDs used in early disease

Six studies used data from RCTs of conventional DMARDs used in early disease to inform the treatment effects of these drugs used later in the sequence.[114 169 172 174 179 181 183 191] Three studies (Chen, 2006; Clark, 2004; Malottki, 2011) adjusted the treatment effects to account for their use in later disease (Figures D1 and D3, respectively) as discussed at the end of Section 6.5.3.3.[169 183 191]

The use of data from the methotrexate-placebo control arm of RCTs of TNF-inhibitors in late disease

Pivotal RCTs of TNF-inhibitors for established rheumatoid arthritis generally included methotrexate as part of the placebo control. Patients with an inadequate response to conventional DMARDs, including methotrexate, were randomised to a TNF-inhibitor plus methotrexate or placebo plus methotrexate (Appendix Volume I, Section C3.3). Methotrexate was therefore frequently used as the reference treatment in network meta-analyses of TNF-inhibitors. Three studies used a summary effect estimate for methotrexate, taken from such meta-analyses to inform the treatment effects of subsequent conventional DMARDs.[95 189 192] The meta-analysis conducted by Kielhorn *et al*. was based on four pivotal RCTs, which included an RCT of leflunomide versus methotrexate in patients who were methotrexate-naïve, two RCTs of TNF-inhibitors in patients who had previously failed at least two conventional DMARDs, and one RCT of an alternative biological agent, rituximab, in patients who had inadequate response to a TNF-inhibitor and prior conventional DMARDs.[192] Clearly the patient population in these trials varied considerably in terms of both disease duration and treatment history, which was reflected in heterogeneity in the baseline rates. Treatment responses were therefore adjusted using a reference placebo response rate, which was based on the weighted average effect for methotrexate in all four trials.

The use of data obtained from patient registries of early or late disease

An alternative data source used to inform the treatment effects of conventional DMARDs was patient registries,[99 168 170 171 175 176] which were also used to obtain individual patient level data.[99 168 175 176] Three studies obtained data from a registry held by the South Swedish Arthritis Treatment Group (SSATG), which included patients with established rheumatoid arthritis treated with leflunomide, etanercept or infliximab, and who had failed on average four previous conventional DMARDs.[168 175 176] Two studies (Schipper, 2011 and Tran-Duy, 2014) used data from the Nijmegen rheumatoid arthritis inception cohort in the Netherlands, which included patients with early disease and no prior DMARD use.[99 171] Finally, one study (Kobelt, 2011) used individual patient data obtained from an observational study of patients in Malmo, Sweden, to inform the effects of conventional DMARDs, representing standard treatment, used later in the modelled sequences (Figure D1).[170] Interestingly the same study obtained data for patients switching biological agents from the South Swedish Arthritis Treatment Group (SSATG) registry.

E. AN EVALUATION OF THE KEY FEATURES OF MODELLING TECHNIQUES TO INFORM THE ASSESSMENT OF INCLUDED MODELLING STUDIES

E1 Introduction

This Appendix provides an overview of the key features of different modelling techniques, which are likely to impact or inform the choice of which approach to use for modelling treatment sequencing. This was based primarily on two published taxonomies of modelling techniques[197 198] and was used to inform the review of modelling studies. The finding of the review of modelling approaches is presented in Chapter 7.

E2 Taxonomies of modelling approaches

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. This section describes the two published taxonomies developed for categorising different modelling techniques according to their key features.[197 198] The key features of modelling techniques described in these taxonomies are described in more detail in Section E3.

A taxonomy of 14 different modelling techniques for economic evaluation of health technologies has been developed by Brennan *et al*.[197] This is presented in Table E1. It illustrates the relationship between the different modelling techniques, where each grid cell in the taxonomy is related to its neighbour by varying some of the basic assumptions that underlie each model type. The columns (A-D) separate cohort from individual level models, and disentangle the assumptions about the role of expected values, randomness, the heterogeneity of entities, ~~a~~nd the degree of non-Markovian structure. The rows (1-4) describe features involving both time and interactions between individuals, and distinguish between discrete time and continuous models. Brennan *et al*. noted that most modelling techniques currently used by health economists fall into the top half of the table,[197] that is, models that assume independence between individuals, and where time may or may not be modelled explicitly. Such models include cohort based deterministic and simulated decision tree modelling, as well as simulated patient-level decision tree modelling, and both deterministic and individual-level simulation for Markov modelling. Interaction between patients is not considered an important issue for modelling treatment sequences as such. This is generally more pertinent for modelling infectious diseases, or when the choice of treatment for one patient affects what can be given for another, for example in organ transplantation.[199] However, the distinction between discrete and continuous time models is relevant, and models using interactions, such as discrete event simulation have been advocated for modelling chronic conditions and treatment sequences.[199-202]

Table E2: Taxonomy of model structures published by Brennan, 2016

			A	B	C	D
			Cohort/Aggregate Level/Counts		**Individual Level**	
			Expected value, Continuous state, Deterministic	**Markovian, Discrete State, Stochastic**	**Markovian, Discrete State, Individuals**	**Non-Markovian, Discrete-State, Individuals**
1	**No Interaction Allowed**	**Untimed**	Decision Tree Rollback	Simulated Decision Tree (SDT)	Individual Sampling Model (ISM): Simulated Patient-Level Decision Tree (SPLDT)	
2		**Timed**	Markov Model (Evaluated Deterministically)	Simulated Markov Model (SMM)	Individual Sampling Model (ISM): Simulated Patient-Level Markov Model (SPLMM) (variations as in quadrant below for patient-level models with interaction)	
3	**Interaction Allowed**	**Discrete Time**	System Dynamics (Finite Difference Equations, FDE)	Discrete Time Markov Chain Model (DTMC)	Discrete-Time Individual Event History Model (DT, IEH)	Discrete Individual Simulation (DT, DES)
4		**Continuous Time**	System Dynamics (Ordinary Differential Equations, ODE)	Continuous Time Markov Chain Model (CTMC)	Continuous Time Individual Event History Model (CT, IEH)	Discrete Event Simulation (CT, DES)

Taken from: Brennan, A., S. E. Chick, et al. (2006) A taxonomy of model structures for economic evaluation of health technologies. Health Economics, 15(12), 1295-1310.

The structural features used within the taxonomy presented by Brennan *et al*.[197] are also included in a framework developed by Kim and Goldie for categorising theoretical models used for evaluating the cost-effectiveness analyses of vaccination programmes.[198] This is presented in Table E2. Here features were divided into six domains, the first three were reported as being central for categorising model structures, whilst the remaining three were more related to the question of how analysts measure or compute model outputs. The six domains included whether:[198]

i. The main features of the model change over time (dynamic) or not (static)
ii. The changes in the model occur randomly (stochastic or probabilistic) or the rules of changes are pre-specified (deterministic)
iii. The population's behaviour in the model is simulated using aggregate variables of which values are population averages (aggregate) or the behaviours of individuals in the population are tracked (individual based)
iv. Events are assumed to occur at a discrete time interval (discrete) or at a point on a continuum (continuous)
v. The model allows individuals to continually enter the model (open) or not (closed)
vi. The model is expressed in equations that are functions of linearly linked parameters (linear) or not (non-linear)

Table E2: A framework for categorising theoretical models published by Kim & Goldie, 2008

		STATIC	DYNAMIC
DETERMINISTIC	**Aggregate level (Compartmental / cohort)**	***TYPE 1*** Deterministic aggregate-level static model 1.1. Decision trees 1.2 State-transition models (e.g. Markov model) 1.3 Hybrid models (e.g. a decision tree embedded within Markov models)	***TYPE 2*** Deterministic aggregate-level (compartmental) dynamic model 2.1 Discrete difference equations model (discrete time) 2.2 ODE model (continuous time) 2.3 PDE model (continuous time) 2.4 Other types of models that allow for interaction e.g. transmission dynamics at the aggregate level
	Individual level	(Not impossible, but uncommon[a])	(Not impossible, but uncommon[a])
STOCHASTIC (PROBABILISITIC)	**Aggregate level (compartmental / cohort)**	***TYPE 3*** Stochastic aggregate-level static model e.g. Monte Carlo simulation (sampling of outcomes) of a decision tree or a state-transition model	***TYPE 4*** Stochastic aggregate-level dynamic model e.g. individual sampling of compartmental dynamic model
	Individual level	***TYPE 5*** Static microsimulation model e.g. Monte Carlo microsimulation (individual sampling) of a decision tree or a state-transition model	***TYPE 6*** Dynamic microsimulation model 6.1 Monte Carlo simulation (individual sampling) of a Markov model with interaction 6.2 Discrete-event simulation model 6.3 Agent-based model e.g. transmission dynamics at the individual level

Taken from: Kim, S.Y. and Goldie S.J. (2008) Cost-effectiveness analyses of vaccination programmes: a focused review of modelling approaches. Pharmacoeconomics, 26(3), 191-215.
Comment: Using classification along several dimensions, this figure presents general types of (mathematical) models that can be used in projecting the health and economic consequences of vaccination programmes.
a Most individual-based models are stochastic.
Abbreviations: ODE ordinary differential equations; PDE partial differential equations.

Some of the key features used by Brennan *et al*.[197] and Kim and Goldie[198] for categorising different modelling techniques are discussed in more detail in the next section. This includes an evaluation of their significance in choosing the most appropriate method for modelling treatment sequencing. The extent to which they actually influenced the selection of modelling approaches used in practice was then evaluated as part of the review of included modelling studies, presented in the subsequent section.

E3 Brief overview of the key features of modelling approaches

E3.1 Event versus state

One key difference between modelling techniques is how they conceptualise what is happening to the target population. The decision problem may be represented as a series of events that the patients experience or a series of states that the patients move through.[202] Thus treatment

sequencing could potentially be conceptualised as a series of events that trigger treatment switching, or health states representing each treatment that patients progress through.

E3.2 Markovian property

State-transition models are often synonymously referred to as 'Markov models'. A defining characteristic, of Markov models is that the future development of the process is not dependent on the history of the process, just on the present.[203] This is known as the Markov property, or Markovian assumption, and means that the probability of a patient transiting to another state is dependant only on the state in which the patient is in at the start of the model cycle, and is not allowed to depend on either the time a patient has spent in a given state, or the patient's previous history before entering that state.[199 204] Another requirement of a Markov model is that that patients are only in one state at any given time.[205]

The Markovian property could potentially impede the modelling of treatment sequences, as it implies that treatment success is independent of the number or type of previous treatments, and that the time spent in a given health state or on a specific treatment, is not determined by previous treatment history, which is likely to be unrealistic. However, there are a number of ways of introducing 'memory' to the standard Markov cohort model. This can be achieved by using temporary states that patients can only enter for one cycle, or by a series of temporary states that must be visited in a fixed sequence, known as tunnel states.[206] However, if extensive history and or time-dependency is required, a large number of tunnel states may be necessary,[207] and Markov models with multi states can-become non-transparent and very complicated,[202] as the number of states increases exponentially.[208] It also increases the difficulty of estimating transition probabilities for the different subsets of the cohort with different prognoses, for whom temporary states have been developed.[209] A newer representation of the Markov model, which could aid the implementation of treatment sequences, is the Markovian cycle-tree,[210] where events that can occur within a cycle are modelled as a series of chance nodes.[211]

E3.3 Discrete versus continuous

Modelling approaches can be distinguished by their time framework, which can be either discrete or continuous. How time is interpreted in the model is likely to be an important consideration for modelling treatment sequences. The extent of the challenge, of modelling timing and duration of clinical events, as well as treatment sequencing, will vary according to the type of treatment and disease being investigated and the extent of the treatment sequences being considered. The simplest form of state-transition models are discrete time (homogenous) Markov chains, where a single fixed-transition matrix is applied at every cycle.[141] The transition probabilities are assumed to occur at discrete intervals and to be constant over time. A semi-Markov process model is an example of a continuous time formulation of a state-transition model. A Markov decision process is a Markov chain in which state-transitions depend on both the current state and an action vector that is applied to the system. In other words, they can account for the time a patient has spent in a given state, but

not the patient's previous history before entering that state (Markovian assumption). A semi-Markov framework can therefore be used to circumvent dependency on time in a specific state.[203] Long-term health interventions or complex, time-varying courses of treatments may be better captured using Markov process-based methods, which permit a more straightforward and flexible sequencing of outcomes, including recurring outcomes through time.[204 212] However, they are not yet widely applied in health economic evaluation to assess the cost-effectiveness of alternative health care interventions.[199 213] One of the main reasons why continuous time models are rarely employed in practice is that closed form solutions for the expected time in the states often do not exist and, where they do exist, they can be mathematically demanding as, for example, they are not constant through time.[203] The need to identify an appropriate continuous function is an important disadvantage. An alternative approach is to emulate continuous time using shorter discrete time steps to model.

In the taxonomy presented by Brenan *et al*. (Table E1), which distinguished between discrete and continuous time models, Markovian distributions include exponential distribution for continuous time models and geometric distribution for discrete time models.[197] Individual level approaches based on non-Markovian distributions, such as discrete event simulation, can be implemented as either continuous or discrete time models.

E3.4 Cohort versus continuous

Both Brenan *et al*.[197] and Kim and Goldie[198] make a clear distinction between cohort and individual based models. The choice between the two relates to whether the model should seek to characterise the experience of the 'average' patient from a population sharing the same characteristics, or whether the aim is to explicitly consider the individual patient and allow for variability between patients.[207] In macrosimulation models, the unit of analysis is a hypothetical cohort,[214] which needs to be homogenous, as the decision problem is to establish the most cost-effective option for specifically defined groups of individuals.[207] Microsimulation models, on the other hand, simulate one individual at a time; the progression of potentially heterogeneous individuals is tracked through the model and an average effect over all the values is developed.[204] The individual patients are subject to the same probability of transition as the cohort of patients[205] and, providing the number of simulations over which the results are averaged is very large, the mean value of this distribution will be similar to the expected utility obtained by a cohort simulation.[215] A large number of simulated patient histories are generally used in order to limit Monte Carlo error, and the results evaluated using sampling algorithms.

Individual based models can be used to account for different treatment sequencing effects in various subgroups. Cohort models can also account for different patient covariates, for example stages of natural history of disease, by subdividing the number of states in a state-transition model or branches in a decision tree. However, the number of dimensions will rise exponentially with each attribute, for example, M binary attributes imply 2^m dimensions.[197] Another important advantage of using individual patient simulations for modelling treatment sequences is that the patient's future in the model is

conditional on their past, and progression to a certain treatment can be allowed if a specific condition, for example the number of previous treatments or disease level, is met.[216] The fact that the model is tracking each individual's past, also means that it is possible to allow transition probabilities to vary according to the time patients have spent in a particular state, or the total time for which they have been ill.[199 207]

E3.5 Deterministic versus stochastic

Both Brenan *et al*. and Kim and Goldie distinguish between deterministic and stochastic models. Deterministic models are based on single point estimates of each parameter, and thus assume certainty in all aspects, and will always produce the same output for a given input, or set of initial conditions. Stochastic models, on the other hand, represent uncertainty by using ranges in the form of probability distributions, rather than fixed values. The use of simulation enables the model to take into account the entire distribution of a predefined parameter according to specific distribution laws.[188] It can be argued that the relationships between actions and outcomes in clinical practice are usually more probabilistic than deterministic, where chance alone can turn a good decision into a poor outcome, or vice versa.[217] Incorporating randomness, using simulation methods such as Monte Carlo simulations, allows the model to mimic reality and to capture the uncertainly that occurs when the true value of a parameter is not known, due to imperfect knowledge or measurement.[197 218]

Individual patient-level simulation models are naturally stochastic, where a random number is used to determine the path of each individual within the model, however, the model parameters, such as response rates and transition probabilities are fixed.[198 219] The implementation of random variation here is known as *first order simulation*, which propagates variability and heterogeneity into the model, but not uncertainty in the individual parameters or variables. Random variation can be implemented not just at the level of the patient characteristics, representing patient heterogeneity, but also at the decision node, state-transition, or time-to-event level. However, many replications with different random numbers, or simulation runs, are likely to be needed, to quantify the output mean and variance with sufficient accuracy,[197] and thus reduce stochastic uncertainty. A Markov cohort model would only need to be run once to generate expected values, whilst an individual patient-level simulation requires multiple iterations to develop accurate and precise estimates.

E3.6 Representing parameter uncertainty

Uncertainty arising from the limited evidence base regarding treatment sequencing effects will result in parameter uncertainty. This relates to the uncertainty in the model data inputs. Parameter uncertainty reflects the fact that we can never know for certain what the mean expected costs and effects would be if the treatment is provided for a particular population of patients.[220 221] It can arise from a number of sources,[220 221] such as heterogeneity and sampling error within the data used to derive the parameter estimate, the existence of multiple conflicting studies, generalisability of the data to a real-world setting, and the lack of empirical data.[222]

Sensitivity analyses are used for assessing the level of uncertainty associated with the optimum decision, due to the parameter uncertainty. These can include standard deterministic sensitivity analysis, where each variable is varied separately and independently, or probabilistic sensitivity analysis, in which all variables are varied simultaneously using probability distributions informed by estimates of the sample mean and sampling error from the best available evidence. Deterministic sensitivity analysis does not give a complete picture of the effects of joint uncertainty and correlation between variables.[223 224] Probabilistic sensitivity analysis is therefore considered the preferred way of assessing or characterising parameter uncertainty, and recommended by NICE.[225 226] The use of probabilistic sensitivity analysis is especially recommended for non-linear models or multilinear models with correlated parameters.[141 219 225]

The uncertainties in the inputs to a deterministic model are external, and can be handled by analysing the model using a Monte Carlo simulation, which is also known as *second-order simulation* or probabilistic sensitivity analysis. However, this does not make the model stochastic. Probabilistic sensitivity analysis has also been described as probabilistic decision analysis.[221]

For individual patient simulation, conducting formal probabilistic sensitivity analysis can conflict with the computationally expensive structures, as it requires a second level of simulation.[219] The need for probabilistic sensitivity analysis[225] has therefore been used as the justification for using a cohort model instead.[219 227] Implementing probabilistic sensitivity analysis within a patient-level simulation requires two nested simulation loops: an inner loop that evaluates the outcomes across the simulated population, and an outer loop that samples the parameter values to reflect uncertainty in the model inputs.[141] Although the increased power of modern computers and specialist software can facilitate this, it may still require specialist training.[141 219] The computation time can also be reduced by increasing the efficacy of running the inner loop. The use of discrete event simulation rather than a state-transition model may also be more efficient.

E3.7 Dynamic versus static, and open versus closed populations

The framework presented by Kim and Goldie distinguishes between static and dynamic models. Dynamic models directly incorporate time into the structure. Whilst static models try to show what happens over time, or as time passes, but time itself is not represented or embodied directly into the model. Static models typically focus on a single cohort that ages as it progresses through the model, whilst dynamic models are run over many years on the basis of multiple cohorts.[228] These can be closed models, defined as models that simulate those cohorts present in the initial period, or open models, which allow new cohorts to enter over time.[229] Most health economic models are static, with changes in the population's risk of progression, or recurrence, over time characterised as a fixed state or distribution for a single cohort.

The distinction between a static and a dynamic model is considered important for infectious diseases because the rate of infection in the modelled population is unlikely to be fixed.[198 207] A similar

argument may also hold for modelling treatment sequences, as the impact on disease-severity or progression over time may be variable. Patients can also experience an event, such as lack of efficacy or side effects of treatment, at any point during the course of a treatment or disease, and the probability of these events may change with time. When using a static framework, disease-severity is modelled as a constant over time. However, the patients' distribution over states of disease-severity is likely to vary as a result of treatment effect or disease progression.[230] In a dynamic model, the probability of a patient having disease progression or recurrence would be dependent on the patient's current treatment and the evolving clinical course of the disease over time. Treatment in clinical practice could also potentially be viewed as a dynamic process, which involves decisions made sequentially over time based on accruing observations on the patient.[217] This was first highlighted in Chapter 1, Section 1.2.1.

Treatment sequences can be characterised within a model as a fixed sequence of treatments that the patients progress through or, alternatively, can be developed as part of the modelling process using random draws. Modelling fixed treatment sequences may not reflect the dynamic real-life clinical decision process, but the choice of next treatment is also rarely random. The modelling 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 dynamic decision making should not be included as part of the state-transition models, but in the Markov decision tree.[211] Discrete event simulation and agent based modelling are categorised as dynamic microsimulation models within the framework presented by Kim and Goldie.[198] They also allow the flexibility of modelling both open and closed populations.[231]

Open models, which allow new cohorts to enter over time, are sometimes referred to as population models.[229] The open approach can be used to represent an ongoing intervention program, and is often the basis for budget impact calculations.[232] The closed approach, where members enter only at the beginning of the model, is reported to correspond more closely to the medical sector perspective and is often used in health technology assessments.[232] Hypothetical populations can also be generated by combining multiple simulated cohorts.

REFERENCES FOR APPENDIX VOLUME I

1. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *Journal of Clinical Nursing* 2003;12(1):77-84.
2. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evidence-based medicine* 2016;21(4):125-7.
3. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.
4. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials *BMJ* 2011;342:d4002
5. Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Statistical Methods in Medical Research* 2005;14(5):515-24.
6. Rubin DB. For objective causal inference, design trumps analysis. *Annals of Applied Statistics* 2008;2(3):808-40.
7. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999;319(7205):312-5.
8. 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.
9. Wood L, Egger M, Gluud LL, Schulz K, Juni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008;336:601
10. Page M, Higgins JPT, Clayton G, Sterne JAC, Hróbjartsson A, Savović J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PLoS One* 2016;11(7):e0159267.
11. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. Chapter 21: Notes on Subgroup Analyses and Meta-Regression. In: Borenstein M, Hedges LV, Higgins JP, Rothstein HR, editors. *Introduction to Meta-Analysis (eds)*. Chichesater, UK: Wiley, 2009.
12. Waddington H, White H, Snilstevit B, Hombrados JG, Vojtkova M, Davies P, et al. How to do a good systematic review of effects in international development: a tool kit *Journal of Development Effectiveness* 2012;4(3):359-87.
13. 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.
14. 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.
15. 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.
16. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to Meta-Analysis*. Chichesater, UK: Wiley, 2009.
17. Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value in Health* 2011;14(4):417-28.
18. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005;365(9453):82-93.

19. 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.
20. 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.
21. Singh S, Loke YL. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13(1):138.
22. Suvarna V. Phase IV of Drug Development. *Perspectives in clinical research* 2010;1(2):57-60.
23. Ford I, Norrie J. Pragmatic Trials. *New England journal of medicine* 2016;375(5):454-63.
24. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in medicine* 2005;24(10):1455-81.
25. 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.
26. 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.
27. 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.
28. Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data *Clinical Pharmacology & Therapeutics* 2016;99:262-5.
29. Schneeweiss S. Learning from big health care data. *New England Journal of Medicine* 1014;370:2161-3.
30. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, et al. Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *Journal of Clinical Epidemiology* 2018;97:26-34.
31. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open* 2018;8:e019703.
32. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials. In: Chandler J, McKenzie JE, Boutron I, Welch V, editors. *Cochrane Methods Cochrane Database of Systematic Reviews* 2016:29–31.
33. 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.
34. Page MJ, Boutron I, Hansen C, Altman DG, Hróbjartsson A. Assessing risk of bias in studies that evaluate health care interventions: recommendations in the misinformation age. *Journal of Clinical Epidemiology* 2018;97(133-136).
35. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *BMJ* 2016;353:i2016.
36. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *Journal of Clinical Epidemiology* 2013;66(2):151-7.
37. Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE Working Group.GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. . *Journal of Clinical Epidemiology* 2018;pii:S0895-4356(17)31031-4 [Epub ahead of print].
38. 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.
39. 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.
40. Chakraborty B. Dynamic treatment regimes for managing chronic health conditions: a statistical perspective. *American Journal of Public Health* 2011;101(1):40-5.
41. Braun MS, Richman SD, Thompson L, Daly CL, Meade AM, Adlard JW, et al. Association of molecular markers with toxicity outcomes in a randomized trial of chemotherapy for advanced colorectal cancer: the FOCUS trial. *Journal of Clinical Oncology* 2009.
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. Almirall D, Nahum-Shani I, Sherwood NE, Murphy SA. Introduction to SMART designs for the development of adaptive interventions: with application to weight loss research. *Translational Behavioral Medicine* 2014;4(3):260-74.
48. 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.
49. Schneider LS, Tariot PN, Lyketsos CG, Dagerman KS, Davis KL, Davis S, et al. National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE): Alzheimer disease trial methodology. *American Journal of Geriatric Psychiatry* 2001;9(4):346-60.
50. 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.
51. 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.
52. Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine* 2000;19(8):1011-28.
53. Thall P, Logothetis C, Pagliaro L, Wen S, Brown M, Williams D, et al. Adaptive Therapy for Androgen-Independent Prostate Cancer: A Randomized Selection Trial of Four Regimens. *Journal of the National Cancer Institute* 2007;99(21):1613-22.
54. Wahed AS, Tsiatis AA. Optimal estimator for the survival distribution and related quantities for treatment policies in two-stage randomization designs in clinical trials. *Biometrics* 2004;60(1):124-33.
55. Thall PF, Wooten LH, Logothetis CJ, Millikan RE, Tannir NM. Bayesian and frequentist two-stage treatment strategies based on sequential failure times subject to interval censoring. *Statistics in Medicine* 2007;26(26):4687-702.

56. Lokhnygina Y, Helterbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics* 2007;63(2):422-8.
57. Lee J, Feng L. Randomised phase II designs in cancer clinical trials: current status and future directions. *Journal of Clinical Oncology* 2005;23(19):4450-7.
58. Mander AP, Thompson SG. Two-stage designs optimal under the alternative hypothesis for phase II cancer clinical trials. *Contemporary Clinical Trials* 2010;31(6):572-8.
59. Banerjee AT, A. A. Adaptive two-stage designs in phase II clinical trials. *Statistics in medicine* 2006;25(19):3382-95.
60. Collins LM, Murphy SA, Bierman KL. A conceptual framework for adaptive preventive interventions. *Prevention Science* 2004;5(3):185-96.
61. Thorlund K, Haggstrom J, Park JJH, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians *BMJ* 2018;360:k698.
62. Chakraborty B, Murphy SA. Dynamic Treatment Regimes. *Annual Review of Statistics and Its Application* 2014;1:447-64.
63. Beckman RA, Antonijevic Z, Kalamegham R, Chen C. Adaptive Design for a Confirmatory Basket Trial in Multiple Tumor Types Based on a Putative Predictive Biomarker. *Clin Pharmacol Ther* 2016;100(6):617-25.
64. 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.
65. Korn EL, Freidlin B. Adaptive Clinical Trials: Advantages and Disadvantages of Various Adaptive Design Elements. *Journal of the National Cancer Institute* 2017;109(6).
66. Berry SM, Connor JT, Lewis RJ. The platform trial: an efficient strategy for evaluating multiple treatments. *JAMA* 2015;313(16):1619-20.
67. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *New England Journal of Medicine* 2017;377(1):62-70.
68. Lin JA, He P. Reinventing clinical trials: a review of innovative biomarker trial designs in cancer therapies. *British Medical Bulletin* 2015;114(1):17-27.
69. Renfro LA, An MW, Mandrekar SJ. Precision oncology: A new era of cancer clinical trials. . *Cancer Letters* 2017;387:121-26.
70. Simon R. Critical Review of Umbrella, Basket, and Platform Designs for Oncology Clinical Trials. *Clinical Pharmacology & Therapeutics* 2017;102(6):934-41.
71. Kaplan R, Maughan TS, Crook A, Fisher D, Wilson R, Brown L, et al. Evaluating many treatments and biomarkers in oncology: a new design *Journal of Clinical Oncology* 2013;31(36):4562-8.
72. James ND, Sydes MR, Clarke NW, Mason MD, Dearnaley DP, Spears MR, et al. Addition of docetaxel, zoledronic acid, or both to first-line long-term hormone therapy in prostate cancer (STAMPEDE): survival results from an adaptive, multiarm, multistage, platform randomised controlled trial. *Lancet* 2016;387(10024):1163-77.
73. Schmoll HJ. FOCUS4: a new trial design for evaluation of targeted drugs in colorectal cancer? *Lancet Gastroenterology & Hepatology* 2018;3(3)(3):143-45.
74. Hey SP, Weijer C, Taljaard M, Kesselheim AS. Research ethics for emerging trial designs: does equipoise need to adapt? . *BMJ.* 2018;360:k226.
75. Mandrekar SJ, Dahlberg SE, Simon R. Improving Clinical Trial Efficiency: Thinking outside the Box. *American Society of Clinical Oncology Educational Book*, 2015:e141-7.
76. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in Multiple Nonmelanoma Cancers with BRAF V600 Mutations. *New England Journal of Medicine* 2015;373(8):726-36.
77. Heinrich MC, Joensuu H, Demetri GD, Corless CL, Apperley J, Fletcher JA, et al. Imatinib Target Exploration Consortium Study B2225. Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. *Clinical Cancer Research* 2008;14(9):2717-25.
78. Conley BA, Doroshow JH. Molecular analysis for therapy choice: NCI MATCH. . *Semin Oncol* 2014;41(3):297-9.

79. National Cancer Institute at the National Institutes of Health. Biological Therapies for Cancer, 2018.
80. Cancer Research UK. Targeted cancer drugs, 2017.
81. Scott DL. Biologics-Based Therapy for the Treatment of Rheumatoid Arthritis. *Clinical Pharmacology & Therapeutics* 2012;91(1):30–43.
82. Macedo T. Management of inflammatory joint disease. *New Zealand family physician* 2004;31(2):94-100.
83. Ledingham J, Snowden N, Ide Z. Diagnosis and early management of inflammatory arthritis. *BMJ* 2017;358:j3248.
84. Keystone EC, Smolen J, van Riel P. Developing an effective treatment algorithm for rheumatoid arthritis. *Rheumatology* 2012;51(Suppl 5):v48-v54.
85. Smolen JS, Landewé R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann Rheum Dis* 2014;73(3):492–509.
86. Smolen JS, Landewé R, Bijlsma J, Burmester G, Chatzidionysiou K, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann Rheum Dis* 2017;76(6):960-77.
87. Prince FH, Bykerk VP, Shadick NA, Lu B, Cui J, Frits M, et al. Sustained rheumatoid arthritis remission is uncommon in clinical practice. *Arthritis Research & Therapy* 2012;14(2):R68.
88. Reynolds A, Koenig AS, Bananis E, Singh A. When is switching warranted among biologic therapies in rheumatoid arthritis? *Expert Rev Pharmacoeconmics & Outcomes Research* 2012;12(3):319-33.
89. Govoni M, Bortoluzzi A, Lo Monaco A, Adami S, Addimanda O, Caimmi C, et al. Therapeutic options after treatment failure in rheumatoid arthritis or spondyloarthritides. *Advances in Therapy* 2014;31(8):780-802.
90. 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.
91. 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.
92. 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.
93. Finckh A, Simard JF, Gabay C, Guerne PA, SCQM physicians. Evidence for differential acquired drug resistance to anti-tumour necrosis factor agents in rheumatoid arthritis. *Ann Rheum Dis* 2006;65(6):746-52.
94. Jani M, Barton A, Warren RB, Griffiths CE, Chinoy H. The role of DMARDs in reducing the immunogenicity of TNF inhibitors in chronic inflammatory diseases. *Rheumatology* 2014;53(2):213-22.
95. 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.
96. Furst DE, Gaylis N, Bray V, Olech E, Yocum D, Ritter J, et al. Open-label, pilot protocol of patients with rheumatoid arthritis who switch to infliximab after an incomplete response to etanercept: the opposite study. *Ann Rheum Dis* 2007;66(7):893-9.
97. Emery P. Optimizing outcomes in patients with rheumatoid arthritis and an inadequate response to anti-TNF treatment. *Rheumatology* 2012;51(Suppl 5):v22-v30.

98. Maneiro JR, Salgado E, Gomez-Reino JJ. Immunogenicity of monoclonal antibodies against tumor necrosis factor used in chronic immune-mediated Inflammatory conditions: systematic review and meta-analysis. *JAMA Internal Medicine* 2013;173(15):1416-28.
99. 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 rhematoid arthritis *Pharmacoeconomics* 2014;32(10):1015 - 28.
100. Raza K, Saber TP, Kvien TK, Tak PP, Gerlag DM. Timing the therapeutic window of opportunity in early rheumatoid arthritis: proposal for definitions of disease duration in clinical trials. *Ann Rheum Dis* 2012;71(12):1921-3.
101. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *The New England Journal of Medicine* 2000;343(22):1586-93.
102. St Clair EW, Heijde van der DM, Smolen JS, Maini RN, Bathon JM, Emery P, et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: A randomized, controlled trial. 2004;50(11):3432-43.
103. Maini R, St Clair E, Breedveld F, Furst D, Kalden J, Weisman M, et al. Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. ATTRACT Study Group. *Lancet* 1999;354(9194):1932-9.
104. Weinblatt ME, Kremer JM, Bankhurst AD, Bulpitt KJ, Fleischmann RM, Fox RI, et al. A trial of etanercept, a recombinant tumor necrosis factor receptor:Fc fusion protein, in patients with rheumatoid arthritis receiving methotrexate. *The New England Journal of Medicine* 1999;340(4):253-9.
105. Smolen JS, Kay J, Doyle MK, Landewé R, Matteson EL, Wollenhaupt J, et al. Golimumab in patients with active rheumatoid arthritis after treatment with tumour necrosis factor alpha inhibitors (GO-AFTER study): a multicentre, randomised, double-blind, placebo-controlled, phase III trial. *Lancet* 2009;374(9685):210-21.
106. Genovese MC, Becker JC, Schiff M, Luggen M, Sherrer Y, Kremer J, et al. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor alpha inhibition. *New England Journal of Medicine* 2005;353(11):1114-23.
107. 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.
108. van den Broek M, Lems W, Allaart CF. BeSt practice: the success of early-targeted treatment in rheumatoid arthritis. *Clinical and Experimental Rheumatology* 2012;30(4 Suppl 73):S35-8.
109. 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.
110. 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.
111. Emery P, Gottenberg JE, Rubbert-Roth A, Sarzi-Puttini P, Choquette D, Taboada VM, et al. Rituximab versus an alternative TNF inhibitor in patients with rheumatoid arthritis who failed to respond to a single previous TNF inhibitor: SWITCH-RA, a global, observational, comparative effectiveness study. *Ann Rheum Dis* 2015;74(6):979-84.
112. Lloyd S, Bujkiewicz S, Wailoo AJ, Sutton AJ, Scott DA. The effectiveness of anti-TNF-a therapies when used sequentially in rheumatoid arthritis patients: a systematic review and meta-analysis. *Rheumatology* 2010;49:2313-21.
113. Silman A, Symmons D, Scott DG, Griffiths I. British Society for Rheumatology Biologics Register. *Ann Rheum Dis* 2003;62 (Suppl 2):ii28-9.

114. 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.
115. 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.
116. Geborek P, Crnkic M, Petersson I, Saxne T. Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southern Sweden. *Ann Rheum Dis* 2002;61(9):793-98.
117. Gómez-Reino JJ, Carmona L, Angel Descalzo M, Biobadaser Group. Risk of tuberculosis in patients treated with tumor necrosis factor antagonists due to incomplete prevention of reactivation of latent infection. *Arthritis & Rheumatology* 2007;57(5):756-61.
118. Hjardem E, Østergaard M, Pødenphant J, Tarp U, Andersen LS, Bing J, et al. Do rheumatoid arthritis patients in clinical practice benefit from switching from infliximab to a second tumor necrosis factor alpha inhibitor? *Ann Rheum Dis* 2007;66(9):1184-89.
119. Kvien TK, Heiberg MS, Lie E, Kaufmann C, Mikkelsen K, Nordvåg BY, et al. A Norwegian DMARD register: prescriptions of DMARDs and biological agents to patients with inflammatory rheumatic diseases. *Clinical and Experimental Rheumatology* 2005;23(5 Suppl 39):S188-94.
120. Silman AJ. Use of biologic agents in rheumatoid arthritis: introduction. *Rheumatology* 2011;50 (Suppl 4):iv3-4.
121. Messner DA. Evaluating the comparative effectiveness of treatment sequences in oncology: a new approach. *Journal of Comparative Effectiveness Research* 2015;4(6):537–54.
122. Porta C, Tortora G, Linassier C, Papazisis K, Awada A, Berthold D, et al. Maximising the duration of disease control in metastatic renal cell carcinoma with targeted agents: an expert agreement. *Medical Oncology* 2012;29(3):1896-907.
123. Freidlin B, Little RF, Korn EL. Design Issues in Randomized Clinical Trials of Maintenance Therapies. *Journal of the National Cancer Institute* 2015;107(11):djv225.
124. Gill S, Berry S, Biagi J, Butts C, Buyse M, Chen E, et al. Progression-free survival as a primary endpoint in clinical trials of metastatic colorectal cancer. *Current Oncology* 2011;18(Suppl 2):S5-S10.
125. Schmidinger M. Improving outcomes in metastatic clear cell renal cell carcinoma by sequencing therapy. *American Society of Clinical Oncology educational book* 2014:e228-38.
126. Dhawan A, Nichol D, Kinose F, Abazeed ME, Marusyk A, Haura EB, et al. Collateral sensitivity networks reveal evolutionary instability and novel treatment strategies in ALK mutated non-small cell lung cancer *Scientific Reports* 2017;7(1):1232.
127. Hammond WA, Swaika A, Mody K. Pharmacologic resistance in colorectal cancer: a review *Therapeutic Advances in Medical Oncology* 2016;8(1):57-84.
128. Groenendijk F, Bernards R. Drug resistance to targeted therapies: Deja vu all over again. *Molecular oncology* 2014;8(6):1067-83.
129. Dranitsaris G, Zhu X, Adunlin G, Vincent MD. Cost effectiveness vs. affordability in the age of immuno-oncology cancer drugs. *Expert Review of Pharmacoeconomics & Outcomes Research* 2018;18(4):351-57.
130. Evans WK, de Oliveira C. Evaluating Cost-Effectiveness: An Essential Part of New Cancer Drug Approvals. *Journal of Thoracic Oncology* 2017;12(10):1461-63.
131. Jönsson B, Ramsey S, Wilking N. Cost effectiveness in practice and its effect on clinical outcomes. *Journal of Cancer Policy* 2014;2(1):12-21.
132. 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.

133. Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist* 2008;13(Suppl 2):19-21.
134. Davis S, Tappenden P, Cantrell A. *A review of studies examining the relationship between progression-free survival and overall survival in advanced or metastatic cancer*. Sheffield: NICE Decision Support Unit, ScHARR, 2012.
135. 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.
136. European Medicines Agency (EMA). Guideline on the evaluation of anticancer medicinal products in man, 22 September 2017 2017.
137. Matulonis UA, Oza AM, Ho TW, Ledermann JA. Intermediate clinical endpoints: A bridge between progression-free survival and overall survival in ovarian cancer trials. *Cancer* 2014;121(11):1737-46.
138. Dimopoulos MA, Petrucci MT, Foà R, Catalano J, Kropff M, Terpos E, et al. Impact of maintenance therapy on subsequent treatment in patients with newly diagnosed multiple myeloma: use of “progression-free survival 2” as a clinical trial end-point. *Haematologica* 2015;100(8):e328-e30.
139. 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.
140. Allegra C, Blanke C, Buyse M, Goldberg R, Grothey A, Meropol NJ, et al. End points in advanced colon cancer clinical trials: a review and proposal. *Journal of Clinical Oncology* 2007;25(24):3572-5.
141. 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.
142. 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.
143. Rini BI, Escudier B, Tomczak P, Kaprin, A., Szczylik C, Hutson TE, et al. Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* 2011;378(9807):1931-9.
144. Fischer S, Gillessen S, Rothermundt C. Sequence of treatment in locally advanced and metastatic renal cell carcinoma. *Translational Andrology and Urology* 2015;4(3):310-25.
145. Calvo E, Ravaud A, Bellmunt J. What is the optimal therapy for patients with metastatic renal cell carcinoma who progress on an initial VEGFr-TKI? *Cancer Treat Rev* 2013;39(4):366-74.
146. 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.
147. Pal SK, Signorovitch JE, Li N, Zichlin ML, Liu Z, Ghate SR, et al. Patterns of care among patients receiving sequential targeted therapies for advanced renal cell carcinoma: A retrospective chart review in the USA. *International Journal of Urology* 2017;24(4):272-78.
148. Bozionelou V, Kakolyris S. Metastatic renal cell cancer: What is the sequential treatment for long-term clinical benefit? *Forum of Clinical Oncology* 2013;4(1):48-56.
149. Wells JC, Stukalin I, Norton C, Srinivas S, Lee JL, Donskov F, et al. Third-line Targeted Therapy in Metastatic Renal Cell Carcinoma: Results from the International Metastatic Renal Cell Carcinoma Database Consortium. *European Urology* 2017;71(2):204-09.
150. Sánchez-Gastaldo A, Kempf E, González Del Alba A, Duran I. Systemic treatment of renal cell cancer: A comprehensive review. *Cancer Treat Rev* 2017;60(77-89).

151. Rodriguez-Vida A, Hutson TE, Bellmunt J, Strijbos MH. New treatment options for metastatic renal cell carcinoma. *ESMO Open* 2017;2:e000185.
152. NCT01217931 Sequential Two-agent Assessment in Renal Cell Carcinoma Therapy: The START Trial, 2018.
153. 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.
154. Michel MS, Vervenne W, Goebell PJ, Fischer Von Weikersthal L, Freier W, De Santis M, et al. Phase III randomized sequential open-label study to evaluate efficacy and safety of sorafenib (SO) followed by sunitinib (SU) versus sunitinib followed by sorafenib in patients with advanced/metastatic renal cell carcinoma without prior systemic therapy (SWITCH Study): Safety interim analysis results. *Journal of Clinical Oncology* 2012;30:15_suppl, 4539-39.
155. Motzer RJ, Barrios CH, Kim TM, Falcon S, Cosgriff T, Harker WG, et al. Phase II randomized trial comparing sequential first-line everolimus and second-line sunitinib versus first-line sunitinib and second-line everolimus in patients with metastatic renal cell carcinoma. *Journal of Clinical Oncology* 2014;32(25):2765-72.
156. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, et al. Phase 3 trial of everolimus for metastatic renal cell carcinoma : final results and analysis of prognostic factors. *Cancer* 2010;116(18):4256-65.
157. Thompson Coon J, Hoyle M, Green C, Liu Z, Welch K, Moxham T, et al. Bevacizumab, sorafenib tosylate, sunitinib and temsirolimus for renal cell carcinoma: a systematic review and economic evaluation. *Health Technology Assessment* 2010;14(2):1-184, iii-iv.
158. Tappenden P, Jones R, Paisley S, Carroll C. Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer. *Health Technology Assessment* 2007;11(12):1-146.
159. Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, et al. Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation. *Health technology assessment* 2006;10(9):1-161.
160. Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, et al. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. *Lancet* 2007;369(9566):1000-15.
161. Chadwick D. Safety and efficacy of vigabatrin and carbamazepine in newly diagnosed epilepsy: a multicentre randomised double-blind study. Vigabatrin European Monotherapy Study Group. *Lancet* 1999;354(9172):13-19.
162. 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).
163. 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.
164. Cockerell OC, Johnson AL, Sander JW, Hart YM, Shorvon SD. Remission of epilepsy: results from the National General Practice Study of Epilepsy. *Lancet* 1995;346(8968):140-44.
165. Powell GA, Bonnett LJ, Tudur-Smith C, Hughes DA, Williamson PR, Marson AG. Using routinely recorded data in the UK to assess outcomes in a randomised controlled trial: The Trials of Access. *Trials* 2017;18(1):389.
166. Chadwick D, Shukralla A, Marson T. Comparing drug treatments in epilepsy. *herapeutic Advances in Neurological Disorders* 2009;2(3):181-7.
167. Kwan P, Brodie MJ. Early identification of refractory epilepsy. *New England Journal of Medicine* 2000;342(5):314-19.

168. 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.
169. 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.
170. Kobelt G, Lekander I, Lang A, Raffeiner 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.
171. 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.
172. 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.
173. 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).
174. 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.
175. 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.
176. 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.
177. 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.
178. 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.
179. 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.
180. 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.
181. 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.
182. Cimmino MA, Leardini G, Salaffi F, Intorcia 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.

183. 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.
184. 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.
185. 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.
186. 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.
187. 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.
188. 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.
189. 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.
190. 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.
191. Malottki 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.
192. 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.
193. 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.
194. 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.
195. 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.
196. Jobanputra P. A clinician's critique of rheumatoid arthritis health economic models. *Rheumatology* 2011;50 (suppl 4):iv48-iv52.
197. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Economics* 2006;15(12):1295-310.
198. Kim SY, Goldie SJ. Cost-effectiveness analyses of vaccination programmes: a focused review of modelling approaches. *Pharmacoeconomics* 2008;26(3):191-215.
199. 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.
200. 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.

201. 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.
202. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? . *Value in Health* 2010;13(8):1056-60.
203. Soares MO, Canto e Castro L. *Simulation or cohort models? Continuous time simulation and discretized Markov models to estimate cost-effectiveness. CHE Research Paper 56*. York: Centre for Health Economics, University of York, 2010.
204. Petrou S, Gray A. Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting. *BMJ* 2011;342:d1766.
205. Briggs A, Sculpher M. An Introduction to Markov Modelling for Economic Evaluation. *Pharmacoeconomics* 1998;13(4):397-409.
206. Gray A, Clarke P, Wolstenholme J, Wordsworth S. *Applied methods of cost-effectiveness analysis in health care*. Oxford: University Press, 2010.
207. Briggs A, Sculpher M, K. C. Chapter 3: Further developments in decision analytic models for economic evolution. *Decision Modelling for Health Economic Evaluation*. Oxford: University Press, 2006.
208. 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.
209. Petitti DB. *Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine* New York, NY: Oxford University Press 1994.
210. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993;13(4):322-38.
211. 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.
212. 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.
213. 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.
214. 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.
215. Sonnenberg A. Diminishing returns in sequential interventions of gastroenterology. *European Journal of Gastroenterology and Hepatology* 2008;20(5):465-68.
216. Tosh J, Brennan A, Wailoo A, Bansback N. The Sheffield rheumatoid arthritis health economic model. *Rheumatology* 2011;50(4):26-31.
217. Tavakoli M, Davies HTO, Thomson R. Decision analysis in evidence-based decision making. *Journal of Evaluation of Clinical Practice* 2000;6(2):111-20.
218. 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.
219. 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.
220. Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;26(9):781-98.
221. National Institute for Health and Clinical Excellence. NICE Briefing paper for methods review workshop on exploring uncertainty. London: NICE, 2008.
222. 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.

223. Ades AE, Claxton K, Sculpher M. Evidence synthesis, parameter correlation and probabilistic sensitivity analysis. *Health Econonomics* 2006;15(4):373-81.
224. Briggs AH. Handling Uncertainty in Cost-Effectiveness Models. *Pharmacoeconomics* 2000;17(5):479–500.
225. National Institute for Health and Care Excellence (NICE). *Guide to the methods of technology appraisal 2013*. London: NICE, 2013.
226. Claxton K, Sculpher M, McCabe C, Briggs AH, Akehurst R, Buxton M, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health Economics* 2005;14(4):339-47.
227. 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.
228. 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.
229. 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.
230. Ethgen O, Standaert B. Population- versus cohort-based modelling approaches. *Pharmacoeconomics* 2012;30(3):171-81.
231. Chhatwal J, He T. Economic evaluations with agent-based modelling: an introduction. *Pharmacoeconomics* 2015;33(5):423-33.
232. 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.

APPENDIX VOLUME II

The thesis includes a series of integrated reviews, the findings of which are presented in separate chapters. **Appendix volume II** includes the supplementary data for each of these chapters, listed below. A list of contents is included at the start of each appendix (A-E).

LIST OF APPENDICES INCLUDED IN VOLUME II:

A. APPENDIX FOR CHAPTER 2: NETWORK META-ANALYSES OF SCIATICA TREATMENTS

LIST OF CONTENTS:

Table A7.3: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for global effect which included only RCTs or quasi-RCTs
Table A7.4: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for pain intensity which included all study types
Table A7.5: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for pain intensity which included only RCTs or quasi-RCTs

APPENDIX A1: INTRODUCTION

Chapter 2 summarises the findings of the network meta-analysis published in the Spine Journal (Lewis, 2015), which represent updated analyses that were undertaken in response to peer review comments. The original health technology assessment (HTA) (Lewis, 2011) also included the initial evaluation of the evidence on clinical effectiveness for each individual treatment category based on a series of conventional pairwise meta-analysis. I performed all the analyses for the health technology appraisal, including the pair-wise meta-analyses and funnel plots.

The publication of the network meta-analysis in the Spine Journal included substantial supplementary data which are presented in this appendix (Sections A3-7). This Appendix also includes a brief summary of some of the pair-wise meta-analysis conducted as part of the evaluation for disc surgery in the original HTA (Section A2). This provides an illustration of the limited available evidence that precluded met-regression analysis for evaluating the effect of previous treatments.

References

Lewis R, Williams NH, Sutton AJ, Burton K, Ud Din N, Matar HE, Hendry M, Phillips CJ, Nafees S, Fitzsimmons D, Rickard I, Clare 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. doi: 10.1016/j.spinee.2013.08.049. Epub 2013 Oct 4.

Lewis R, Williams N, Matar HE, Din N, Fitzsimmons D, Phillips C, Jones M, Sutton A, Burton K, Nafees S, Hendry M, Rickard I, Chakraverty R, Wilkinson C. The effectiveness and cost effectiveness of management strategies for sciatica: systematic review and economic model. Health Technology Assessment 2011: 15(39)

APPENDIX A2: EXAMPLES OF PAIRWISE META-ANALYSES AND A FUNNEL PLOT PERFORMED AS PART OF THE HEALTH TECHNOLOGY ASSESSMENT OF SCIATICA TREATMENTS

Methods

The data were analysed according to three follow-up periods: short (≤6 weeks), medium (6 weeks to 6 months) and long (>6 months). Studies were pooled using the random effects model in STATA (StataCorp LP, College Station TX, USA) with between study heterogeneity examined using I^2 and Chi^2 statistics. For all comparisons for which there were more than eight studies, funnel plots together with associated statistical tests were used to assess the potential publication bias. We had originally planned to evaluate the effect of study level covariates (such as symptom duration, previous treatments, and study quality criteria) on between study heterogeneity using meta-regression, where ten or more studies were included in the pairwise meta-analysis. However, only one comparison (disc surgery vs chemonucleolysis for global effect at long term follow-up) included sufficient studies. The poor reporting of data relevant to these covariates, such as previous treatment, also hampered the feasibility of undertaking the meta-regression analyses.

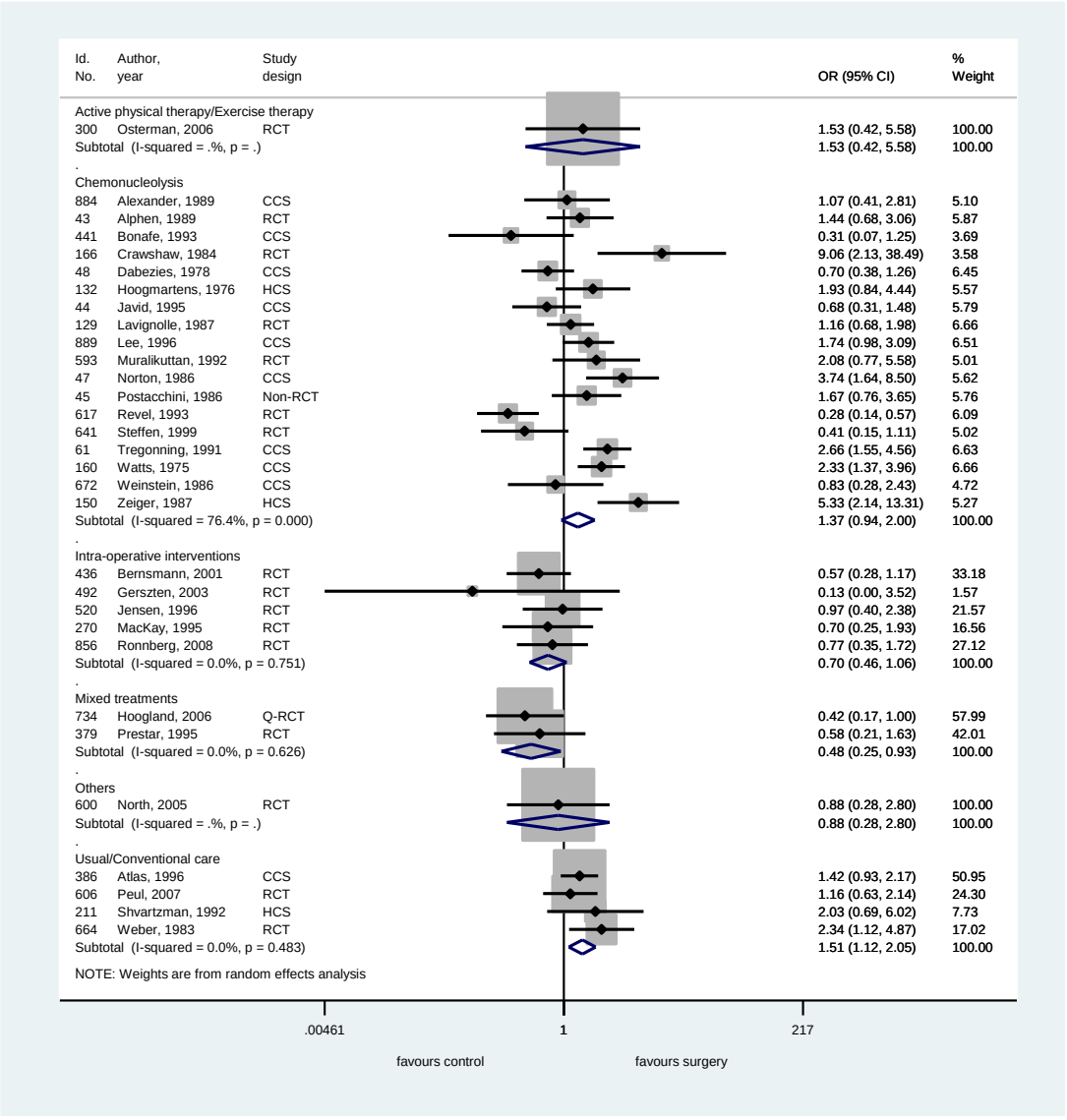
Disc surgery

The forest plots for the comparison of disc surgery versus alternative treatment strategies for the outcome global effect at long tern follow-up are presented in Figure A2.1, and the funnel plot for studies comparing disk surgery and chemonucleolysis is presented in Figure A2.2. Disc surgery was compared to usual care, active physical therapy, chemonucleolysis, intra-operative interventions, mixed treatments, and spinal cord stimulation (categorised within 'Others'). Duration of follow-up ranged from one to 10 years. Most studies included patients with chronic (>3 months) sciatica or a mixture of chronic and acute (≤3 months) symptoms. Just over half (16/31, 52%) of the studies were RCTs of which only one (606) was good quality overall (comparing disc surgery to usual care).

Eighteen studies (884, 43, 441, 166, 48, 132, 44, 129, 889, 593, 47, 45, 617, 641, 61, 160, 672, 150) compared disc surgery and chemonucleolysis, for which the findings were very heterogeneous, giving a pooled result that was borderline statistically significant in favour of surgery. A summary of the sciatica type and study population for the studies are presented in Table A2.1. The majority of studies (n=17) included patients who had received previous treatment for their current episode of sciatica; the type of previous treatments used and the associated number of patients were not well reported in most studies. Three studies (132, 617, 160) included patients who had received previous disc surgery, and 8 included patients who had not. The duration of follow-up ranged from one to ten years and duration of sciatica varied between studies. Eight studies included patients with chronic sciatica (884, 43, 132, 44, 641, 61, 160, 672), five considered either chronic or acute sciatica (441, 593, 47, 45, 150), and 5 did

not report this information (166, 48, 129, 889, 617). Two studies (45, 672) included patients who had sciatica for the first time and one study (47) only included patients with recurrent sciatica. The remaining studies included patients with either first episode or recurrent sciatica (48, 132, 160), or did not report this information. There was a mixture of study designs. When only the six RCTs (43, 166, 129, 593, 617, 641) were considered the findings were still heterogeneous, although most reported findings in favour of disc surgery (pooled analysis: OR 1.12, 95% CI: 0.51, 2.49). One moderate quality RCT (617) found chemonucleolysis to be more effective than disc surgery but the study had a high withdrawal rate in the surgery group (at least 41%), compared with chemonucleolysis (at least 19%), with drop-outs being given a poor outcome in the analysis. The funnel plot, for publication and other biases, did not appear to show asymmetry, but did indicate a lack of large studies.

Figure A2.1: Summary findings of global effect at long term follow-up for studies comparing disc surgery with alternative interventions



CCS, concurrent cohort studies (cohort study with concurrent controls); CI, confidence interval; HCS, historical cohort studies (cohort study with historical controls); id No, unique identifier number for each study; Non-RCT, non-randomised controlled trials; OR, odds ratio; RCT, randomised controlled trial.

Figure A2.2: Funnel plot with pseudo 95% confidence intervals for studies comparing disc surgery with chemonucleolysis at long term follow-up

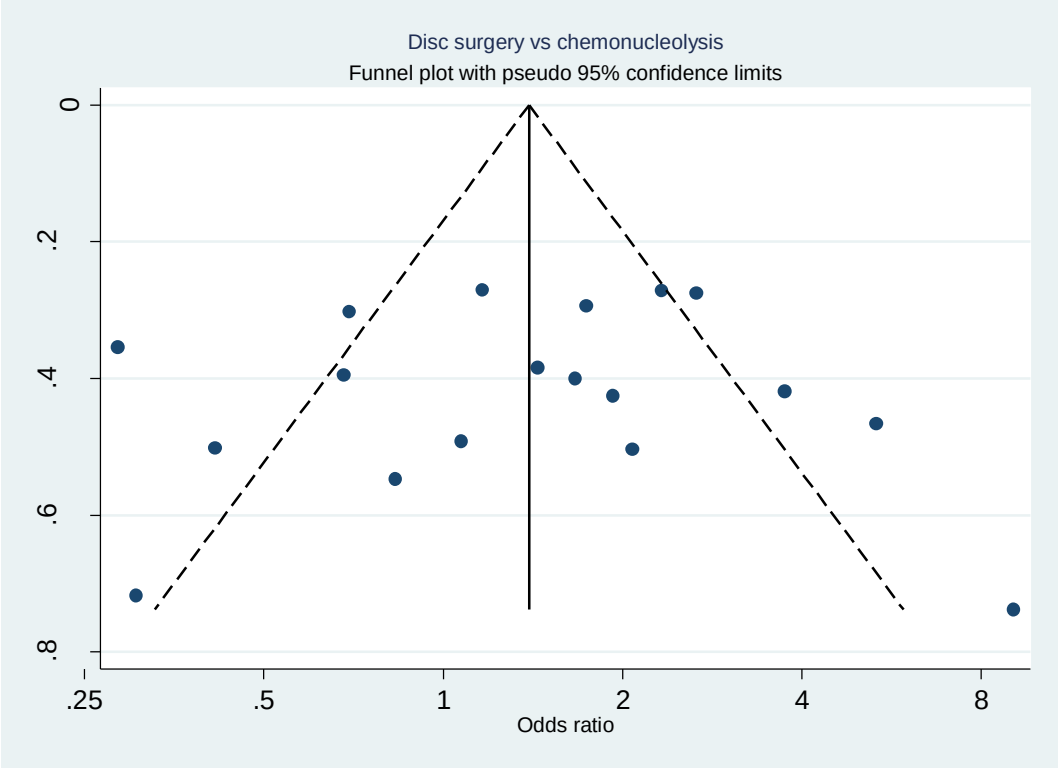


Table A2.1: Summary of sciatica type and study population details for studies that were included in the meta-analysis comparing disc surgery with chemonucleolysis for the outcome global effect at long term follow-up *(ordered alphabetically by author)*

ID No.	Author, year	Study design	No. of patients	Age	No. of men (%)	Symptom duration	Type of sciatica	Confirmed by imaging	Recurrent episode	Included patients with stenosis[a]	Included patients with sequestered disc (or extruded)	Previous treatment for sciatica	Previous back surgery for sciatica
884	Alexander, 1989	CCS	100	Mean 33.5 (range 18-65) yrs	90 (90%)	Mean 5.5 mths	Nerve root pain	Yes	NR	No	Yes	Yes	No
43	Alphen, 1989	RCT	151	Mean 34 (range 18-45) yrs	99 (66%)	<6 mths 55%, >6 mths 45%	Nerve root pain	Yes	NR	No	No	Yes	No
441	Bonafe, 1993 [French language]	CCS	40	Mean 46 (range 27-68) yrs.	28 (70%)	Mean 3 mths (range several dys to 15 mths)	Nerve root pain	Yes	NR	No	No	Yes	NR
166	Crawshaw, 1984	RCT	52	Mean 37 yrs	NR	NR	Nerve root pain	Yes	NR	No	No	Yes	No
48	Dabezies, 1978	CCS	200	Mean 39 yrs	132 (66%)	NR	Nerve root & referred pain	No	Recurrent and first episode	No	No	Yes	NR
132	Hoogmartens, 1976	HCS	97	Mean 35.5 yrs	48 (49%)	25-35 mths	Nerve root pain	NR	Recurrent and first episode	No	No	Yes	Yes
44	Javid, 1995	CCS	200	Mean 39 (range 17-81) yrs	134 (67%)	Mean 7.2 mths	Nerve root pain	Yes	NR	No	No	Yes	No
129	Lavignolle, 1987 [French language]	RCT	358	Mean 41 (SD 12.03 yrs	225 (63%)	NR	Nerve root pain	NR	NR	No	No	NR	NR
889	Lee, 1996 [German language]	CCS	300	<30 yrs 50%, >40 yrs 25%	213 (71%)	NR	Nerve root pain	Yes	NR	No	No	Yes	NR
593	Muralikuttan, 1992	RCT	92	Mean 35 (range 19-60) yrs	55 (60%)	Mean 24 wks	Nerve root pain	Yes	NR	No	No	Yes	NR
47	Norton, 1986	CCS	105	Mean 40(range 20-67) yrs	86 (82%)	Mean 18.5 mths (range 5 dys-128 mths)	Nerve root pain	Yes	Recurrent	No	No	Yes	No
45	Postacchini, 1986	Non-RCT	161	NR	NR	Mean 8.75 range 1.2-36 mths	Nerve root & referred pain	Yes	First episode	No	No	Yes	NR
617	Revel, 1993	RCT	165	Mean 39 (SD 9; range 21-65) yrs	96 (68%)	NR	Nerve root pain	Yes	NR	No	No	Yes	Yes
641	Steffen, 1999 [German language]	RCT	69	NR	NR	10.6 mths	Nerve root pain	Yes	NR	No	No	Yes	NR
61	Tregonning, 1991	CCS	268	Mean 40.4 (range 20-65) yrs	135 (68%)	NR	Nerve root pain	Yes	NR	No	No	Yes	No
160	Watts, 1975	CCS	274	Range 24-62 yrs	55 (55%)	NR	Nerve root & referred pain	Yes	Recurrent and first episode	No	No	Yes	Yes

ID No.	Author, year	Study design	No. of patients	Age	No. of men (%)	Symptom duration	Type of sciatica	Confirmed by imaging	Recurrent episode	Included patients with stenosis[a]	Included patients with sequestered disc (or extruded)	Previous treatment for sciatica	Previous back surgery for sciatica
672	Weinstein, 1986	CCS	159	Mean 41 (range 28-57) yrs	64 (41%)	Minimum period of 3 mths	Nerve root pain	Yes	First episode	No	No	Yes	No
150	Zeiger, 1987	CCS	126	NR	NR	4 or more wks	Nerve root pain	Yes	NR	No	No	Yes	No

CCS, concurrent cohort studies; HCS, historical cohort studies; mths, months; NA, Not applicable; NR, Not reported; RCT, randomised controlled trial; SD, standard deviation; wks, weeks; yrs, years

APPENDIX A3: WINBUGS CODES

These are based on those published on the Bristol University network meta-analysis webpage (https://www.bris.ac.uk/cobm/research/mpes/mtc.html)

GLOBAL

```
#Random-effects model for multi-arm trials
model{
for(i in 1:NS){
        w[i,1] <-0
                delta[i,t[i,1]]<-0
                mu[i] ~ dnorm(0,.0001)            # vague priors for 95 study baselines
                for (k in 1:na[i])  {
                      r[i,k] ~ dbin(p[i,t[i,k]],n[i,k])   # binomial likelihood
                                  logit(p[i,t[i,k]])<-mu[i] + delta[i,t[i,k]]
             rhat[i,k] <- p[i,t[i,k]] * n[i,k]
           dev[i,k] <- 2*(r[i,k] * (log(r[i,k]/rhat[i,k])) + (n[i,k]-r[i,k]) * (log((n[i,k]-r[i,k])/(n[i,k]-
rhat[i,k])))
                                                }
                     sumdev[i] <- sum(dev[i,1:na[i]])

# model
                        for (k in 2:na[i]) {
                delta[i,t[i,k]] ~ dnorm(md[i,t[i,k]],taud[i,t[i,k]])  # trial-specific LOR distributions
                md[i,t[i,k]] <-  d[t[i,k]] - d[t[i,1]]  + sw[i,k]       # mean of LOR distributions
                 taud[i,t[i,k]] <- tau *2*(k-1)/k                        #precision of LOR distributions
                 w[i,k] <- (delta[i,t[i,k]]  - d[t[i,k]] + d[t[i,1]])     #adjustment, multi-arm RCTs
                   sw[i,k] <-sum(w[i,1:k-1])/(k-1) }            # cumulative adjustment for multi-arm
                                                                                trials
  }
ssumdev <-sum(sumdev[])
d[1]<-0
for (k in 2:NT){d[k] ~ dnorm(0,.0001) }     #  vague priors for basic parameters

sd~dunif(0,2)                                  #  vague prior for random-effects standard deviation
tau<-1/pow(sd,2)
tau.squared <- sd*sd


# Absolute log odds(success) on Treatment A,  based on a separate model on the
#  29 studies Treatment A arms.
mA ~ dnorm(-0.476, 40.076)
# Absolute pr(success) Treatments B,C,D based on T[1] and the
#   MEAN  Relative treatment effects
for (k in 1:NT)   { logit(T[k])<- mA  +d[k] }


# Ranking and prob{treatment k is best}
for (k in 1:NT) { rk[k]<- NT+1 - rank(T[],k)
                     best[k]<-equals(rk[k],1)}


# pairwise ORs
for (c in 1:(NT-1))
        {  for (k in (c+1):NT)
              {  lor[c,k] <- d[k] - d[c]
                 log(or[c,k]) <- lor[c,k]
              }
         }

}
```

PAIN

```
#Random-effects model combining study- and arm-based summaries
model{
        for (i in 1:N.trial){
                prec[i]<- 1/var[i]          #Precision of differences = 1/var
                diff[i]~dnorm(delta[i],prec[i]) #'Likelihood for mean differences between arms

                delta[i]~dnorm(md[i],tau)          #Random-effects model for delta's
                md[i]<- d[t.trial[i]] - d[b.trial[i]]     #Define functional parameters for t[i] vs b[i]


        #       dev2[i] <- (diff[i]-delta[N.arm+i])*(diff[i]-delta[N.arm+i])/var[i]}
    #    sumdev2 <- sum(dev2[1:N.trial])


dev2[i] <- (diff[i]-delta[i])*(diff[i]-delta[i])/var[i]}
        sumdev2 <- sum(dev2[1:N.trial])

 for(i in 1:N.arm){
                prec.y[i]<- n[i]/(sd[i]*sd[i])
                y[i] ~ dnorm(my[i],prec.y[i])
                my[i]<-mu[s[i]]+ delta[i+N.trial]*(1-equals(t.arm[i],b.arm[i]))


#Random-effects model for treatment effects
                delta[i+N.trial] ~ dnorm(md[i+N.trial],tau)
                md[i+N.trial] <- d[t.arm[i]] - d[b.arm[i]]

        #       dev[i] <- (y[i]-my[i])* (y[i]-my[i])*prec.y[i]        }
   #  sumdev <- sum(dev[1:N.arm])

        dev[i] <- (y[i]-my[i])* (y[i]-my[i])*prec.y[i]        }
     sumdev <- sum(dev[1:N.arm])

tot.sumdev <- sumdev  + sumdev2


        for(j in 2:53){ mu[j]~dnorm(0,.0001)}

        d[1]<-0
        for (k in 2:NT) {d[k] ~ dnorm(0,.00001) }      # vague priors for basic parameters
        sd.d~dunif(0,50)
                                                # vague prior for random-effects sd
        tau<-1/pow(sd.d,2)

        tau.squared <- sd.d*sd.d

# Ranking and prob{treatment k is best}
        for (k in 1:NT) {
                rk[k]<- rank(d[],k)
                best[k]<-equals(rk[k],1)
                }

# pairwise mean difference comparisons
for (c in 1:(NT-1)) {  for (k in (c+1):NT)  { SMD[c,k] <- (d[k] - d[c] ) }  }

}
```

APPENDIX A4: INCLUDED STUDIES (n=122)

REFERENCE DETAILS FOR INCLUDED STUDIES

Alexander AH, Burkus JK, Mitchell JB, Ayers WV. Chymopapain chemonucleolysis versus surgical discectomy in a military population. Clin Orthop Relat Res. 1989(244):158-165.

Aminmansour B, Khalili HA, Ahmadi J, Nourian M. Effect of high-dose intravenous dexamethasone on postlumbar discectomy pain. Spine. Oct 1 2006;31(21):2415-2417.

Atlas SJ, Deyo RA, Keller RB, et al. The Maine Lumbar Spine Study .2. 1-year outcomes of surgical and nonsurgical management of sciatica. Spine. Aug 1996;21(15):1777-1786.

Becker C, Heidersdorf S, Drewlo S, de Rodriguez SZ, Kramer J, Willburger RE. Efficacy of epidural perineural injections with autologous conditioned serum for lumbar radicular compression: an investigator-initiated, prospective, double-blind, reference-controlled study. Spine. Aug 1 2007;32(17):1803-1808.

Bernsmann K, Kramer J, Ziozios I, Wehmeier J, Wiese M. Lumbar micro disc surgery with and without autologous fat graft - A prospective randomized trial evaluated with reference to clinical and social factors. Arch. Orthop. Trauma Surg. Sep 2001;121(8):476-480.

Bonafe A, Tremoulet M, Sabatier J, et al. Foraminal and Lateroforaminal Herniated Disks - Midterm Results of Percutaneous Treatment with Chemonucleolysis or Nucleotomy. Neurochirurgie. 1993;39(2):110-115.

Bontoux D, Alcalay M, Debiais F, et al. Treatment of Lumbar-Disk Herniation with Intradiscal Injection of Chymopapain or Triamcinolone Hexacetonid - a Comparative-Study of 80 Cases. Rev Rhum Engl Ed. Apr 1990;57(4):327-331.

Bourgeois P, Benoist M, Palazzo E, et al. Multicenter, randomized and double-blind study of triamcinolone hexacetonide versus chymopapain in the treatment of lumbar sciatica. Preliminary results at six months. Rev Rhum Mal Osteo Articulaires. 1988(10):767-769.

Brown MD, Tompkins JS. Pain response post-chemonucleolysis or disc excision. Spine. Mar 1989;14(3):321-326.

Buchner M, Zeifang F, Brocai DR, Schiltenwolf M. Epidural corticosteroid injection in the conservative management of sciatica. Clin Orthop. Jun 2000(375):149-156.

Buric J. Ozone chemyonucleolysis vs microdiscectomy. Prospective controlled study with 18 months follow-up. Rivista Italiana di Ossigeno-Ozonoterapia. Apr 2005;4(1):49-54.

Burton AK, Tillotson KM, Cleary J. Single-blind randomised controlled trial of chemonucleolysis and manipulation in the treatment of symptomatic lumbar disc herniation. European Spine Journal. Jun 2000;9(3):202-207.

Bush K, Hillier S. A controlled study of caudal epidural injections of triamcinolone plus procaine for the management of intractable sciatica. Spine. May 1991;16(5):572-575.

Carette S, Leclaire R, Marcoux S, et al. Epidural corticosteroid injections for sciatica due to herniated nucleus pulposus. New England Journal of Medicine. Jun 1997;336(23):1634-1640.

Cengiz SL, Baysefer A. Efficacy of Adcon-L gel or Healon-GV in epidural fibrosis after lumbar microdiscectomy. Neurosciences. Apr 2007;12(2):109-113.

Chatterjee S, Foy PM, Findlay GF. Report of a controlled clinical trial comparing automated percutaneous lumbar discectomy and microdiscectomy in the treatment of contained lumbar disc herniation. Spine. Mar 15 1995;20(6):734-738.

Chen M-r, Wang P, Cheng G, Guo X, Wei G-w, Cheng X-h. The warming acupuncture for treatment of sciatica in 30 cases. J Tradit Chin Med. 2009;29(1):50-53.

Cohen S, Bogduk N, Dragovich A, et al. Randomized, double-blind, placebo-controlled, dose-response, and preclinical safety study of transforaminal epidural etanercept for the treatment of sciatica. Anesthesiology. 2009;110(5):1116-1126.

Coomes EN. Comparison between Epidural Anaesthesia and Bed Rest in Sciatica. British Medical Journal. 1961;1(5218):20-24.

Crawshaw C, Frazer AM, Merriam WF, Mulholland RC, Webb JK. A comparison of surgery and chemonucleolysis in the treatment of sciatica. A prospective randomized trial. Spine. Mar 1984;9(2):195-198.

Dabezies EJ, Brunet M. Chemonucleolysis vs laminectomy. Orthopedics. Jan-Feb 1978;1(1):26-29.

Dabezies EJ, Langford K, Morris J, Shields CB, Wilkinson HA. Safety and efficacy of chymopapain (Discase) in the treatment of sciatica due to a herniated nucleus pulposus. Results of a randomized, double-blind study. Spine. May 1988;13(5):561-565.

Debi R, Halperin N, Mirovsky Y. Local application of steroids following lumbar discectomy. J Spinal Disord Tech. Aug 2002;15(4):273-276.

Dei-Anang K, Weigand H, Mader U. Is the Percutaneous Discectomy an Alternative to Chemonucleolysis. Radiologe. 1990;30(2):70-74.

Dilke TF, Burry HC, Grahame R. Extradural corticosteroid injection in management of lumbar nerve root compression. British Medical Journal. Jun 16 1973;2(5867):635-637.

Dincer U, Kiralp MZ, Cakar E, Yasar E, Dursan H. Caudal epidural injection versus non-steroidal anti-inflammatory drugs in the treatment of low back pain accompanied with radicular pain. Joint Bone Spine. Oct 2007;74(5):467-471.

Dreiser RL, Le Parc JM, Velicitat P, Lleu PL. Oral meloxicam is effective in acute sciatica: two randomised, double-blind trials versus placebo or diclofenac. Inflammation Research. Mar 2001;50 Suppl 1:S17-23.

Dubourg G, Rozenberg S, Fautrel B, et al. A pilot study on the recovery from paresis after lumbar disc herniation. Spine. Jul 1 2002;27(13):1426-1431; discussion 1431.

Duplan B, Cabanel G, Piton JL, Grauer JL, Phelip X. Acupuncture and sciatica in the acute phase. Double-blind study of 30 cases. Semaine des Hopitaux. Dec 8 1983;59(45):3109-3114.

Feldman J, Menkes CJ, Pallardy G, et al. Double-blind study of the treatment of disc lumbosciatica by chemonucleolysis. Rev Rhum Mal Osteoartic. Mar 1986;53(3):147-152.

Finckh A, Zufferey P, Schurch M-A, Balague F, Waldburger M, So AKL. Short-term efficacy of intravenous pulse glucocorticoids in acute discogenic sciatica. A randomized controlled trial. Spine. Feb 15 2006;31(4):377-381.

Gallucci M, Limbucci N, Zugaro L, et al. Sciatica: treatment with intradiscal and intraforaminal injections of steroid and oxygen-ozone versus steroid only. Radiology. Mar 2007;242(3):907-913.

Genevay S, Stingelin S, Gabay C. Efficacy of etanercept in the treatment of acute, severe sciatica: a pilot study. Ann Rheum Dis. Sep 2004;63(9):1120-1123.

Gerszten PC, Moossy JJ, Flickinger JC, Welch WC. Low-dose radiotherapy for the inhibition of peridural fibrosis after reexploratory nerve root decompression for postlaminectomy syndrome. Journal of Neurosurgery. Oct 2003;99(3 Suppl):271-277.

Geurts JWM, van Wijk R, Wynne HJ, et al. Radiofrequency lesioning of dorsal root ganglia for chronic lumbosacral radicular pain: a randomised, double-blind, controlled trial. Lancet. Jan 2003;361(9351):21-26.

Ghoname EA, White PF, Ahmed HE, Hamza MA, Craig WF, Noe CE. Percutaneous electrical nerve stimulation: an alternative to TENS in the management of sciatica. Pain. Nov 1999;83(2):193-199.

Glasser RS, Knego RS, Delashaw JB, Fessler RG. The perioperative use of corticosteroids and bupivacaine in the management of lumbar disc disease. Journal of Neurosurgery. Mar 1993;78(3):383-387.

Gogan WJ, Fraser RD. Chymopapain - a 10-Year, Double-Blind-Study. Spine. Apr 1992;17(4):388-394.

Goldie I. A clinical trial with indomethacin (indomee(R)) in low back pain and sciatica. Acta Orthop Scand Suppl. 1968;39(1):117-128.

Graham CE. Chemonucleolysis: a double blind study comparing chemonucleolysis with intra discal hydrocortisone: in the treatment of backache and sciatica. Clin Orthop. Jun 1976(117):179-192.

Grevsten S, Johansson H. Phenylbutazone in treatment of acute lumbago-sciatica. Zeitschrift fur Rheumatologie. Nov-Dec 1975;34(11-12):444-447.

Hansson E, Hansson T. The cost-utility of lumbar disc herniation surgery. European Spine Journal. Mar 2007;16(3):329-337.

Hedeboe J, Buhl M, Ramsing P. Effects of using dexamethasone and placebo in the treatment of prolapsed lumbar disc. Acta Neurologica Scandinavica. Jan 1982;65(1):6-10.

Helliwell M, Robertson JC, Ellis RM. Outpatient treatment of low pack pain and sciatica by a single extradural corticosteroid injection. Br J Clin Pract. 1985;39(6):228-231.

Herrmann WA, Geertsen MS. Efficacy and safety of lornoxicam compared with placebo and diclofenac in acute sciatica/lumbo-sciatica: an analysis from a randomised, double-blind, multicentre, parallel-group study. Int J Clin Pract. 2009;63(11):1613-1621.

Hofstee DJ, Gijtenbeek JMM, Hoogland PH, et al. Westeinde sciatica trial: randomized controlled study of bed rest and physiotherapy for acute sciatica. Journal of Neurosurgery. Jan 2002;96(1 Suppl):45-49.

Holve RL, Barkan H. Oral steroids in initial treatment of acute sciatica. J Am Board Fam Med. 2008;21(5):469-474.

Hoogmartens M, Van Loon L, Demedts D, Desmet L. Comparative study of the results obtained with discectomy and chemonucleolysis. Acta Orthop Belg. Mar-Apr 1976;42(2):134-138.

Javid MJ, Nordby EJ, Ford LT, et al. Safety and efficacy of chymopapain (Chymodiactin) in herniated nucleus pulposus with sciatica. Results of a randomized, double-blind study. Journal of the American Medical Association. May 13 1983;249(18):2489-2494.

Javid MJ. Chemonucleolysis Versus Laminectomy - a Cohort Comparison of Effectiveness and Charges. Spine. Sep 1995;20(18):2016-2022.

Jensen TT, Asmussen K, Berg-Hansen EM, et al. First-time operation for lumbar disc herniation with or without free fat transplantation. Prospective triple-blind randomized study with reference to clinical factors and enhanced computed tomographic scan 1 year after operation. Spine. May 1 1996;21(9):1072-1076.

Jirarattanaphochai K, Jung S, Thienthong S, Krisanaprakornkit W, Sumananont C. Peridural methylprednisolone and wound infiltration with bupivacaine for postoperative pain control after posterior lumbar spine surgery - A randomized double-blinded placebo-controlled trial. Spine. 2007;32(6):609-616.

Karppinen J, Malmivaara A, Kurunlahti M, et al. Periradicular infiltration for sciatica - A randomized controlled trial. Spine. May 2001;26(9):1059-1067.

Karppinen J, Korhonen T, Malmivaara A, et al. Tumor necrosis factor-alpha monoclonal antibody, infliximab, used to manage severe sciatica. Spine. Apr 15 2003;28(8):750-753; discussion 753-754.

Khoromi S, Patsalides A, Parada S, Salehi V, Meegan JM, Max MB. Topiramate in chronic lumbar radicular pain. The Journal of Pain. 2005;6(12):829-836.

Khoromi S, Cui L, Nackers L, Max MB. Morphine, nortriptyline and their combination vs. placebo in patients with chronic lumbar root pain. Pain. Jul 2007;130(1-2):66-75.

Kim KD, Wang JC, Robertson DP, et al. Reduction of radiculopathy and pain with Oxiplex/SP gel after laminectomy, laminotomy, and discectomy: a pilot clinical study. Spine. May 15 2003;28(10):1080-1087; discussion 1087-1088.

Kim M-J, Lee S-H, Jung E-S, et al. Targeted percutaneous transforaminal endoscopic diskectomy in 295 patients: comparison with results of microscopic diskectomy. Surg Neurol. Dec 2007;68(6):623-631.

Klenerman L, Greenwood R, Davenport HT, White DC, Peskett S. Lumbar epidural injections in the treatment of sciatica. Br. J. Rheumatol. Feb 1984;23(1):35-38.

Koranda I, Sefrna F. Effectiveness of conservative and surgical treatment of lumboischiadic syndrome. Cas Lek Cesk. Aug 2 1995;134(15):474-477.

Korhonen T, Karppinen J, Paimela L, et al. The treatment of disc herniation-induced sciatica with infliximab - Results of a randomized, controlled, 3-month follow-up study. Spine. Dec 2005;30(24):2724-2728.

Kwasucki J, Olbrych K, ska B, Stepie A, Dec S, Maksymiuk G. Assessment of dexamethasone effectiveness in the treatment of ischalgia. Neurol Neurochir Pol. 1993;27(4):515-522.

Kwasucki J, Stepien A, Maksymiuk G, Olbrych-Karpinska B. Evaluation of analgesic action of fluvoxamine compared with efficacy of imipramine and tramadol for treatment of sciatica--open trial. Wiad Lek. 2002;55(1-2):42-50.

Lagarrigue J, Lazorthes Y, Verdie JC, Richaud J. Comparative Results of Surgery and Nucleolysis in 1085 Cases of Herniated Lumbar Disc. Neurochirurgie. 1991;37(2):96-105.

Laiq N, Khan MN, Iqbal MJ, Khan S. Comparison of Epidural Steroid Injections with conservative management in patients with lumbar radiculopathy. J Coll Physicians Surg Pak. 2009;19(9):539-543.

Larsson U, Choler U, Lidstrom A, et al. Auto Traction for Treatment of Lumbago Sciatica a Multicenter Controlled Investigation. Acta Orthop Scand Suppl. 1980;51(5):791-798.

Lavignolle B, Vital JM, Baulny D, Grenier F, Castagnera L. Comparative study of surgery and chemonucleolysis in the treatment of sciatica caused by a herniated disk. Acta Orthop Belg. 1987;53(2):244-249.

Lee SH, Lee SJ, Park KH, et al. Comparison of percutaneous manual and endoscopic laser diskectomy with chemonucleolysis and automated nucleotomy. Orthopade. 1996;25(1):49-55.

Lee SH, Chung SE, Ahn Y, Kim TH, Park JY, Shin SW. Comparative radiologic evaluation of percutaneous endoscopic lumbar discectomy and open microdiscectomy: A matched cohort analysis. Mount Sinai Journal of Medicine. Sep 2006;73(5):795-801.

Lidstrom A, Zachrisson M. Physical therapy on low back pain and sciatica. An attempt at evaluation. Scand J Rehabil Med. 1970;2(1):37-42.

Ljunggren AE, Walker L, Weber H, Amundsen T. Manual traction versus isometric exercises in patients with herniated intervertebral lumbar discs. Physiotherapy Theory and Practice. Dec 1992;8(4):207-213.

Luijsterburg PAJ, Verhagen AP, Ostelo RWJG, et al. Physical therapy plus general practitioners' care versus general practitioners' care alone for sciatica: a randomised clinical trial with a 12-month follow-up. European Spine Journal. Apr 2008;17(4):509-517.

Lundin A, Magnuson A, Axelsson K, Kogler H, Samuelsson L. The effect of perioperative corticosteroids on the outcome of microscopic lumbar disc surgery. European Spine Journal. Dec 2003;12(6):625-630.

Mackay MA, Fischgrund JS, Herkowitz HN, Kurz LT, Hecht B, Schwartz M. The Effect of Interposition Membrane on the Outcome of Lumbar Laminectomy and Diskectomy. Spine. Aug 1995;20(16):1793-1796.

Mathews JA, Mills SB, Jenkins VM, et al. Back pain and sciatica: controlled trials of manipulation, traction, sclerosant and epidural injections. Br J Rheumatol. 1987;26(6):416-423.

Mayer HM, Brock M. Percutaneous Endoscopic Discectomy - Surgical Technique and Preliminary-Results Compared to Microsurgical Discectomy. Journal of Neurosurgery. Feb 1993;78(2):216-225.

Moret NC, van der Stap M, Hagmeijer R, Molenaar A, Koes BW. Design and feasibility of a randomized clinical trial to evaluate the effect of vertical traction in patients with a lumbar radicular syndrome. Manual Therapy. 1998;3(4):203-211.

Muralikuttan KP, Hamilton A, Kernohan WG, Mollan RA, Adair IV. A prospective randomized trial of chemonucleolysis and conventional disc surgery in single level lumbar disc herniation. Spine. Apr 1992;17(4):381-387.

Murata Y, Kato Y, Miyamoto K, Takahashi K. Clinical study of low back pain and radicular pain pathways by using l2 spinal nerve root infiltration: a randomized, controlled, clinical trial. Spine. 2009;34(19):2008-2013.

North RB, Kidd DH, Farrokhi F, Piantadosi SA. Spinal cord stimulation versus repeated lumbosacral spine surgery for chronic pain: a randomized, controlled trial. Neurosurgery. 2005;56(1):98-106; discussion 106-107.

Norton WL. Chemonucleolysis versus surgical discectomy. Comparison of costs and results in workers' compensation claimants. Spine. Jun 1986;11(5):440-443.

Osterman H, Seitsalo S, Karppinen J, Malmivaara A. Effectiveness of microdiscectomy for lumbar disc herniation: a randomized controlled trial with 2 years of follow-up. Spine. Oct 1 2006;31(21):2409-2414.

Ozturk B, Gunduz OH, Ozoran K, Bostanoglu S. Effect of continuous lumbar traction on the size of herniated disc material in lumbar disc herniation. Rheumatology International. May 2006;26(7):622-626.

Peul WC, Van Houwelingen HC, Van Den Hout WB, et al. Surgery versus prolonged conservative treatment for sciatica. New England Journal of Medicine. 31 2007;356(22):2245-2256.

Popiolek A, Domanik A, Mazurkiewicz G. Epidural injections of steroids in the treatment of patients with chronic sciatica in discopathy. Neurol Neurochir Pol. Sep-Oct 1991;25(5):640-646.

Porsman O, Friis H. Prolapsed lumbar disc treated with intramuscularly administered dexamethasonephosphate. A prospectively planned, double-blind, controlled clinical trial in 52 patients. Scandinavian Journal of Rheumatology. 1979;8(3):142-144.

Postacchini F, Lami R, Massobrio M. Chemonucleolysis versus surgery in lumbar disc herniations: correlation of the results to preoperative clinical pattern and size of the herniation. Spine. Mar 1987;12(2):87-96.

Price C, Arden N, Coglan L, Rogers P. Cost-effectiveness and safety of epidural steroids in the management of sciatica. Health Technology Assessment. Aug 2005;9(33):1-58.

Rasmussen S, Krum-Moller DS, Lauridsen LR, et al. Epidural steroid following discectomy for herniated lumbar disc reduces neurological impairment and enhances recovery: a randomized study with two-year follow-up. Spine. 2008;33(19):2028-2033.

Rattanatharn R, Sanjaroensuttikul N, Anadirekkul P, Chaivisate R, Wannasetta W. Effectiveness of lumbar traction with routine conservative treatment in acute herniated disc syndrome. Journal of the Medical Association of Thailand. Sep 2004;87 Suppl 2:S272-277.

Reust P, Chantraine A, Vischer TL. Treatment of lumbar sciatica with or without neurological deficit using mechanical traction. A double-blind study. Schweizerische Medizinische Wochenschrift. Feb 27 1988;Journal Suisse de Medecine. 118(8):271-274.

Revel M, Payan C, Vallee C, et al. Automated Percutaneous Lumbar Discectomy Versus Chemonucleolysis in the Treatment of Sciatica - a Randomized Multicenter Trial. Spine. Jan 1993;18(1):1-7.

Richter HP, Kast E, Tomczak R, Besenfelder W, Gaus W. Results of applying ADCON-L gel after lumbar discectomy: the German ADCON-L study. Journal of Neurosurgery. Oct 2001;95(2 Suppl):179-189.

Ridley MG, Kingsley GH, Gibson T, Grahame R. Outpatient lumbar epidural corticosteroid injection in the management of sciatica. Br. J. Rheumatol. Aug 1988;27(4):295-299.

Ronnberg K, Lind B, Zoega B, et al. Peridural scar and its relation to clinical outcome: a randomised study on surgically treated lumbar disc herniation patients. European Spine Journal. 2008;17(12):1714-1720.

Santilli V, Beghi E, Finucci S. Chiropractic manipulation in the treatment of acute back pain and sciatica with disc protrusion: a randomized double-blind clinical trial of active and simulated spinal manipulations. Spine J. Mar-Apr 2006;6(2):131-137.

Schwetschenau PR, Ramirez A, Johnston J. Double blind evaluation of intradiscal chymopapain for herniated lumbar discs; early results. J Neurosurg. 1976;45(6):622-627.

Shvartzman L, Weingarten E, Sherry H, Levin S, Persaud A. Cost-effectiveness analysis of extended conservative therapy versus surgical intervention in the management of herniated lumbar intervertebral disc. Spine. Feb 1992;17(2):176-182.

Snoek W, Weber H, Jorgensen B. Double-Blind Evaluation of Extradural Methyl Prednisolone for Herniated Lumbar Discs. Acta Orthop Scand Suppl. 1977;48(6):635-641.

Steffen R. Laser discotomy/Chemonucleolysis ini the therapy of displaced lumbal intervertebral disc - prospective randomised comparison. Hefte zur der Unfallchirurg. 1999:215-221.

Stula D. Chemonucleolysis with Chymopapain in Intervertebral-Disk Hernia - a Randomized Comparative-Study of Patients Following Surgery. Neurochirurgia (Stuttg). Sep 1990;33(5):169-172.

Styczynski T, Zarski S, Drozdowska Z, et al. Clinical Evaluation of the Effectiveness of Traction Treatment of Patients with Herniation of Lumbar Intervertebral Discs. Reumatologia (Warsaw). 1991;29(3-4):275-281.

Tassi GP. Comparison of results of 500 microdiscectomies and 500 percutaneous laser disc decompression procedures for lumbar disc herniation. Photomed Laser Surg. Dec 2006;24(6):694-697.

Tregonning GD, Transfeldt EE, McCulloch JA, Macnab I, Nachemson A. Chymopapain versus conventional surgery for lumbar disc herniation. 10-year results of treatment. J Bone Joint Surg Br. May 1991;73(3):481-486.

de Tribolet N, Porchet F, Lutz TW, et al. Clinical assessment of a novel antiadhesion barrier gel: prospective, randomized, multicenter, clinical trial of ADCON-L to inhibit postoperative peridural fibrosis and related symptoms after lumbar discectomy. Am J Orthop. 1998;27(2):111-120.

Unlu Z, Tasci S, Tarhan S, Pabuscu Y, Islak S. Comparison of 3 physical therapy modalities for acute pain in lumbar disc herniation measured by clinical evaluation and magnetic resonance imaging. J Manipulative Physiol Ther. Mar 2008;31(3):191-198.

Vad VB, Bhat AL, Lutz GE, Cammisa F. Transforaminal epidural steroid injections in lumbosacral radiculopathy: a prospective randomized study. Spine. Jan 1 2002;27(1):11-16.

Valat JP, Giraudeau B, Rozenberg S, et al. Epidural corticosteroid injections for sciatica: a randomised, double blind, controlled clinical trial. Ann Rheum Dis. Jul 2003;62(7):639-643.

van Alphen HA, Braakman R, Bezemer PD, Broere G, Berfelo MW. Chemonucleolysis versus discectomy: a randomized multicenter trial. Journal of Neurosurgery. Jun 1989;70(6):869-875.

Veihelmann A, Devens C, Trouillier H, Birkenmaier C, Gerdesmeyer L, Refior HJ. Epidural neuroplasty versus physiotherapy to relieve pain in patients with sciatica: a prospective randomized blinded clinical trial. J Orthop Sci. Jul 2006;11(4):365-369.

Vroomen PC, de Krom MC, Wilmink JT, Kester AD, Knottnerus JA. Lack of effectiveness of bed rest for sciatica. New England Journal of Medicine. Feb 11 1999;340(6):418-423.

Watters WC, Mirkovic S, Boss J. Treatment of the isolated lumbar intervertebral disc herniation: microdiscectomy versus chemonucleolysis. Spine (Phila Pa 1976). 1988;13(3):360-362.

Watts C, Hutchison G, Stern J, Clark K. Comparison of intervertebral disc disease treatment by chymopapain injection and open surgery. Journal of Neurosurgery. Apr 1975;42(4):397-400.

Weber H. Lumbar disc herniation. A controlled, prospective study with ten years of observation. Spine. Mar 1983;8(2):131-140.

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. Sep 1 1993;18(11):1433-1438.

Wehling P, Reinecke J. Acupuncture together with cytokine depressing herbs in comparison to injection therapy with steroids in sciatic pain. Schmerz. Jun 13 1997;11(3):180-184.

Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. Journal of the American Medical Association. Nov 22 2006;296(20):2441-2450.

Weinstein JN, Lurie JD, Tosteson TD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT) observational cohort. Journal of the American Medical Association. 22 2006;296(20):2451-2459.

Weinstein J, Spratt KF, Lehmann T, McNeill T, Hejna W. Lumbar disc herniation. A comparison of the results of chemonucleolysis and open discectomy after ten years. J Bone Joint Surg Am. Jan 1986;68(1):43-54.

Yildirim K, Sisecioglu M, Karatay S, et al. The effectiveness of gabapentin in patients with chronic radiculopathy. The Pain Clinic. 2003;15(3):213-218.

Zeiger HE, Jr. Comparison of chemonucleolysis and microsurgical discectomy for the treatment of herniated lumbar disc. Spine. Oct 1987;12(8):796-799.

Table A4.1: Summary of overall quality of included studies

Author, year	Study size	Overall follow-up	Study design	Adeq R [a]	All Con [b]	>80% FU [c]	Blind OA [d]	Overall quality rating	Overall external validity rating
Epidural injections/nerve block vs inactive control (D vs A)									
Bush, 1991	23	1 year	RCT	Unclear	Unclear	60-79%	Yes	Moderate	Weak
Snoek, 1977	51	ranged from 8 to 20 months	RCT	Unclear	Unclear	80-100%	Yes	Weak	Weak
Carette, 1997	158	3 months	RCT	Yes	Partial	60-79%	Yes	Strong	Moderate
Valat, 2003	85	35 days	RCT	Yes	Partial	80-100%	Yes	Strong	Weak
Dilke, 1973	100	3 months	RCT	Unclear	Unclear	60-79%	Yes	Moderate	Weak
Vad, 2002	50	Mean 16 months (Range 12-21 months)	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Helliwell, 1985	39	3 months	RCT	Unclear	Unclear	80-100%	Unclear	Moderate	Weak
Klenerman, 1984	74	2 months	RCT	Yes	Partial	80-100%	Yes	Weak	Weak
Ridley, 1988	39	6 months	RCT	Yes	Unclear	80-100%	Yes	Moderate	Weak
Karppinen, 2001	160	1 year	RCT	Yes	Partial	80-100%	Yes	Strong	Strong
Price, 2005	228	12 months	RCT	Yes	Partial	80-100%	Yes	Strong	Moderate
Mathews, 1987	57	12 months	RCT	Partial	Unclear	60-79%	Yes	Moderate	Moderate
Chemonucleolysis vs inactive control (E vs A)									
Gogan, 1991	60	10 years	RCT	Yes	Unclear	80-100%	Yes	Moderate	Moderate
Schweschenau, 1976	66	1 year	RCT	Yes	Yes	80-100%	Yes	Moderate	Moderate
Feldman, 1986	39	3 months	RCT	Unclear	Unclear	80-100%	Unclear	Moderate	Moderate
Dabezies, 1988	173	6 months	RCT	Partial	Yes	60-79%	Yes	Moderate	Weak
Javid, 1983	108	6 months	RCT	Yes	Partial	80-100%	Yes	Moderate	Weak
Non-opioids vs inactive control (F vs A)									
Goldie, 1968	50	14 days	RCT	Unclear	Yes	80-100%	Yes	Moderate	Weak
Hedeboe, 1982	39	3 months	RCT	Partial	Partial	80-100%	Yes	Moderate	Moderate
Porsman, 1979	52	9 days	RCT	Unclear	Yes	80-100%	Yes	Weak	Moderate
Weber, 1993	214	4 weeks	RCT	Unclear	Unclear	80-100%	No	Moderate	Weak
Dreiser, 2001	532	7 days	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Finckh, 2006	65	30 days	RCT	Yes	Partial	80-100%	Yes	Moderate	Weak
Grevsten, 1975	36	2 weeks	RCT	Unclear	Unclear	80-100%	Yes	Weak	Weak
Herrmann, 2009	171	5 days	RCT	Yes	Yes	80-100%	Yes	Moderate	Weak
Holve, 2008	29	6 months	Q-RCT	No	Partial	80-100%	Yes	Moderate	Weak
Traction vs inactive control (H vs A)									
Rattanatharn, 2004	120	4 weeks	RCT	Yes	Partial	60-79%	NA	Moderate	Weak
Larsson, 1980	84	3 months	RCT	Unclear	Unclear	80-100%	Unclear	Moderate	Weak
Reust, 1988	60	12 days	RCT	Yes	Unclear	<60%	Yes	Moderate	Weak
Manipulation vs inactive control (I vs A)									
Santilli, 2006	102	6 months	RCT	Yes	Yes	80-100%	Yes	Strong	Strong
Acupuncture vs inactive control (J vs A)									
Duplan, 1983	30	5 days	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Moderate
Passive physical therapy vs inactive control (L vs A)									
Ghoname, 1999	64	11 weeks	RCT	Unclear	Unclear	Can't tell	NA	Weak	Weak
Biological agents vs inactive control (M vs A)									
Karppinen, 2003	72	3 months	Non-RCT	No	No	Can't tell	No	Weak	Weak
Korhonen, 2005	41	1 year	RCT	Yes	Unclear	80-100%	Unclear	Moderate	Weak
Cohen, 2009	24	6 months	RCT	Yes	Partial	80-100%	Yes	Moderate	Weak
Opioids vs inactive control (O vs A)									
Khoromi, 2007	55	36 weeks	Cross-over RCT	Yes	Yes	<60%	Yes	Moderate	Strong
Neuropathic pain-modulators vs inactive control (R vs A)									
Khoromi, 2007	55	36 weeks	Cross-over RCT	Yes	Yes	<60%	Yes	Moderate	Strong
Yildirim, 2003	50	2 months	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Khoromi, 2005	42	8 weeks	Cross-over RCT	Yes	Yes	70%	Yes	Strong	Weak
Radiofrequency treatment vs inactive control (U vs A)									
Geurts, 2003	83	12 months but data for three months' follow up presented	RCT	Yes	Partial	80-100%	Yes	Strong	Moderate
Disc surgery vs conventional care (C vs B)									
Shvartzman, 1992	55	2 years	HCS	No	No	NA	NA	Weak	Weak
Koranda, 1991	100	3 months	Q-RCT	No	No	80-100%	Unclear	Weak	Moderate
Atlas, 1996	507	10 years	CCS	No	No	60-79%	NA	Moderate	Moderate

Author, year	Study size	Overall follow-up	Study design	Adeq R [a]	All Con [b]	>80% FU [c]	Blind OA [d]	Overall quality rating	Overall external validity rating
Peul, 2007	283	1 year (main follow-up visits at 8 weeks, 6 months and 1 year). 2 year data reported later	RCT	Yes	Partial	80-100%	NA	Strong	Strong
Weber, 1983	126	10 years	RCT	Unclear	Partial	60-79%	No	Weak	Moderate
Hansson, 2007	184	2 years	CCS	No	No	80-100%	NA	Weak	Moderate
Weinstein, 2006a	501	2 years	RCT	Yes	Yes	80-100%	NA	Strong	Weak
Weinstein, 2006b	743	2 years	CCS	No	No	80-100%	NA	Moderate	Weak
Epidural injections/nerve block vs conventional care (D vs B)									
Buchner, 2000	36	6 month	RCT	Partial	Partial	80-100%	Unclear	Moderate	Weak
Popiolek, 1991	60	21 days	Non-RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Laiq, 2009	52	6 months	Q-RCT	No	No	80-100%	No	Weak	Weak
Traction vs conventional care (H vs B)									
Styczynski, 1991	157	After treatment	Non-RCT	No	No	80-100%	Unclear	Weak	Weak
Exercise therapy vs conventional care (K vs B)									
Luijsterburg, 2008	135	12 months	RCT	Yes	Yes	80-100%	NA	Strong	Strong
Chemonucleolysis vs disc surgery (E vs C)									
Alphen, 1989	151	12 months	RCT	Partial	Unclear	80-100%	No	Moderate	Strong
Postacchini, 1986	161	Chemo: mean 2.9 years range 20-38 months Surgery: mean 2.8 years range 21-42 months	Non-RCT	No	No	80-100%	No	Weak	Moderate
Norton, 1986	105	At least 1 year	CCS	No	No	NA	Unclear	Weak	Weak
Dabezies, 1978	200	2 years	CCS	No	No	Can't tell	No	Weak	Moderate
Stula, 1990	69	Postoperative	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Tregonning, 1991	268	10 years	CCS	No	No	80-100%	No	Weak	Moderate
Lagarrigue, 1991	1085	Mean 17.2, range 12-84 months	CCS	No	No	80-100%	Unclear	Weak	Moderate
Lavignolle, 1987	358	Mean 25 months for surgery and 22 months for Chemonucleolysis	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Hoogmartens, 1976	97	58 months for discectomy and 38 months for Chemonucleolysis	HCS	No	No	NA	NA	Weak	Moderate
Watts, 1975	274	2 years	CCS	No	No	80-100%	Unclear	Weak	Weak
Crawshaw, 1984	52	1 year	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Moderate
Brown, 1989	85	3 months	CCS	No	No	80-100%	Yes	Weak	Weak
Buric, 2005	45	18 months	Non-RCT	No	No	80-100%	NA	Weak	Weak
Muralikuttan, 1992	92	1 year	RCT	Yes	Unclear	80-100%	Unclear	Moderate	Moderate
Weinstein, 1986	159	Range 10-13.5 years (mean 10.3 years)	CCS	No	No	80-100%	NA	Weak	Weak
Non-opioids vs disc surgery (F vs C)									
Dubourg, 2002	67	6 months	CCS	No	No	80-100%	No	Weak	Weak
Intra-operative interventions vs disc surgery (G vs C)									
Aminmansour, 2006	61	2 months	Q-RCT	No	No	80-100%	Yes	Weak	Moderate
MacKay, 1995	190	1 year	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Lundin, 2003	80	2 years	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Moderate
Cengiz, 2007	60	12 months	RCT	Unclear	Yes	80-100%	Unclear	Moderate	Weak
Kim, 2003	35	6 months	RCT	Yes	Yes	80-100%	NA	Moderate	Weak
Bernsmann, 2001	200	Median of 24.2 months after surgery	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Debi, 2002	70	1 year	RCT	Unclear	Partial	80-100%	No	Weak	Weak
Gerszten, 2003	10	1 year	RCT	Yes	Unclear	80-100%	NA	Moderate	Weak
Glasser, 1993	32	1 month	RCT	Unclear	Unclear	60-79%	Unclear	Weak	Weak

Author, year	Study size	Overall follow-up	Study design	Adeq R [a]	All Con [b]	>80% FU [c]	Blind OA [d]	Overall quality rating	Overall external validity rating
Jensen, 1996	118	Median 376, range 276-501 days	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Moderate
Richter, 2001	398	6 months	RCT	Yes	Yes	80-100%	Yes	Moderate	Weak
Rasmussen, 2008	200	2 years	RCT	Yes	Unclear	80-100%	Yes	Moderate	Weak
Ronnberg, 2008	128	2 years	RCT	Unclear	Partial	80-100%	Yes	Weak	Weak
Jirarattanaphochai, 2007	103	3 months	RCT	Yes	Partial	80-100%	Yes	Moderate	Moderate
Tribolet, 1998	298	6 months	RCT	Yes	Unclear	80-100%	Yes	Moderate	Moderate
Exercise therapy vs disc surgery (K vs C)									
Osterman, 2006	57	2 year	RCT	Yes	Yes	80-100%	NA	Moderate	Weak
Percutaneous discectomy vs disc surgery (Q vs C)									
Lee, 2006	60	Mean 37.5 (32-45) months	CCS	No	No	NA	Unclear	Weak	Weak
Tassi, 2006	1000	2 years	HCS	No	No	80-100%	No	Weak	Weak
Chatterjee, 1995	71	6 months	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Kim, 2007	915	Mean 23.6 (18-36) months	CCS	No	No	80-100%	No	Weak	Weak
Mayer, 1993	40	2 years	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Spinal cord stimulation vs disc surgery (T vs C)									
North, 2005	60	2 years	RCT	Yes	Partial	60-79%	No	Weak	Moderate
Non-opioids vs epidural injections/nerve block (F vs D)									
Dincer, 2007	64	3 months, assessment at 15th day, 1st and 3rd month	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Moderate
Murata, 2009	246 (136 radicular pain)	7 days	RCT	Unclear	Partial	80-100%	Unclear	Weak	Weak
Acupuncture vs epidural injections/nerve block (J vs D)									
Wehling, 1997	278	5 weeks	CCS	No	No	80-100%	No	Weak	Weak
Passive physical therapy vs epidural injections/nerve block (L vs D)									
Veihelmann, 2006	99	12 months	RCT	Partial	Yes	<60%	Yes	Moderate	Weak
Biological agents vs epidural injections/nerve block (M vs D)									
Becker, 2007	90	22 weeks	RCT	Yes	Partial	80-100%	Yes	Moderate	Weak
Bed rest vs epidural injections/nerve block (N vs D)									
Coomes, 1961	40	9 weeks	Non-RCT	No	No	80-100%	No	Weak	Weak
Intra-operative interventions vs chemonucleolysis (G vs E)									
Javid, 1995	200	1 year	CCS	No	No	80-100%	No	Weak	Moderate
Zeiger, 1987	126	ranged from 6 to 46 months, with an average time from treatment procedure to follow-up evaluation of 18 months	CCS	No	No	NA	Yes	Weak	Weak
Alexander, 1989	100	Mean 14 and range of 6-35 months	CCS	No	No	80-100%	Unclear	Weak	Weak
Watters,1988	100	3 years	Non-RCT	No	No	80-100%	No	Weak	Weak
Manipulation vs chemonucleolysis (I vs E)									
Burton, 2000	40	12 months	RCT	No	No	60-79%	Yes	Moderate	Weak
Percutaneous discectomy vs chemonucleolysis (Q vs E)									
Bonafe, 1993	40	Mean 15 months, range 3 to 36 months	CCS	No	No	80-100%	Unclear	Weak	Weak
Dei-Anang, 1990	201	1 year	CCS	No	No	NA	Unclear	Weak	Weak
Revel, 1993	165	1 year	RCT	Yes	Unclear	80-100%	Unclear	Moderate	Weak
Steffen, 1999	69	1 year	RCT	Unclear	Unclear	80-100%	Yes	Weak	Weak
Lee, 1996	300	1 year	CCS	No	No	Can't tell	Unclear	Weak	Weak
Epidural injections/nerve block vs chemonucleolysis (S vs E)									
Graham, 1976	40 (23 with sciatica)	2 years	Non-RCT	No	No	80-100%	Yes	Weak	Weak
Bourgeois, 1988	60	6 months	RCT	Yes	Partial	80-100%	Yes	Moderate	Weak
Bontoux, 1990	80	3 months	RCT	Yes	Unclear	80-100%	Yes	Moderate	Weak
Gallucci, 2007	159	6 months	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Acupuncture vs non-opioids (J vs F)									

Author, year	Study size	Overall follow-up	Study design	Adeq R [a]	All Con [b]	>80% FU [c]	Blind OA [d]	Overall quality rating	Overall external validity rating
Chen, 2009	90	1 year	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Moderate
Biological agents vs non-opioids (M vs F)									
Genevay, 2004	10	6 weeks	HCS	No	No	80-100%	No	Weak	Moderate
Opioids vs non-opioids (O vs F)									
Kwasucki, 1993	43	2 weeks	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Exercise therapy vs traction (K vs H)									
Ljunggren, 1992	50	1 week	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Passive physical therapy vs traction (L vs H)									
Unlu, 2008	60	3 month	RCT	Unclear	Unclear	80-100%	Yes	Moderate	Weak
Ozturk, 2005	46	2 weeks	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Mathews, 1987	143		RCT	Partial	Unclear	<60%	Yes	Moderate	Moderate
Bed rest vs traction (N vs H)									
Moret, 1998	16	3 weeks	RCT	Yes	Partial	80-100%	No	Moderate	Strong
Bed rest vs exercise therapy (N vs K)									
Lidstrom, 1970	62	1 month	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Education/advice vs bed rest (P vs N)									
Vroomen, 1999	183	12 weeks	RCT	Yes	No	80-100%	Yes	Strong	Strong
Hofstee, 2002	250	6 months	RCT	Yes	No	80-100%	No	Moderate	Moderate
Neuropathic pain-modulators vs opioids (R vs O)									
Kwasucki, 2002	70	19 days	RCT	Unclear	Unclear	80-100%	Unclear	Weak	Weak
Khoromi, 2007	55	36 weeks	Cross-over RCT	Yes	Yes	<60%	Yes	Moderate	Strong

RCT randomised controlled trial; Q-RCT, quasi-randomised controlled trial; CCS, concurrent cohort study; HCS, historical cohort study; NA, not applicable.

a Adequate randomisation

b Allocation concealment

c Follow up

d Blind outcome assessment

APPENDIX A5: RESULTS OF INACTIVE CONTROL COMPARISONS FROM NETWORK META-ANALYSES

Table A5.1: Probability of being best and the ORs of global effect for different treatment strategies compared with inactive control from the network meta-analysis and pairwise meta-analyses (*treatments ordered according to the probability of being the best*)

Treatment category	**Code**	**Probability of being 'best' (Mean)**	**Median OR (95% CrI)**	***Results of pairwise meta-analysis: Number of studies ORs (95% CIs)***
Biological agents	**M**	0.57	16.83 (0.76, 946.7)	N=1 10.0 (0.65, 166.67)
Acupuncture	**J**	0.26	7.92 (1.11, 66.65)	-
Manipulation	**I**	0.10	4.88 (1.07, 22.8)	N=1 4.72 (1.95, 11.37)
Spinal cord simulation	**T**	0.04	2.79 (0.45, 17.2)	-
Epidural injections / nerve block	**D**	0.01	3.48 (2.14, 5.78)	N=9 2.58 (1.25, 5.29)
Intra-operative interventions	**G**	0.01	3.64 (1.70 7.81)	-
Education / advice	**P**	0.01	1.75 (0.34, 9.01)	-
Conventional care	**B**	0.00	0.82 (0.42, 1.59)	-
Traction	**H**	0.00	1.28 (0.59, 2.73)	N=2 1.11 (0.60, 2.05)
Intradiscal injections	**S**	0.00	0.69 (0.25, 1.90)	-
Chemonucleolysis	**E**	0.00	1.57 (0.87, 2.82)	N=5 2.56 (1.59, 4.11)
Percutaneous discectomy	**Q**	0.00	1.23 (0.58, 2.59)	-
Disc surgery	**C**	0.00	2.46 (1.33, 4.51)	-
Non-opioids	**F**	0.00	2.18 (1.27, 3.78)	N=7 1.68 (1.12, 2.51)
Exercise therapy	**K**	0.00	1.09 (0.39, 3.01)	-
Passive physical therapy	**L**	0.00	1.49 (0.58, 3.9)	N=1 4.27 (1.47, 12.35)
Bed rest	**N**	0.00	1.35 (0.39, 4.73)	-
Opioids	**O**	0.00	1.15 (0.38, 3.41)	N=1 1.37 (0.5, 3.76)
Neuropathic painmodulators	**R**	0.00	1.41 (0.49, 4.04)	N=2 2.01 (0.77, 5.24)
Radiofrequency treatment	**U**	0.00	0.56 (0.1, 2.94)	N=1 0.57 (0.19, 1.72)

OR, odds ratio; CrI, credible interval; CI, confidence interval; N, number of studies

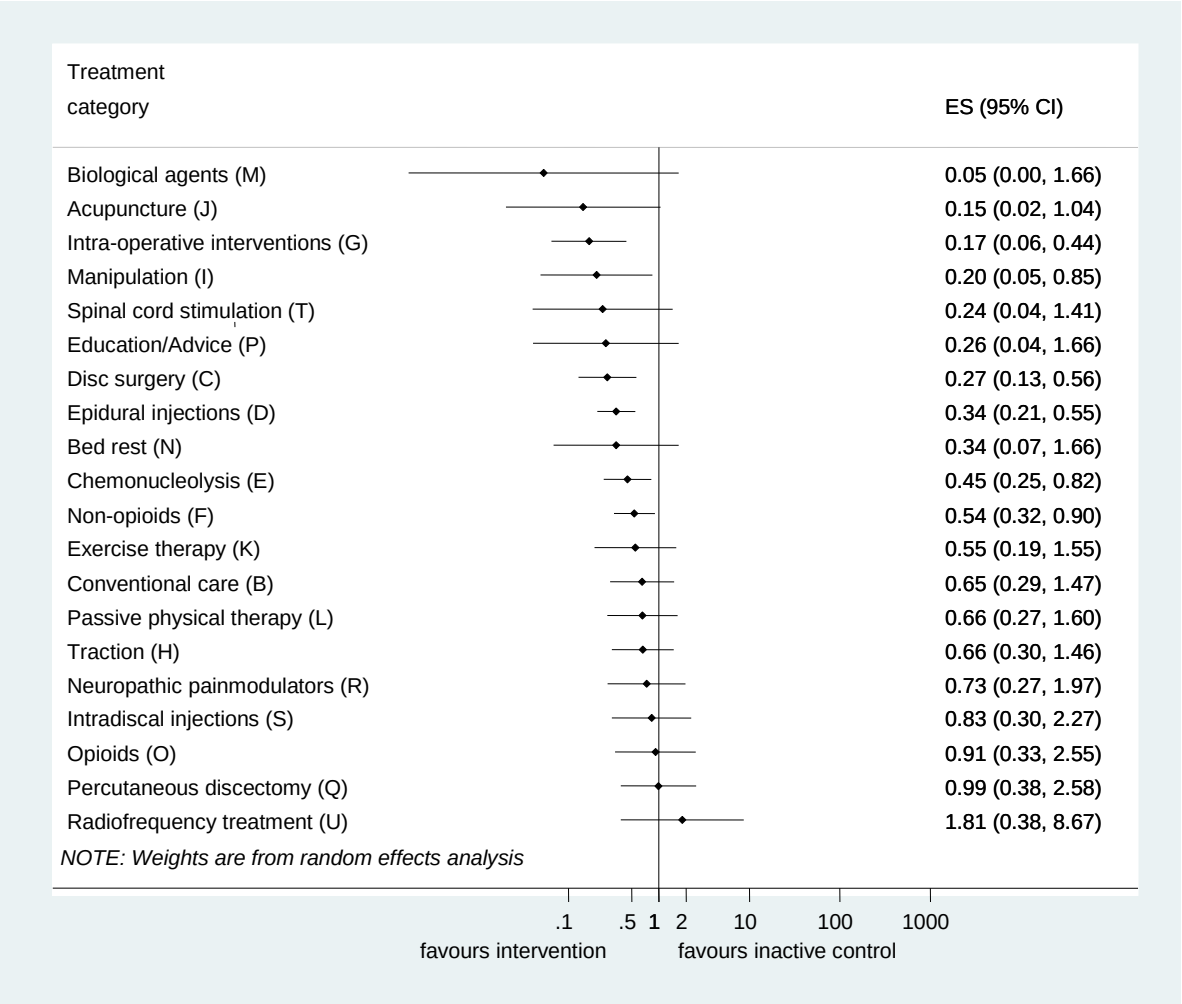
Table A5.2: Probability of being best and the ORs for global effect of different treatment strategies compared with inactive control from the network meta-analysis and pairwise meta-analyses based on RCTs and Q-RCTs only

(treatments ordered according to the probability of being the best)

Treatment category	**Code**	**Prob of being 'best' (Mean)**	**Median OR (95% CrI)**	***Results of pairwise meta-analysis: Number of studies ORs (95% CIs)***
Biological agents	**M**	0.55	16.63 (0.84, 839.8)	N=1 10.0 (0.65, 154.40)
Acupuncture	**J**	0.18	6.70 (1.02, 53.38)	-
Spinal cord simulation	**T**	0.07	4.2 (0.71, 25.47)	-
Manipulation	**I**	0.07	4.88 (1.18, 20.68)	N=1 4.71 (1.95, 11.37)
Education / advice	**P**	0.06	3.85 (0.60, 24.76)	-
Intra-operative interventions	**G**	0.06	5.90 (2.3, 15.59)	-
Bed rest	**N**	0.01	2.96 (0.60, 14.67)	-
Conventional care	**B**	0.00	1.54 (0.67, 3.48)	-
Non-opioids	**F**	0.00	1.87 (1.11, 3.15)	N=7 1.68 (1.12, 2.51)
Percutaneous discectomy	**Q**	0.00	1.01 (0.39, 2.68)	-
Disc surgery	**C**	0.00	3.70 (1.78, 7.85)	-
Epidural injections / nerve block	**D**	0.00	2.96 (1.85, 4.89)	N=9 2.58 (1.25, 5.29)
Chemonucleolysis	**E**	0.00	2.23 (1.21, 4.08)	N=5 2.56 (1.59, 4.12)
Traction	**H**	0.00	1.51 (0.69, 3.34)	N=2 1.11 (0.60, 2.05)
Exercise therapy	**K**	0.00	1.83 (0.64, 5.15)	-
Passive physical therapy	**L**	0.00	1.52 (0.63, 3.76)	N=1 4.27 (1.47, 12.42)
Opioids	**O**	0.00	1.1 (0.39, 3.03)	N=1 1.37 (0.5, 3.76)
Neuropathic painmodulators	**R**	0.00	1.37 (0.50, 3.69)	N=2 2.01 (0.77, 5.24)
Intradiscal injections	**S**	0.00	1.21 (0.44, 3.34)	-
Radiofrequency treatment	**U**	0.00	0.56 (0.11, 2.64)	N=1 0.57 (0.19, 1.71)

OR, odds ratio; CrI, credible interval; C,I confidence interval; N, number of studies

Figure A5.1: Plot of the ORs of global effects for the different treatment strategies compared with inactive control from the network meta-analysis, based on RCTs and Q-RCTs only



RCT, randomised controlled trial; Q-RCT, quasi-randomised controlled trial; OR, odds ratio; ES, effect size; CI, confidence interval

NB 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 OR <1 represents a decrease in the number of patients *not* showing overall improvement in favour of the intervention

Table A5.3: Probability of being best and the weighted mean difference for pain intensity for different treatment strategies compared with inactive control from the network meta-analysis and pairwise meta-analyses *(treatments ordered according to the probability of being the best)*

Treatment category	Code	Prob of being 'best' (Mean)	Median of the posterior (95% Credible interval)	*Results of meta-analysis: Number of studies WMD (95% CIs)*
Biological agents	**M**	0.33	-19.51 (-32.85, -6.43)	N=2 -9.91 (-43.23, 23.41)
Acupuncture	**J**	0.19	-14.82 (-33.81, 3.87)	N=1 -25.0 (-41.75, -8.25)
Intra-operative interventions	**G**	0.14	-14.64 (-32.93, 3.29)	
Manipulation	**I**	0.13	-7.24 (-39.15, 24.56)	
Neuropathic painmodulators	**R**	0.12	-11.6 (-31.63, 8.38)	N=1 -26.66 (-38.35, -14.97)
Traction	**H**	0.02	-1.81 (-21.48, 18.29)	N= 1 3.36 (-14.49, 21.21)
Exercise therapy	**K**	0.02	-2.62 (-25.0, 19.85)	
Epidural injections / nerve block	**D**	0.01	-11.4 (-19.19, -3.75)	N=7 -8.11 (-19.34, 3.12)
Chemonucleolysis	**E**	0.01	-6.62 (-24.56, 11.15)	N=1 -5.40 (-23.66, 12.86)
Passive physical therapy	**L**	0.01	-0.46 (-18.41, 18.18)	N=1 -7.00 (-13.58, -0.42)
Education / advice	**P**	0.01	16.22 (-19.01, 51.84)	
Percutaneous discectomy	**Q**	0.01	11.6 (-18.9, 41.77)	
Conventional care	**B**	0.0	-2.44 (-17.91, 12.89)	
Disc surgery	**C**	0.0	-9.54 (-25.41, 6.0)	
Bed rest	**N**	0.0	17.13 (-13.83, 48.52)	
Opioids	**O**	0.0	4.92 (-15.68, 25.3)	
Radiofrequency treatment	**U**	0.0	12.98 (-12.05, 37.83)	N=1 13.00 (2.04, 23.96)
Non-opioids	**F**	0.0	-2.54 (-11.79, 6.42)	N=4 -6.26 (-15.41, 2.89)

WMD, weighted mean difference; Cr,I credible interval; CI, confidence interval; N, number of studies

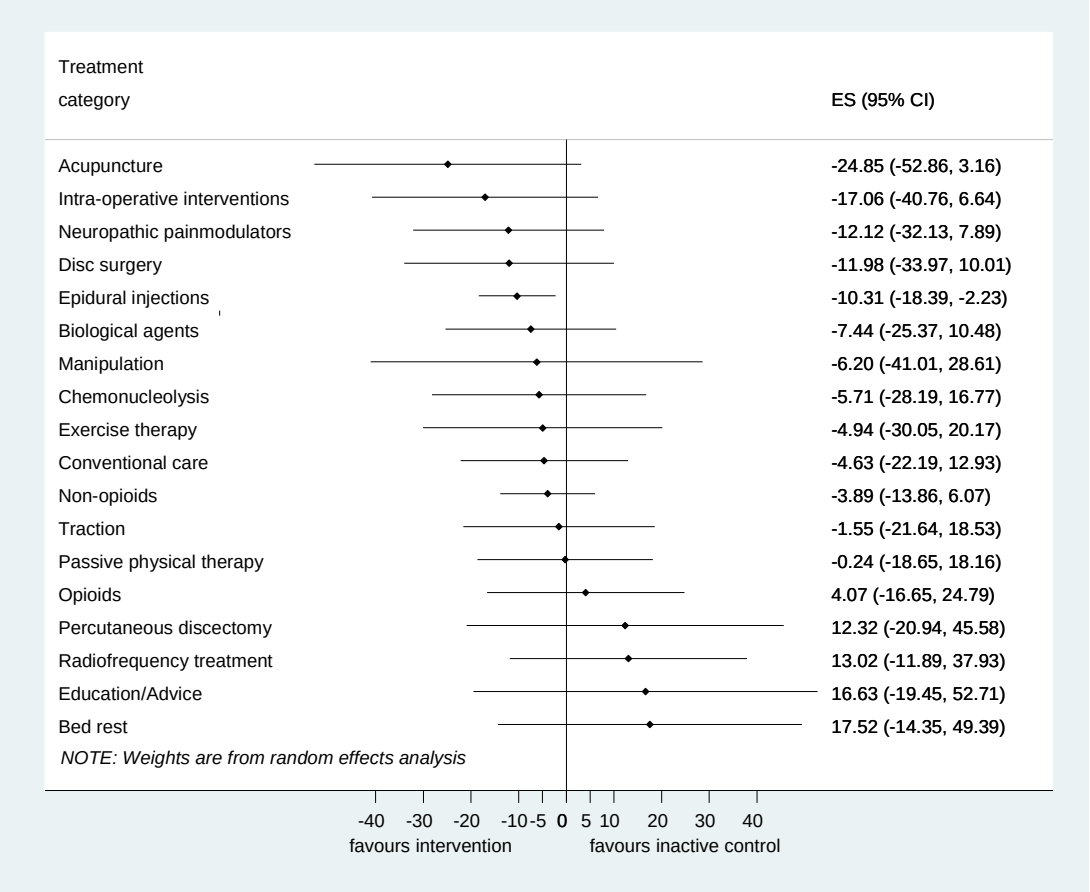
Table A5.4: Probability of being best and weighted mean difference for pain intensity of different treatment strategies compared with inactive control from the network meta-analysis and pairwise meta-analyses based on RCTs and Q-RCTs only

(treatments ordered according to the probability of being the best)

Treatment category	Code	Prob of being 'best' (Mean)	Median of the posterior (95% Credible interval)	*Results of meta-analysis: Number of studies WMD (95% CIs)*
Acupuncture	**J**	0.51	-24.8 (-53.02, 3.26)	N=1 -25.0 (-41.75, -8.25)
Intra-operative interventions	**G**	0.18	-17.04 (-40.87, 6.79)	
Manipulation	**I**	0.09	-6.26 (-41.09, 28.88)	
Neuropathic painmodulators	**R**	0.09	-12.11 (-32.45, 7.95)	N=1 -26.66 (-38.35, -14.97)
Biological agents	**M**	0.03	-7.49 (-25.31, 10.7)	N=1 7.0 (-5.25, 19.25)
Exercise therapy	**K**	0.02	-4.95 (-30.33, 20.39)	
Disc surgery	**C**	0.01	-11.96 (-34.09, 10.07)	
Epidural injections / nerve block	**D**	0.01	-10.3 (-18.5, -2.24)	N=7 -8.11 (-19.34, 3.12)
Chemonucleolysis	**E**	0.01	-5.68 (-28.31, 16.96)	N=1 -5.40 (-23.66, 12.86)
Traction	**H**	0.01	-1.64 (-21.46, 19.08)	N= 1 3.36 (-14.49, 21.21)
Education / advice	**P**	0.01	16.51 (-19.29, 53.29)	
Conventional care	**B**	0.0	-4.61 (-22.34, 13.05)	
Non-opioids	**F**	0.0	-3.82 (-14.11, 5.89)	N=4 -6.26 (-15.41, 2.89)
Passive physical therapy	**L**	0.0	-0.36 (-18.45, 18.57)	N=1 -7.0 (-13.58, -0.42)
Bed rest	**N**	0.0	17.35 (-14.16, 49.98)	
Opioids	**O**	0.0	4.19 (-16.96, 24.67)	
Percutaneous discectomy	**Q**	0.0	12.26 (-20.99, 45.88)	
Radiofrequency treatment	**U**	0.0	13.03 (-12.18, 38.18)	N=1 13.00 (2.04, 23.96)

RCT, randomised controlled trial; Q-RCT, quasi-randomised controlled trial; WMD, weighted mean difference; CrI, credible interval; CI, confidence interval; N, number of studies

Figure A5.2: Plot of the weighted mean difference for pain intensity for the different treatment strategies compared with inactive control from the network meta-analysis, based on RCTs and Q-RCTs only



RCT, randomised controlled trial; Q-RCT, quasi-randomised controlled trial; ES, effect size; CI, confidence interval

NB: A WMD > 0 represents a reduction in pain intensity in favour of the intervention

APPENDIX A6: RESULTS OF NETWORK META-ANALYSES RESTRICTED TO RCTs AND Q-RCTs

Table A6.1: Results (odds ratios ORs, with 95% confidence intervals/credible intervals) of network meta- analysis for RCTs/Q-RCTs reporting global effect

A			N=9, 2.58 (1, 5)	N=5, 2.56 (2, 4)	N=7, 1.68 (1, 3)		N=2, 1.11 (0.6, 2)	N=1, 4.71 (2, 11)			N=1, 4.27 (1, 12)	N=1, 10.0 (1, 154)		N=1, 1.37 (0.5, 4)			N=2, 2.01 (0.8, 5)			N=1, 0.57 (0.2, 2)
1.54 (1, 3)	**B**	N=4, 2.21 (1, 4)	N=2, 1.99 (0.7, 6)							N=1, 1.46 (0.7, 3)										
3.70 (2, 8)	2.41 (1, 4)	**C**		N=5, 0.57 (0.3, 1)		N=7, 1.49 (1, 2)				N=1, 0.77 (0.2, 3)							N=2, 0.31 (0, 3)			N=1, 1.13 (0.4, 4)	
2.96 (2, 5)	1.92 (0.8, 5)	0.80 (0.4, 2)	**D**		N=1, 0.45 (0.2, 1)						N=1, 0.20 (0.1, 1)										
2.23 (1, 4)	1.45 (0.7, 3)	0.6 (0.3, 1)	0.75 (0.4, 2)	**E**												N=2, 0.46 (0.3, 1)		N=3, 0.55 (0.3, 1)		
1.87 (1, 3)	1.22 (0.5, 3)	0.50 (0.2, 1)	0.63 (0.3, 1)	0.84 (0.4, 2)	**F**				N=1, 3.27 (0.8, 14)					N=1, 0.18 (0.1, 1)						
5.90 (2, 16)	3.84 (2, 9)	1.6 (0.9, 3)	2.0 (0.7, 5)	2.65 (1, 6)	3.20 (1, 9)	**G**														
1.51 (0.7, 3)	0.98 (0.4, 3)	0.41 (0.2, 1)	0.51 (0.2, 1)	0.68 (0.3, 2)	0.81 (0.3, 2)	0.25 (0.1, 0.8)	**H**			N=1, 0.88 (0.3, 3)	N=1, 0.93 (0.5, 2)		N=1, 1.0 (0.1, 7)							
4.88 (1, 21)	3.18 (0.6, 17)	1.32 (0.3, 7)	1.65 (0.4, 8)	2.19 (0.5, 11)	2.63 (0.6, 12)	0.83 (0.1, 5)	3.25 (0.6, 17)	**I**												
6.70 (1, 53)	4.37 (0.6, 40)	1.81 (0.2, 16)	2.27 (0.3, 19)	3.01 (0.4, 26)	3.58 (0.6, 26)	1.14 (0.1, 11)	4.47 (0.6, 41)	1.38 (0.1, 17)	**J**											
1.83 (0.6, 5)	1.20 (0.5, 3)	0.49 (0.2, 1)	0.6 (0.2, 2)	0.82 (0.3, 2)	0.98 (0.3, 3)	0.31 (0.1, 0.96)	1.21 (0.4, 3)	0.37 (0.1, 2)	0.27 (0.0, 2)	**K**			N=1, 2.2 (0.6, 8)							
1.52 (0.6, 4)	0.99 (0.3, 3)	0.41 (0.1, 1)	0.51 (0.2, 1)	0.68 (0.2, 2)	0.8 (0.3, 2)	0.26 (0.1, 0.9)	1.01 (0.4, 3)	0.31 (0.1, 2)	0.23 (0.0, 2)	0.83 (0.2, 3)	**L**									
16.63 (0.8, 840)	10.87 (0.5, 603)	4.48 (0.2, 247)	5.6 (0.3, 294)	7.49 (0.4, 402)	8.92 (0.4, 474)	2.81 (0.1, 160)	11.21 (0.5, 600)	3.47 (0.1, 218)	2.5 (0.1, 199)	9.24 (0.4, 518)	11.12 (0.5, 608)	**M**								
2.96 (1, 15)	1.94 (0.4,10)	0.80 (0.2, 4)	1.00 (0.2, 5)	1.33 (0.3, 7)	1.58 (0.3, 8)	0.50 (0.1, 3)	1.97 (0.4, 9)	0.60 (0.1, 5)	0.44 (0.0, 5)	1.63 (0.4, 7)	1.95 (0.3, 11)	0.18 (0.0, 5)	**N**		N= 2, 1.32 (0.7, 2)					
1.1 (0.4, 3)	0.71 (0.2, 3)	0.30 (0.1, 1)	0.37 (0.1, 1)	0.49 (0.1, 2)	0.59 (0.2, 2)	0.19 (0, 0.7)	0.73 (0.2, 3)	0.22 (0, 1)	0.16 (0.0, 1)	0.60 (0.1, 3)	0.72 (0.2, 3)	0.06 (0.0, 2)	0.37 (0.1, 2)	**O**			N=2, 0.78 (0.4, 2)			
3.85 (0.6, 25)	2.51 (0.4, 17)	1.04 (0.2, 7)	1.3 (0.2, 9)	1.73 (0.3, 11)	2.06 (0.3, 14)	0.65 (0.1, 5)	2.57 (0.4, 16)	0.79 (0.1, 8)	0.57 (0.0, 8)	2.1 (0.4, 12)	2.54 (0.4, 18)	0.23 (0.0, 8)	1.3 (0.5, 3)	3.52 (0.4, 30)	**P**					
1.01 (0.4, 3)	0.66 (0.3, 2)	0.27 (0.1, 0.6)	0.34 (0.1, 0.96)	0.46 (0.2, 1)	0.54 (0.2, 2)	0.17 (0.1, 0.4)	0.68 (0.2, 2)	0.21 (0.0, 1)	0.15 (0.0, 1)	0.56 (0.2, 2)	0.67 (0.2, 2)	0.06 (0.0, 1)	0.34 (0.1, 2)	0.93 (0.2, 4)	0.26 (0.0, 2)	**Q**				
1.38 (0.5, 4)	0.89 (0.2, 3)	0.37 (0.1, 1)	0.46 (0.2, 1)	0.62 (0.2, 2)	0.73 (0.2, 2)	0.23 (0.1, 0.9)	0.91 (0.3, 3)	0.28 (0.1, 2)	0.20 (0.0, 2)	0.75 (0.2, 3)	0.90 (0.2, 3)	0.08 (0.0, 2)	0.46 (0.1, 3)	1.25 (0.5, 3)	0.36 (0.0, 3)	1.36 (0.3, 5)	**R**			
1.21 (0.4, 3)	0.78 (0.3, 2)	0.33 (0.1, 0.9)	0.41 (0.1, 1)	0.54 (0.2, 1)	0.65 (0.2, 2)	0.2 (0.1, 0.6)	0.80 (0.2, 3)	0.25 (0.0, 1)	0.18 (0.0, 2)	0.66 (0.2, 3)	0.79 (0.2, 3)	0.07 (0.0, 2)	0.40 (0.1, 3)	1.10 (0.3, 5)	0.31 (0.0, 2)	1.19 (0.4, 4)	0.88 (0.2, 4)	**S**		
4.20 (0.7, 25)	2.74 (0.5, 15)	1.13 (0.2, 6)	1.42 (0.2, 9)	1.89 (0.3, 11)	2.25 (0.4, 15)	0.71 (0.1, 4)	2.79 (0.4, 19)	0.85 (0.1, 9)	0.62 (0.04, 8)	2.30 (0.3, 16)	2.76 (0.4, 20)	0.25 (0.0, 9)	1.41 (0.1, 14)	3.83 (0.5, 30)	1.09 (0.1, 13)	4.14 (0.7, 26)	3.07 (0.4, 24)	3.48 (0.5, 24)	**T**	
0.56 (0.1, 3)	0.36 (0.1, 2)	0.15 (0, 0.8)	0.19 (0, 0.95)	0.25 (0. , 1)	0.3 (0.1, 2)	0.09 (0, 0.58)	0.37 (0.1, 2)	0.11 (0.0, 0.9)	0.08 (0.0, 0.96)	0.30 (0.0, 2)	0.36 (0.1, 2)	0.03 (0.0, 1)	0.19 (0.0, 2)	0.51 (0.1, 3)	0.14 (0, 2)	0.55 (0.1, 3)	0.4 (0.1, 3)	0.46 (0.1, 3)	0.13 (0.0, 1)	**U**

RCT, randomised controlled trial; Q-RCT, quasi-randomised controlled trial; OR, odds ratio

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.

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 intradiscal injections; T Spinal cord simulations; U Radiofrequency treatment ; N number of studies included in conventional pairwise meta-analysis.

Table A6.2: Results (weighted mean difference WMD, with 95% confidence intervals/credible intervals) of network meta-analysis for RCTs/Q-RCTs reporting pain intensity

A			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=1, -7.0 (-5, 19)					N=1, -26.66 (-38, -15)	N=1,13.00 (2, 24)
-4.61 (-22, 13)	**B**	N=1, -6.1 (-11, -0.8)	N=2, -5.32 (-12, 1)							N=1, -2.00 (-12, 8)								
-11.96 (-34, 10)	-7.35 (-25, 10)	**C**		N=1, 6.0 (-4, 16)		N=8,-5.17 (-12, 2)				N=1, 9.00 (-4, 22)								
-10.3 (-19, -2)	-5.70 (-22, 11)	1.65 (-20, 23)	**D**		N=2,18.01 (6, 30)						N=1, 35.00 (-25, 95)	N=1, -9.3 (-23, 4.9)						
-5.68 (-28, 17)	-1.07 (-24, 22)	6.23 (-14, 27)	4.56 (-18, 28)	**E**				N=1, -0.63 (-15, 14)								N=1, 18.00 (8, 28)		
-3.82 (-14, 6)	0.81 (-19, 20)	8.12 (-16, 32)	6.49 (-5, 17)	1.84 (-23, 26)	**F**									N=1, 22.50 (11, 35)				
-17.04 (-41, 7)	-12.46 (-32, 7)	-5.09 (-14, 4)	-6.74 (-30, 17)	-11.36 (-34, 11)	-13.23 (-38, 12)	**G**												
-1.64 (-21, 19)	2.99 (-23, 30)	10.3 (-19, 40)	8.64 (-13, 31)	4.10 (-26, 35)	2.19 (-20, 25)	15.40 (-15, 47)	**H**				N=2, 3.19 (-13, 19)		N=1, 19.00 (8, 30)					
-6.26 (-41, 29)	-1.56 (-37, 34)	5.75 (-28, 40)	4.06 (-31, 39)	-0.5 (-27, 26)	-2.39 (-38, 34)	10.8 (-24, 46)	-4.65 (-45, 35)	**I**										
-24.8 (-53, 3)	-20.16 (-54, 13)	-12.81 (-49, 23)	-14.55 (-44, 15)	-19.04 (-56, 17)	-21.0 (-51, 9)	-7.752 (-45, 29)	-23.2 (-58, 11)	-18.67 (-64, 26)	**J**									
-4.95 (-30, 20)	-0.33 (-20, 20)	7.04 (-13, 27)	5.39 (-19, 30)	0.76 (-26, 28)	-1.12 (-28, 26)	12.16 (-10, 34)	-3.30 (-36, 29)	1.25 (-37, 39)	19.86 (-18, 58)	**K**								
-0.36 (-18, 19)	4.31 (-21, 30)	11.61 (-17, 41)	9.93 (-10, 30)	5.38 (-23, 35)	3.41 (-17, 25)	16.68 (-13, 47)	1.31 (-15, 17)	5.90 (-33, 46)	24.48 (-9, 59)	4.60 (-26, 36)	**L**							
-7.49 (-25, 11)	-2.86 (-27, 21)	4.47 (-23, 32)	2.80 (-15, 21)	-1.73 (-30, 27)	-3.69 (-23, 17)	9.62 (-19, 39)	-5.85 (-33, 21)	-1.27 (-40, 38)	17.37 (-16, 51)	-2.57 (-32, 28)	-7.11 (-33, 18)	**M**						
17.35 (-14, 50)	22.01 (-14, 59)	29.48 (-9, 69)	27.64 (-5, 61)	23.14 (-16, 63)	21.22 (-12, 56)	34.56 (-5, 75)	18.99 (-6, 44)	23.61 (-24, 72)	42.28 (-0.1, 85)	22.38 (-18, 64)	17.73 (-12, 48)	24.8 (-12, 62)	**N**		N=2, -1.09 (-7, 5)			
4.19 (-17, 25)	8.78 (-19, 35)	16.11 (-14, 46)	14.43 (-8, 36)	9.80 (-21, 40)	7.97 (-12, 28)	21.2 (-10, 52)	5.78 (-24, 34)	10.36 (-31, 50)	28.89 (-6, 64)	9.04 (-24, 41)	4.51 (-24, 32)	11.62 (-16, 39)	-13.24 (-53, 24)	**O**			N=2, -8.23 (-18, 2)	
16.51 (-19, 53)	21.17 (-19, 62)	28.5 (-14, 71)	26.83 (-10, 65)	22.23 (-20, 66)	20.3 (-17, 59)	33.6 (-10, 77)	18.15 (-12, 49)	22.64 (-27, 74)	41.4 (-4, 88)	21.5 (-22, 66)	16.81 (-18, 51)	23.98 (-16, 65)	-0.85 (-18, 16)	12.46 (-29, 55)	**P**			
12.26 (-21, 46)	16.97 (-17, 51)	24.26 (-8, 57)	22.64 (-11, 56)	18.02 (-6, 43)	16.11 (-18, 51)	29.37 (-4, 63)	13.93 (-25, 53)	18.54 (-17, 55)	37.07 (-6, 81)	17.26 (-19, 54)	12.66 (-26, 50)	19.84 (-18, 57)	-5.13 (-52, 41)	8.17 (-31, 48)	-4.21 (-54, 45)	**Q**		
-12.11 (-32, 8)	-7.49 (-34, 19)	-0.16 (-30, 30)	-1.79 (-23, 20)	-6.39 (-37, 24)	-8.28 (-29, 13)	4.92 (-26, 36)	-10.38 (-40, 18)	-5.89 (-46, 34)	12.74 (-22, 47)	-7.27 (-40, 25)	-11.73 (-40, 15)	-4.59 (-32, 22)	-29.52 (-68, 8)	-16.23 (-33, 1)	-28.63 (-71, 12)	-24.45 (-64, 15)	**R**	
13.03 (-12, 38)	17.62 (-13, 48)	24.99 (-9, 59)	23.39 (-3, 50)	18.69 (-15, 53)	16.86 (-10, 44)	30.11 (-5, 65)	14.66 (-18, 47)	19.14 (-24, 62)	37.93 (0.2, 76)	17.97 (-18 54)	13.36 (-18, 44)	20.54 (-11, 51)	-4.301 (-45, 36)	8.86 (-24, 42)	-3.49 (-48, 40)	0.70 (-41, 42)	25.09 (-7, 57)	**U**

RCT, randomised controlled trial; Q-RCT, quasi-randomised controlled trial

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

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

RESULTS OF SENSITIVITY ANALYSES

The results of the sensitivity analyses excluding observational studies and non-RCTs 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. The relative effects of epidural, chemonucleolysis, non-opioids, manipulations, and acupuncture with usual care were no longer statistically significant in the network meta-analysis restricted to RCTs and Q-RCTs. A number of other comparisons were also no longer statistically significant, usually with the analyses restricted to RCTs having a wider credible interval (epidural injections vs chemonucleolysis, traction, exercise therapy, and spinal cord simulation (SCS); and SCS vs non-opioids, manipulation, and accupuncture). The following comparison, however, did become statistically significant within the restricted analyses: disc surgery was more effective than radiofrequency treatment (Odds ratio (OR) 6.68, 95% Credible Interval (CrI): 1.19 to 39.53); intra-operative injections were more effective than both non-opioids (OR 3.16, 95% CrI: 1.09 to 9.49), passive physical therapy (OR 3.89, 95% CrI: 1.11 to 13.84), opioids (OR 5.40, 95% CrI: 1.35 to 22.40), and neuropathic pain modulators (OR 4.33, 95% CrI: 1.09 to 17.51); and manipulation was more effective radiofrequency treatment (OR 8.84, 95% CrI: 1.06 to 76.22).

For pain intensity the most notable discrepancies between the network meta-analysis with and without observational studies and non-RCTs only occurred with biological agents (vs inactive control, conventional care, disc surgery, non-opioids, intra-operative interventions, acupuncture, exercise therapy, opioids, and neuropathic pain modulators). Only the mean difference between epidural injections and inactive control (weighted mean difference (WMD) -10.3 95% CrI: -18.5, -2.24) remained statistically significant in the network meta-analysis restricted to RCTs and Q-RCTs, predominantly through the inflation of confidence intervals due to the removal of evidence.

APPENDIX A7: ASSESSMENT OF MODEL FIT AND BETWEEN STUDY HETEROGENEITY

ASSESSMENT OF MODEL FIT

Table A7.1: Residual deviance data for all network meta-analyses

Model	No. of data points	Posterior mean deviance
Global effect including all studies	190	195
Global effect including RCTs and Q-RCTs	136	140
Pain intensity including all studies	107	105
Pain intensity including RCTs and Q-RCTs	92	90

ASSESSMENT OF BETWEEN STUDY HETEROGENEITY

Global effect

For the network meta-analyses of global effect that included all study designs, the median between-trial variance (tau-squared) observed in the posterior distributions was 0.38 (95% CrI: 0.22 to 0.65). However, this does not give an indication of the between study heterogeneity within each intervention comparison. This information has been derived from the standard pair-wise meta-analyses shown in the Table F II below. There was high a level of between study heterogeneity for the following comparisons: disc surgery vs conventional care; and percutaneous discectomy vs disc surgery.

Table A7.2: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for global effect which included all study types

Treatment comparators	Chi-squared statistic	Degrees of freedom	P-value	I-squared**	Tau-squared
CB	10.5	4	0.038	60.6%	0.1806
DA	26.58	8	0.001	69.9%	0.7921
DB	8.82	2	0.012	77.3%	2.2995
EA	5.01	4	0.286	20.2%	0.0604
EC	41.08	13	0.000	68.4%	0.2582
FA	7.74	6	0.258	22.5%	0.0659
FC	0.00	0	.	.%	0.0659
FD	0.00	0	.	.%	0.0659
GC	2.32	6	0.888	0.0%	0.0000
GE	13.10	3	0.004	77.1%	0.7505
HA	0.18	1	0.669	0.0%	0.0000
HK	0.00	0	.	.%	0.0000
IA	0.00	0	.	.%	0.0000
JF	0.00	0	.	.%	0.0000
KB	0.00	0	.	.%	0.0000
KC	0.00	0	.	.%	0.0000
KH	0.00	0	.	.%	0.0000
LA	0.00	0	.	.%	0.0000
LD	0.00	0	.	.%	0.0000
LH	0.00	0	.	.%	0.0000
MA	0.00	0	.	.%	0.0000
ND	0.00	0	.	.%	0.0000
NH	0.00	0	.	.%	0.0000
NK	0.00	0	.	.%	0.0000
OA	0.00	0	.	.%	0.0000
OF	0.00	0	.	.%	0.0000
PN	0.76	1	0.384	0.0%	0.0000
QC	15.43	4	0.004	74.1%	0.3051
QE	7.85	4	0.097	49.0%	0.1483
RA	1.61	1	0.205	37.7%	0.1814

Treatment comparators	Chi-squared statistic	Degrees of freedom	P-value	I-squared**	Tau-squared
RO	0.29	1	0.593	0.0%	0.0000
SE	6.97	3	0.073	57.0%	0.3459
TC	0.00	0	.	.%	0.3459
UA	0.00	0	.	.%	0.3459
Overall	*396.34*	*94*	*0.000*	*76.3%*	*0.5917*

** I-squared: the variation in OR attributable to heterogeneity

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 intradiscal injections; T Spinal cord simulations; U Radio frequency treatment.

For the network meta-analyses of global effect that included only the RCTs and quasi-RCTs, the median between-trial variance (tau-squared) observed in the posterior distributions was 0.30 (95% CrI: 0.12 to 0.62). There was high a level of between study heterogeneity for the following comparisons: disc surgery vs conventional care; and percutaneous discectomy vs disc surgery (see Table F III).

Table A7.3: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for global effect which included only RCTs or quasi-RCTs

Treatment comparators	Chi-squared statistic	Degrees of freedom	P-value	I-squared**	Tau-squared
CB	8.96	2	0.011	77.7%	0.4647
DA	26.58	8	0.001	69.9%	0.7921
DB	0.17	1	0.680	0.0%	0.0000
EA	5.01	4	0.286	20.2%	0.0604
EC	7.26	4	0.123	44.9%	0.1757
FA	7.74	6	0.258	22.5%	0.0659
FD	0.00	0	.	.%	2.2417
GC	2.32	6	0.888	0.0%	0.0000
HA	0.18	1	0.669	0.0%	0.0000
KH	0.00	0	.	.%	0.3077
IA	0.00	0	.	.%	0.0000
JF	0.00	0	.	.%	0.3077
KB	0.00	0	.	.%	0.0000
KC	0.00	0	.	.%	0.0000
LA	0.00	0	.	.%	0.0000
LD	0.00	0	.	.%	2.2417
LH	0.00	0	.	.%	0.3077
MA	0.00	0	.	.%	0.0000
NH	0.44	1	0.506	0.0%	0.0000
OA	0.00	0	.	.%	0.0000
OF	0.00	0	.	.%	0.3077
PN	0.76	1	0.384	0.0%	0.0000
QC	7.25	1	0.007	86.2%	2.2417
QE	0.08	1	0.783	0.0%	0.0000
RA	1.61	1	0.205	37.7%	0.1814
RO	0.29	1	0.593	0.0%	0.0000
SE	5.13	2	0.077	61.0%	0.3077
TC	0.00	0	.	.%	2.2417
UA	0.00	0	.	.%	0.1814
Overall	*201.29*	*68*	*0.000*	*66.2%*	*0.4619*

** I-squared: the variation in OR attributable to heterogeneity

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 intradiscal injections; T Spinal cord simulations; U Radiofrequency treatment.

Pain intensity

For the network meta-analyses of pain intensity that included all study designs, the median between-trial variance (tau-squared) observed in the posterior distributions was 119.9 (95% CrI: 64.34 to 227.9). There was high a level of between study heterogeneity for the following comparisons: disc surgery vs inactive control; non-opioids vs inactive control; biological agents vs inactive control; intra-operative interventions vs disc surgery; and non-opioids vs epidural injections (see Table F IV).

Table A7.4: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for pain intensity which included all study types

Treatment comparator (number of studies)	Heterogeneity Statistic	degrees of freedom	P	I-squared**	Tau-squared
DA	56.13	6	0.000	89.3%	198.6672
EA	0.00	0	.	.%	0.0000
FA	18.77	3	0.000	84.0%	65.8051
HA	0.00	0	.	.%	0.0000
JA	0.00	0	.	.%	0.0000
LA	0.00	0	.	.%	0.0000
MA	13.69	1	0.000	92.7%	535.7751
RA	0.00	0	.	.%	0.0000
UA	0.00	0	.	.%	0.0000
CB	1.02	1	0.313	1.8%	0.7251
DB	0.03	1	0.868	0.0%	0.0000
KB	0.00	0	.	.%	0.0000
EC	6.20	2	0.045	67.7%	56.4674
FC	0.00	0	.	.%	0.0000
GC	29.82	7	0.000	76.5%	72.3876
KC	0.00	0	.	.%	0.0000
FD	7.92	1	0.005	87.4%	62.9147
JD	0.00	0	.	.%	0.0000
LD	0.00	0	.	.%	0.0000
IE	0.00	0	.	.%	0.0000
QE	0.00	0	.	.%	0.0000
MF	0.00	0	.	.%	0.0000
OF	0.00	0	.	.%	0.0000
LH	3.86	1	0.049	74.1%	98.4535
NH	0.00	0	.	.%	0.0000
PN	0.16	1	0.689	0.0%	0.0000
RO	0.67	1	0.414	0.0%	0.0000

** I-squared: the variation in WMD attributable to heterogeneity

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.

For the network meta-analyses of pain intensity that included only the RCTs and quasi-RCTs, the median between-trial variance (tau-squared) observed in the posterior distributions was 122.4 (95% CrI: 62.15 to 249.9). There was high a level of between study heterogeneity for the following comparisons: disc surgery vs inactive control; non-opioids vs inactive control; biological agents vs inactive control; intra-operative interventions vs disc surgery; and non-opioids vs epidural injections (see Table F V).

Table A7.5: Test(s) of heterogeneity for studies included in the standard pair-wise analyses for pain intensity which included only RCTs or quasi-RCTs

Treatment comparator (number of studies)	Heterogeneity Statistic	degrees of freedom	P	I-squared**	Tau-squared
DA	56.13	6	0.000	89.3%	198.6672
EA	0.00	0	.	.%	0.0000
FA	18.77	3	0.000	84.0%	65.8051
HA	0.00	0	.	.%	0.0000

Treatment comparator (number of studies)	Heterogeneity Statistic	degrees of freedom	P	I-squared**	Tau-squared
JA	0.00	0	.	.%	0.0000
LA	0.00	0	.	.%	0.0000
MA	13.69	1	0.000	92.7%	535.7751
RA	0.00	0	.	.%	0.0000
UA	0.00	0	.	.%	0.0000
CB	0.00	0	.	.%	0.0000
DB	0.03	1	0.868	0.0%	0.0000
KB	0.00	0	.	.%	0.0000
EC	0.00	0	.	.%	0.0000
GC	29.82	7	0.000	76.5%	72.3876
KC	0.00	0	.	.%	0.0000
FD	7.92	1	0.005	87.4%	62.9147
LD	0.00	0	.	.%	0.0000
IE	0.00	0	.	.%	0.0000
QE	0.00	0	.	.%	0.0000
OF	0.00	0	.	.%	0.0000
LH	3.86	1	0.049	74.1%	98.4535
NH	0.00	0	.	.%	0.0000
PN	0.16	1	0.689	0.0%	0.0000
RO	0.67	1	0.414	0.0%	0.0000

** I-squared: the variation in WMD attributable to heterogeneity

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.

B. APPENDIX FOR CHAPTER 4: REVIEW OF NICE GUIDANCE

INCLUDES:

APPENDIX B1: NICE GUIDANCE WHERE TREATMENT SEQUENCING WAS HIGHLIGHTED AS AN ISSUE

The NICE guidance that highlighted treatment sequencing to be an important consideration somewhere within the document are summarised in Table 2A. This includes an extract of the text, or summary of the reason why it was identified by the search. These are ordered by condition, and the earliest date. Where relevant a reference to the associated health technology assessment (HTA) publication is provided (highlighted as underlined text).

Table B1.1: Extracts from NICE guidance/guidelines identified during the NICE website search as showing treatment sequences to be a pertinent issue (*ordered by condition, and the earliest date*)

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
INFLAMATORY ARTHRITIS		
Ankylosing spondylitis - adalimumab, etanercept and infliximab (TA143) May 2008 [HTA – McLeod, 2007]	Sequential treatments included in recommendations for research ERG appraisal: Evaluation of clinical effectiveness did not consider treatment sequences. Economic model based on single agent framework due to lack of sequencing studies.	Studies to examine whether ankylosing spondylitis responds to more than one **TNF-a inhibitor given sequentially**. The collection of data through a register of people with ankylosing spondylitis receiving TNF-a inhibitor treatment in England
Ankylosing spondylitis - golimumab (TA233) August 2011	Sequential use of TNF-a inhibitors: the Committee heard from the manufacturer that there was no evidence on the efficacy of golimumab when used in sequence with the other TNF-a inhibitors	
Psoriatic arthritis - etanercept, infliximab and adalimumab (TA199) August 2010 [HTA – Rodgers, 2011]	The Committee concluded that there were insufficient data to make a recommendation on the sequential use of TNF inhibitors in psoriatic arthritis. The Committee was aware of registries that collect data for the long-term outcomes ERG appraisal: Evaluation of clinical effectiveness did not consider treatment sequences, but this was prominent in recommendations for future research. Evaluation of cost effectiveness included modelling sequential TNF-inhibitors as part of sensitivity analysis.	
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 [HTA – Malottki, 2011]	The manufacturer's submission included a model of sequential TNF-inhibitors. The Committee noted that, apart from the RCTs of rituximab and abatacept, the available evidence on the effectiveness of treatment with the considered technologies after the failure of a TNF inhibitor was mainly derived from observational studies with short follow-up periods that included relatively small numbers of participants. The Committee noted that many of the studies lacked a comparison group, so it was not clear what would have happened had participants not received therapy. The Committee considered that shortcomings in the design of studies of the sequential use of TNF inhibitors could affect the validity of the results. *Cost-effectiveness:* The Committee examined the cost-effectiveness analysis of sequential use of TNF inhibitors performed by the Assessment Group and the manufacturers of the technologies. The Committee noted that all analyses modelled a sequence of treatments, which it considered appropriate for rheumatoid arthritis. The Committee noted, however, that there were differences in the sequences modelled. The Committee noted that the use of non-randomised comparisons could affect the robustness of the results, but it accepted that the evidence base available for the sequential use of adalimumab, etanercept and infliximab did not currently allow for a robust analysis of the relative treatment effects.	Further clinical trials should be undertaken to compare the clinical effectiveness of adalimumab, etanercept and infliximab used sequentially after the failure of a TNF inhibitor with the clinical effectiveness of management strategies that do not include TNF inhibitors, including strategies that use untried DMARDs or biological DMARDs such as rituximab.
Rheumatoid arthritis (after the failure of previous anti-rheumatic drugs) - golimumab (TA225) June 2011	*Manufacturer's submission*: The manufacturer's economic model evaluated golimumab as part of a sequence of treatments. One model evaluated golimumab in people who had had previous treatment with conventional DMARDs only, and the other in people who had had treatment with both conventional DMARDs and a TNF inhibitor. *Consideration of the evidence*: The committee considered the economic model that evaluated golimumab as part of a sequence of treatments in people who had had previous treatment with conventional DMARDs only and who had not had a previous TNF inhibitor	
Rheumatoid arthritis - tocilizumab (rapid review TA198) (TA247) February 2012	The manufacturer's economic model included treatment sequence that included tocilizumab. The Committee noted the DSU report, which clarified that in order to estimate the cost effectiveness of tocilizumab, each treatment sequence should have been calculated incrementally. In 2010, the DSU was asked to undertake additional cost-effectiveness analyses to validate the manufacturer's ICERs submitted following the third round of consultation, and to conduct sensitivity analyses to address the Appraisal Committee's concerns about key parameter assumptions. The 2010 report highlighted a key issue with the calculation of the ICERs presented by the manufacturer. This concerned the 'pair-wise' calculation of sequences containing tocilizumab plus methotrexate with the same sequence excluding tocilizumab rather than an 'incremental' comparison of all strategies containing tocilizumab plus methotrexate with each other and with a base-case strategy without tocilizumab. .. The DSU's 2010 report explained that an ICER calculated through a pair-wise comparison does not demonstrate that the sequence can be considered cost effective because there are a series of mutually exclusive sequences available and only one can be selected at any one time.	

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
Rheumatoid arthritis - abatacept (2nd line) (rapid review of TA234) (TA280) April 2013 [WITHDRAWN; replaced **by TA375**]	*Cost effectiveness* People who discontinue their allocated treatment either in the initial phase or the long-term phase, regardless of their initial treatment, enter the next phase of treatment with a sequence of conventional DMARDs (leflunomide, gold, azathioprine, ciclosporin, penicillamine), and then palliative treatment.	
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**) January 2016 [HTA: Stevenson, 2016]	*Further analyses by the Assessment Group*: The Assessment Group also explored the effect of sequencing on the ICERs. In one analysis they removed tocilizumab and rituximab from the treatment sequence to test the effect of using only 1 biological DMARD before switching to non-biological therapy. *Cost-effectiveness*: The Assessment Group presented the median ICERs for biological DMARDs for the 3 different populations (the severe active and moderate active disease populations who had been previously treated with methotrexate, and the severe active population who had not been previously treated with methotrexate).	
Stevenson, 2013 [ID537 in progress] Update of TAs 130, 186, 224, 234, 225, and 247 (HTA: Stevenson, 2016)	*Clinical effectiveness: Protocol did not indicate that treatment sequences will be conceded.* *Cost effectiveness: Decision problem was about initiating first biologic, but subsequent treatments were considered in the model limited by the recommended options in the guidance provided for TAs 195, 225, and 247. (RTX followed by cDMARDs; RTX followed by TOC followed by cDMARDs; and where RTX is not suitable, a second biologic followed by cDMARDs.) [Undertaken by Sheffield team]*	
CANCER		
Breast cancer (advanced) (CG81) February 2009	Chemotherapy: Considered the relative clinical and cost effectiveness of currently recommended treatments, either in combination or in sequence. Endocrine therapy: Although there is good evidence to support the use of aromatase inhibitors for postmenopausal women with ER-positive tumours, there is little evidence to determine what is the best sequence of alternative hormone treatments when they progress.	
Breast cancer - bevacizumab (in combination with a taxane) (TA214) February 2011	*Manufacturer's model*: patients were assumed to be in one of three possible discrete health states at any given time: 'progression-free survival', 'progressed' or 'death'. It was assumed that patients would have the same risk of dying after disease progression regardless of the first-line treatment they had received. In addition, the model assumed that patients would have the same sequence of further treatment and resource use after disease progression, regardless of their initial treatment. ERG de Novo economic model included treatment sequences.	
Breast cancer (metastatic) - fulvestrant (TA239) December 2011	There is little or no clinical evidence about the optimal treatment sequence for advanced breast cancer beyond first-line treatment. The Committee considered that the most likely position of fulvestrant in UK clinical practice would remain as a third-line or fourth-line treatment after therapy with aromatase inhibitor The ERG speculated that the apparent increased benefit for fulvestrant after an anti-oestrogen rather than after an aromatase inhibitor may be influenced by where in the treatment sequence most patients received fulvestrant, rather than by whether the last treatment before fulvestrant	
Colorectal cancer (advanced) - irinotecan, oxaliplatin and raltitrexed (TA93) (replaced by CG131) August 2005 [HTA Hind, 2008]	*Cost effectiveness:* considered an indirect comparison of the treatment sequences used in the GERCOR study with the treatment sequence currently recommended by NICE (5-FU/FA monotherapy followed by irinotecan monotherapy on progression). ERG appraisal: Evaluation of clinical effectiveness considered chemotherapies used at single points in the pathway (with the caveat that the results of 1st line treatment needs to be interpreted with caution due to unplanned 2nd-line treatment) and as sequences (2 sequencing studies identified). Evaluation of cost effectiveness included modelling treatment sequences.	The Committee noted the need for the full sequence of treatments to be recorded for all patients in all trials.
Colorectal cancer (metastatic) - bevacizumab and cetuximab (TA118) (partially updated by TA242) January 2007 [HTA (TA118): Tappendern, 2007; HTA (TA242): Hoyle, 2008]	*[Assessment group model]* Data on second-line and subsequent therapies were taken from a study that investigated the optimal sequence of FOLFOX and FOLFIRI as first and second-line therapies, and were applied equally to treatment and control groups. *Consideration of the evidence:* The experts suggested that it would be of merit to add further options and lines of therapy to this sequence. *Manufacturer's model (bevacizumab) was a simple 3 state model: pre-progression, post-progression and death.* Data on progression-free survival for the treatment and control arms were taken from trial data, and an equal risk of death was applied following progression irrespective of treatment group. ERG appraisal: Evaluation of clinical effectiveness did not consider treatment sequences. Main outcome was OS, but subsequent lines of therapy and palliative treatments were not taken into account. Evaluation of cost effectiveness did not consider treatment sequences. Economic evaluation based on updates of the models submitted to NICE by the manufacturers to correct for reported flaws in the survival analyses.	

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
	TA242 (not id. by website search, but checked for inclusion): ERG appraisal: Evaluation of clinical effectiveness did not consider treatment sequences. Economic model did not consider treatment sequences due to lack of data.	
Colorectal cancer (CG131) November 2011	Considered which combination and sequence of chemotherapy to use. ERG de Novo economic model included treatment sequences.	
Leukaemia (lymphocytic) - fludarabine (TA119) [Chronic lymphocytic leukaemia (CLL)] February 2007	The Committee was aware that improvements in PFS ... may not translate directly into OS benefits. However, the Committee heard from the clinical specialists that an international workshop in CLL had agreed that it was appropriate to use PFS as a surrogate endpoint for OS in CLL. This was principally because the prolonged nature of the CLL disease pathway and the use of sequential therapies at different times in the treatment pathways make estimation of differences in overall survival problematic and unreliable. The Committee was persuaded that PFS is a meaningful clinical endpoint for CLL patients. The Committee further considered evidence from clinical specialists that the choice of treatment for CLL and the sequence in which treatments are used is made on an individual patient…	
Leukaemia (chronic myeloid, first line) - dasatinib, nilotinib and standard-dose imatinib (TA251) April 2012 [WITHDRAWAN}	There is extensive uncertainty around the possible treatment sequences following first-line tyrosine kinase inhibitor treatment failure and modelling of short-term survival data. The Committee therefore concluded that the Assessment Group had adequately addressed this structural uncertainty by presenting a range these analyses were an important addition to the Assessment Group's model because they enabled a comparison in scenarios 3 and 4 of all the relevant first- and second-line treatment sequences.	
Multiple myeloma - lenalidomide (TA171) June 2009	The Committee understood that multiple myeloma is an incurable disease. .. and its course are heterogeneous and that for relapsed multiple myeloma the choice of therapy for a particular person is influenced by the initial treatment and their response to it, ... The Committee noted the importance that patients, their carers and physicians placed on having effective options to treat multiple myeloma at presentation and at subsequent relapses. However, it understood that the optimal sequence of agents to use is as yet unclear and depends on several factors, including a person's treatment history, comorbidities and disease characteristics.	
Ovarian cancer (advanced) - paclitaxel, pegylated liposomal doxorubicin hydrochloride and topotecan (TA91) May 2005 [WITHDRAWN] [HTA: Main. 2006]	…when treatments are used sequentially, persistent adverse effects from one previous treatment may necessitate a switch of therapy at the next stage. ERG report: did not consider treatment sequences, but data for the two population subtypes were analyses separately: platinum sensitive, and platinum resistant/refractory.	
Renal cell carcinoma - sunitinib (TA169) March 2009	Recommendations for research acknowledged the presence of sequential treatment trials	There are a number of ongoing trials that are actively recruiting participants and that are relevant to this appraisal. Some of these trials are investigating the optimum sequences of treatment.
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 [HTA: Coon, 2010]	*Second-line treatment for people in whom sunitinib has failed*: [*suggestion from manufacturer of sorafenib*] that consideration should be given to the sequencing of treatments (particularly sunitinib as a first-line treatment followed by sorafenib as second-line treatment). ….. The Committee noted that the evidence base for this treatment pathway was absent, because participants were excluded from the sorafenib RCT if they had received sunitinib as a first-line treatment and the sunitinib RCT only included people who were suitable for immunotherapy. In the absence of robust data, the Committee could not reach any conclusions on whether sorafenib could be considered a clinically effective second-line treatment for people with advanced RCC who had received sunitinib as a first-line treatment. ERG report: did not consider treatment sequences.	Recommendations for future research included acknowledgement of ongoing trials investigating optimum sequences of treatment.
Renal cell carcinoma (advanced, second-line treatment - everolimus (TA219) April 2011	Clinical specialist noted that an increase in PFS [*observed in RECORD-1 placebo-controlled trial*] would be expected to result in an increase in OS because gains in OS had been observed in clinical practice with the introduction of sequential chemotherapy for advanced RCC. 81% of people on placebo had crossed over to receive everolimus.	
MENTAL HEALTH DISORDERS		
Antisocial personality disorder (CG77) January 2009	Sequential treatments included in recommendations for research	*Treatment of comorbid anxiety disorders in antisocial personality disorder* Current treatment guidelines set out clear pathways for the stepped or sequenced care of people with anxiety disorders. An RCT of people with antisocial personality disorder and comorbid anxiety disorders that compares a **sequenced treatment programme**

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
		for the anxiety disorder with usual care should be conducted.
Attention deficit hyperactivity disorder (ADHD) - methylphenidate, atomoxetine and dexamfetamine (review) (TA98) 22 March 2006	*The manufacturer submission*: The manufacturer's model was based on a sequence of up to 4 treatment lines. The first 3 related to treatment with an antipsychotic drug and the fourth included lithium treatment for participants whose condition was resistant to previous therapy. It was assumed the effectiveness of each antipsychotic intervention was not influenced by its position in the treatment pathway *Consideration of the evidence:* The Assessment Group's modelled sequences of the three drugs. This decision was based on their finding that each active treatment was cost effective relative to no treatment. It was therefore considered reasonable to assume that it would always be cost effective to change to the next untried drug, rather than stopping treatment after the first or second drug is found to be ineffective or not tolerated. This analysis relies on modelling assumptions, two of which are that response to one drug is independent of the response to another, and that response and withdrawal rates for second- and third-line treatments are the same as those for first-line treatment. The Committee noted that some of the included sequences might be unsuitable for some individuals because of considerations of adverse events, comorbidities and concordance with therapy. On this basis and given the limitations inherent in the models, the Committee was unable to draw conclusions on the relative cost effectiveness of different drug treatment strategies.	
Bipolar disorder (adolescents) - aripiprazole (TA292) July 2013	Cost-effectiveness was based on a sequence of up to 4 treatment lines. The first 3 related to treatment with an antipsychotic drug and the fourth included lithium treatment for participants whose condition was resistant to previous therapy.	
Common mental health disorders (CG123) May 2011		For people with both anxiety and depression, which disorder should be treated first to improve their outcomes? Comorbidity between depression and anxiety disorders is common. At present there is little empirical evidence to guide healthcare professionals or patients in choosing which disorder should be treated first. Given that for many disorders the treatment strategies, particularly for psychological approaches, can be very different, guidance for healthcare professionals and patients on **the appropriate sequencing of psychological interventions would be likely to significantly improve outcomes**. This should be tested in a randomised trial in which patients who have a dual diagnosis of an anxiety disorder and depression, and where there is uncertainty about the appropriate sequencing of treatment, should be randomised to different sequencing of treatment.
Depression in children and young people (CG28) September 2005	Comorbid diagnoses and developmental, social and educational problems should be assessed and managed, either in sequence or in parallel, with the treatment for depression.	
Depression in adults (update) (CG90) October 2009	Considered sequencing antidepressant treatment after inadequate initial response.	What is the best medication strategy for people with depression who have not had sufficient response to a first SSRI antidepressant after 6 to 8 weeks of adequate treatment?
Obsessive compulsive disorder (OCD) and body dysmorphic disorder (BDD) (CG31) November 2005	It was noted that there is no evidence of the optimal sequence of the listed treatment options [be considered]	
Psychosis with coexisting substance misuse (CG120) March 2011	When developing a treatment plan for a person with psychosis and coexisting substance misuse, tailor the plan and the sequencing of treatments to the person and take account of • the relative severity of both the psychosis and the substance misuse at different times and • the person's social and treatment context and • the person's readiness for change.	
Schizophrenia (update) (CG82) March 2009	Key priorities for implementation: the sequential use of adequate doses of at least two different antipsychotic drugs. At least one of the drugs should be a non-clozapine second-generation antipsychotic.	

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
	DeNovo economic model developed to assess impact of lack of effectiveness in relapse prevention, intolerance, and unacceptability; not to assess or recommend specific drug sequences.	
Schizophrenia - aripiprazole (TA213) January 2011	The Committee considered the manufacturer's updated economic model that compared sequences of treatments starting with aripiprazole with sequences starting with risperidone. The revised model contained four additional treatment sequences specified in the Appraisal Consultation Document: … The Committee was mindful that in people … who are intolerant of or have a contraindication to risperidone, or whose schizophrenia has not been adequately controlled with risperidone, the case for aripiprazole is more plausible. The Committee considered whether there was any evidence to suggest that aripiprazole should be used ahead of, or only after olanzapine or quietapine in the treatment pathway for schizophrenia. It noted that the economic analyses suggest little difference between sequences in which aripiprazole precedes olanzapine and vice versa; and although sequences that contain aripiprazole are suggested to be more cost effective than the sequence that contains quetiapine (sequence D), the Committee was concerned that the cost of quetiapine was unfairly calculated in the manufacturer's economic model, …	
HEPATATIS		
Hepatitis B (chronic) - adefovir dipivoxil and pegylated interferon alfa-2a (TA96) (partially updated by CG165) February 2006 [HTA: Shepherd, 2006]	*[Cost-effectiveness: The Assessment Group therefore produced an analysis that considered more clinically relevant scenarios in which people could receive a sequence of drug treatments as necessary.]* The Committee discussed the cost-effectiveness analysis of the various treatment sequences that could be used over the entire disease process. ERG appraisal: Evaluation of clinical effectiveness did not consider treatment sequences; narrative synthesis conducted due to heterogeneity. Evaluation of cost effectiveness included modelling treatment sequences.	
Hepatitis B - entecavir (TA153) August 2008	the Committee considered that although the totality of the evidence submitted supported the clinical effectiveness of entecavir, it was not in a position to advise on the relative clinical and cost effectiveness of different sequential treatment strategies The Committee considered that, without having reviewed all the evidence on the range of possible treatment sequences, it was not in a position to recommend one treatment algorithm over another and that such a recommendation was beyond the scope of this appraisal.	
Hepatitis B - telbivudine (TA154) August 2008	The manufacturer's seroconversion model evaluated treatment sequences	
Hepatitis B - tenofovir disoproxil fumarate (TA173) July 2009	The manufacturer's model considered sequences of first-, second- and third-line treatments and people were assumed to move on to the next treatment regimen if they developed resistance to their current treatment.	
Hepatitis C (genotype 1) - telaprevir (TA252) April 2012	The Committee noted that the comparison of the cost effectiveness of sequential strategies had not been specified in the scope for this appraisal and therefore concluded that it would not be appropriate to request these analyses from the manufacturer	
Hepatitis B (chronic) (CG165) June 2013	*Introduction:* With multiple treatment options that are efficacious and safe, the key questions are which patients need immediate treatment and what sequence and combination of drug regimens should be used, and which patients can be monitored and delay treatment. The model considered treatment sequences for HBeAg-positive /-negative chronic hepatitis B and compensated liver disease.	
OTHER		
Atrial fibrillation - dronedarone (TA197) August 2010	The treatment pathways evaluated by the manufacturer might not represent the full range of treatment strategies or sequences for dronedarone	
Chronic kidney disease [CKD] (stage 4 or 5): management of hyperphosphataemia (CG157) March 2013	Sequential treatments included in recommendations for research	For, what is the most effective sequence or combination of phosphate binders to control serum phosphate (in adults with stage 4 or 5 CKD, including those on dialysis)? It is thought that the longer people remain on calcium-based binders, the greater their risk of developing hypercalcaemia. However, no evidence was found on the most appropriate sequence or combination of phosphate binders a

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
		person should receive to control serum phosphate and serum calcium.
Cystic fibrosis (pseudomonas lung infection) - colistimethate sodium and tobramycin (TA276) March 2013	Taking into account the sequence of inhaled antibiotics currently used in the treatment pathway in the UK and the clinical specialists' opinion that clinicians would switch from one antibiotic to another (whatever the preparation), the Committee concluded that the most appropriate comparator for colistimethate sodium DPI would be nebulised colistimethate sodium and the most appropriate comparator for tobramycin DPI would be nebulised tobramycin.	
Epilepsy (partial onset seizures) - Retigabine as adjunctive treatment (TA232) *July 2011*	*The ERG stated that the economic model did not consider the sequencing of treatments. The ERG suggested that the model could have included all comparator treatments in different sequences. The ERG did not accept the validity of the modelled assumption that people whose epilepsy responds to treatment with retigabine do not experience any change in clinical response over time.* *…the ERG considered that a Markov model would have been more appropriate because it could have incorporated treatment sequencing, ….* *The Committee heard from the clinical specialists that people with epilepsy are likely to take anti-epileptic drugs over a lifetime, but that they would switch between drugs during this time.*	
Epilepsies: diagnosis and management (CG137) January 2012	Identified by web search, but no sequencing related terms or references identified in the NICE Guidance document. *Recommendations:* It is recommended that combination therapy (adjunctive or 'add-on' therapy) should only be considered when attempts at monotherapy with AEDs have not resulted in seizure freedom.	How do the newer AEDs compare in efficacy to the standard AEDs in the treatment of newly diagnosed epilepsy? [*most newer AEDs have not been evaluated as 1st-line treatments*]
Gout (Hyperuricaemia) - febuxostat (TA164) December 2008	The ERG requested that a sequence of strategies where patients progress to an alternative intervention (allopurinol, febuxostat or no treatment) following lack of response should be evaluated. The manufacturer declined the request, arguing that estimation of a sequential strategy was not feasible because of a lack of clinical data. In addition, the manufacturer argued that it was unethical to consider febuxostat as second-line therapy when it is cost effective as first-line therapy, and that the only appropriate comparison was that investigated in the pivotal RCTs; that is, first-line therapy. The ERG asserted that appropriate modelling assumptions could have been made to allow some exploratory analysis. The committee was mindful that the ERG had requested modelling of sequential use when patients progress to the need for alternative treatments (following lack of response to allopurinol treatment) or no-treatment options, and that the manufacturer had declined the request on the basis of lack of evidence.	
Gout (tophaceous, severe debilitating, chronic) - pegloticase (TA291) June 2013	At the clarification stage [*Manufacturer's submission*], several assumptions in the manufacturer's model were questioned. The model assumed that all patients who responded to pegloticase treatment will progress to maintenance treatment with either allopurinol (70%) or febuxostat (30%). However, the ERG stated that this treatment sequence would not have been appropriate for patients in whom conventional urate-lowering therapies are contraindicated or not tolerated. After pegloticase treatment, 10% of the responders were assumed to switch to best supportive care, instead of maintenance treatment with xanthine oxidase inhibitors.	
Low back pain - Early management of persistent non-specific low back pain (CG88) May 2009 [WITHDRAWN – updated and replaced by NG59]	Sequential treatments included in recommendations for research *Sequential treatments not considered in subsequent NG59 (2012)*	What is the effectiveness and cost effectiveness of sequential interventions (manual therapy, exercise and acupuncture) compared with single interventions on pain, functional disability and psychological distress, in people with chronic non-specific back pain of between six weeks and one year? ... There are substantial cost implications for those who do not respond to initial therapy and receive multiple back care interventions. It is unclear whether there is added health gain for this subgroup from either multiple or sequential use of therapies.
Neuropathic pain in adults: pharmacological management in non-specialist settings (CG96) March 2010 [WITHDRAWN – updated and replaced by CG173]	*Introduction*: there is considerable variation in practice in terms of how treatment is initiated, whether therapeutic doses are achieved and whether there is correct sequencing of therapeutic classes. … There is also uncertainty about which drugs should be used initially (first-line treatment) for neuropathic pain, and the order (sequence) in which the drugs should be used. *The evidence considered* – The GDG had access to an unpublished HTA report, which included a CUA of pharmacological interventions for painful diabetic neuropathy (PDN) or post-herpetic neuralgia (PHN). This was based on ITC analysis of placebo controlled RCTs. The GDG considered that it was not appropriate to use the results of the HTA report to examine sequencing of treatments as the model did not consider class effects, titration practices and treatment switching. These factors resulted in sequences based solely on the outcome of the economic model being clinically inappropriate.	

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
	Recommended sequence based on starting with the most cost-effective (highest mean net benefit) and if this does not provide sufficient pain relief move on to next cost effective drug.	
Neuropathic pain - pharmacological management (CG173) November 2013	Sequencing of therapeutic classes considered an issue (same text included in introduction as CG96) *De novo economic model conducted by GDG (Appendix F) did not consider treatment sequences:* The purpose of the model was not to estimate the cost effectiveness of treatment strategies over more than 1 line. There are insufficient data on the correlation of effectiveness on 1 drug having taken another in a different or same class to model multiple line treatment strategies. The model therefore focussed on the cost effectiveness of individual drugs as monotherapies. (Decision tree model with a time horizon of 20 weeks)	
Osteoporotic fractures - denosumab (TA204) October 2010	The manufacturer's model was not a treatment-sequencing model because of the lack of clinical evidence for such use.	
Peritoneal dialysis (CG125) July 2011	Sequential treatments included in recommendations for research	What is the most effective sequence of treatment? There is limited high-quality evidence on the effectiveness of different sequencing of modalities
Psoriasis - ustekinumab (TA180) September 2009	The ACD suggested the use of ustekinumab after the failure of TNF inhibitors. However, no evidence for the use of ustekinumab after an inadequate response to other TNF inhibitors existed.	
Psoriasis (CG153) October 2012	*Rapid escalation to systemic treatments:* At present the treatment pathway for people with psoriasis follows clinical need as no studies have been conducted to evaluate whether early intervention with systemic treatments alters prognosis. Consequently, patients with more severe disease sequence through all therapies in the treatment pathway… In this guideline, first-line therapy describes traditional topical therapies. Second-line therapy includes the phototherapies and systemic non-biological agents such as ciclosporin, methotrexate and acitretin. Third-line therapy refers to systemic biological therapies such as the tumour necrosis factor antagonists.	
Stroke and systemic embolism (prevention, non-valvular atrial fibrillation) - apixaban (TA275) February 2013	The ERG carried out further exploratory analyses that were not included in its revised base case. These included changes to the treatment sequence to allow second-line treatment with warfarin, apixaban, rivaroxaban or dabigatran.	
Thrombocytopenic purpura - eltrombopag (TA205) (replaced by TA293) October 2010 [WITHDRAWN]	The manufacturer's modelled the cost effectiveness of eltrombopag as part of a sequence of ITP treatments for people with persistent bleeding. In a separate analysis, eltrombopag was considered as part of a long-term treatment sequence with romiplostim, intravenous immunoglobulin, rituximab and anti-D immunoglobulin.	
Thrombocytopenic purpura - eltrombopag (TA293) July 2013	The ERG noted that the manufacturer did not address the optimal positioning of eltrombopag and romiplostim within the treatment sequence in the model (assumed they followed after rituximab, but preceded other drugs used in standard care). The ERG pointed out that there is uncertainty about the optimal place of eltrombopag and romiplostim if one is assumed to be more effective than the other. The ERG stated that the manufacturer should have explored additional sequences of treatment.	
Thrombocytopenic purpura - romiplostim (TA221) April 2011	*Manufacturer's submission*: The manufacturer's model describes a defined sequence of treatments, and questioned whether it was reasonable to exclude some treatments in the comparator arm. *Consideration of the evidence:* No studies were found that compared romiplostim with a specified sequence of active treatments or rescue therapies for the treatment of ITP	
Type 2 diabetes: newer agents (Short CG87) May 2009 [REPLACED by NG28]	*Overview:* Although lifestyle interventions (diet and physical activity) are the first-line treatments for the management of type 2 diabetes, most people subsequently need sequential addition of oral glucose-lowering drugs. Treatment sequences were not specifically evaluated *[The Assessment Group's systematic review of relevant cost and cost-effectiveness studies identified DES model by McEwan et al. 2007, which is included in the methodology review]*	
Type 2 diabetes - Dapagliflozin combination therapy (TA288) June 2013	*Manufacturer's submission*: For the triple therapy analyses, the manufacturers also revised the sequence of treatments in the revised model so that the starting treatment was triple therapy rather than dual therapy	

Condition, TA, and date [Associated HTA reference]	Consideration of the evidence / manufacturer's submission	Recommendations for future research
	Consideration of the evidence: It noted that the sequence of treatments in the manufacturers' revised economic model had been amended so that the approach was consistent with the dual therapy and insulin add-on analyses, with patients in the model starting treatment with triple therapy	
Type 2 diabetes in adults: management (NG28) December 2015 [REPLACES CG87]	Sequential treatments included in recommendations for research	What are the effects of stopping and/or switching drug treatments to control blood glucose levels, and what criteria should inform the decision? There is limited understanding of the short- and long-term effects of stopping a therapy and switching to another in terms of diabetes control (HbA1c levels), hypoglycaemic risk, weight gain, and cardiovascular morbidity and mortality. In addition, there is limited understanding of how quickly consideration should be given to stopping and switching to another drug treatment and, if stopping and switching may be needed, **what the optimal sequencing is of drug treatments**.
Urinary incontinence in women: management (CG171) September 2013	Sequential treatments included in recommendations for research	Sequence of treatment with botulinum toxin A and percutaneous sacral nerve stimulation for the treatment of overactive bladder after failed conservative (including drug) management?

Abbreviations: ACD, Appraisal consultation document; AED, anti-epileptic drug; CG, clinical guideline; CLL, chronic lymphocytic leukaemia; DES, discrete event simulation; DMARD, disease-modifying anti-rheumatic drug; DSU, decision support unit for NICE; ERG, Evidence Review Group; FAD, Final appraisal determination; HTA, Health technology appraisal; ICER, incremental cost-effectiveness ration; ITP, Idiopathic thrombocytopenic purpura; NG, NICE guidance; OS, overall survival; PFS, progression-free survival; RCT, randomised controlled trials; RTX, rituximab; TA, technology appraisal; TNF, tumour necrosis factor; TOC, tocilizumab

APPENDIX B2: NICE TECHNOLOGY APPRAISALS OR CLINICAL GUIDELINES INCLUDED IN THE REVIEW OF METHODS THAT WERE NOT IDENTIFIED BY THE NICE WEBSITE SEARCH

Table B2.1: NICE technology appraisals or clinical guidelines included in the methodology review but not identified by the NICE website search *(Ordered by condition and most recent date)*

NICE TA/CG reference	HTA reference used in methods review (author, year)	Clinical evaluation	Economic evaluation [model name used in methods review]
RHEUMATOLOGY			
Rheumatoid arthritis and juvenile poly-articular idiopathic arthritis - etanercept and infliximab (after failure of previous conventional DMARDs) (TA36) March 2002 [WITHDRAWN Replaced by TA130 and TA195]	Jobanputra, 2002		Modelled sequences [Preliminary Birmingham Model, BPM]
As above	Barton, 2004	Did not include a clinical evaluation; objectives were to overcome some of the limitations of the Birmingham Preliminary Model (BPM) The aim was to restructure the model so that different sequences of treatments could be considered, and to determine the sequence that best represented current practice in the UK.	Modelled sequences [Birmingham Rheumatoid Arthritis Model, BRAM]
Rheumatoid arthritis - anakinra November 2003 (TA72) [WITHDRAWN Replaced by CG79]	Clark, 2004		Modelled sequences [BRAM]
Rheumatoid arthritis - adalimumab, etanercept and infliximab October 2007 (TA130) [Updates TA36. WITHDRAWN updated and replaced by TA375]	Chen, 2006		Modelled sequences *[BRAM]* *(1st-, second- and 3rd-line biologic)*
OTHER			
Epilepsy in adults - newer drugs March 2004 (TA76) [WITHDRAWN replaced by CG137 – included in Table 3A]	Wilby, 2005	Treatment sequences were not specifically considered, but studies were pooled within the participant subtypes: refractory or newly diagnosed epilepsy.	Model considered sequential treatments; used simplifying assumptions. [York epilepsy model]
Epilepsy in children - newer drugs April 2004 (TA79) [WITHDRAWN replaced by CG137 – included in Table 3A]	Connock, 2006	Treatment sequences were not specifically considered, but studies were pooled within the participant subtypes: refractory or newly diagnosed epilepsy.	Modelled sequences [Birmingham epilepsy model]
Psoriasis (severe) in adults - etanercept and efalizumab July 2006 [WITHDRAWN replaced by CG175]	Woolacott, 2006	Did not consider treatment sequences	Model considered sequential treatments; used simplifying assumptions [York psoriasis model]

Abbreviations: aPC, advanced prostate cancer; DMARD, Disease-modifying anti-rheumatic drug; CG, clinical guidelines; RhA, Rheumatoid arthritis; TA, technology appraisal.

C. APPENDIX FOR CHAPTER 5: META-ANALYTIC METHODS FOR ASSESSING TREATMENT SEQUENCES

INCLUDES:

Appendix C1: Description of included studies using meta-regression or subgroup analysis

Table C1.1: Summary of included studies using regression methods

Table C1.2: Summary of included studies using subgroup-analysis

APPENDIX C1: DESCRIPTION OF INCLUDED STUDIES USING META-REGRESSION OR SUBGROUP ANALYSIS

Table C1.1: Summary of included studies using regression methods

Study	Clinical scenario and treatment sequencing	Type of data available	Outcome measures and comparison metric used	Evidence synthesis methods used	Covariates relating to treatment sequencing included in meta-regression
Lloyd, 2011	TNF-inhibitors used as 2nd and subsequent -line treatment for patients with rheumatoid arthritis with an inadequate response to 1-2 TNF inhibitors considered. (Studies of non-TNF-inhibitor biologics used for current or previous treatment, were excluded.)	Only observational studies identified (n=20, 2705 patients): - 5 controlled studies: cohorts receiving 2nd vs 1st line TNF-inhibitors (2nd/3rd biologic was RTX in one study); and - 15 non-comparative observational studies. 1 pilot RCT of pts with partial response to ETA (n=28) randomised to continue ETA or switch to INF was excluded, but sensitivity analyses showed that excluding it from the single-arm analysis did not impact results.	Binary outcomes: proportion of EULAR and ACR20 responders. Continuous outcome: mean overall improvement in DAS-28 and HAQ scores (single-arm analysis based on comparison with baseline data [1st-line biologic] in same patients).	Standard pairwise random effects meta-analyses of TNF inhibitors as a class. Univariate meta-regression used to try to explain heterogeneity using study level demographic covariates. Separate analyses conducted for single-arm data (including all available studies) and comparative studies estimating effect of sequential use of TNF-inhibitor vs 1st-line use (only one outcome with data reported by all 4 studies). Separate analyses were reported to have been conducted for individual agents and according to reasons for discontinuation, but the published results concentrated on biologics analysed as a class, and switching due to inefficiency, combining both primary and secondary inefficiency.	Study level demographic variables included (n = no. of studies): • number of previous conventional DMARDs (n=10); • duration of previous biologic (n=10); • disease duration (n=14); (*plus 3 other covariates*) Subgroup analyses also conducted of type of TNF-inhibitor, and reason for switching.
Schmitz, 2012	Biologics used as 1st-line treatment for patients with rheumatoid arthritis with inadequate response to MTX (*excluded MTX naïve patients and those who had an inadequate response to biologics targeting TNF inhibitors)*	16 placebo controlled placebo controlled RCTs (6,566 patients). The trials included the concurrent use of MTX in one or several treatment arms.	Continuous outcomes: HAQ (a measure of disease severity) improvement modelled as percentage improvement to baseline HAQ score Binary outcomes: relative risk of achieving ACR20, ACR50, ACR70, responder status *ARC criteria also used to develop continuous scales and HAQ scores dichotomised to produce binary data for the purpose of comparing the analyses of binary and continuous outcomes on the same outcome measures.*	Bayesian Hierarchical random effects MTC models used to conduct indirect treatment comparison meta-analyses, which were extended to incorporate multivariate meta-regressions, based on methods described by Nixon 2007*: The models presented by Nixon 2007 are based on binary data (developed to allow for concurrent treatment with MTX and baseline differences in disease severity). These were expanded for use on continuous outcome data using methods by Jensen, 2008, who presented a model for analysing HAQ, a continuous scale, but did not allow for multiple treatment arms and baseline treatment characterises. (*Jensen, 2008*** *presented a MTC model on the HAQ, but did not allow for multiple treatment arms or baseline characteristics*.)	Study level covariates adjusted for: • number of previous conventional DMARDs (ranged 1-3 in 9 RCTs with 4006 patients; not reported in remaining studies); • baseline HAQ; • disease duration As the results were not statistically significant these variables were not included in the final analyses.
Grothey, 2004	Investigated the impact of any subsequent chemotherapy on overall survival in RCTs of 1st-line treatment for advanced	10 treatment arms from 7 recently published phase III RCTs (only data from treatment arms using irinotecan-, or	Median overall survival.	Spearman rank correlation test supplemented by simple linear regression. As a sensitivity analysis, weighted regression was used with weights proportional to the trial's sample size	Arm level proportions correlated with overall survival: • patients receiving any second-line treatment;

	colorectal cancer. The standard 1st-line therapy were FU-LV with either irinotecan or oxaliplatin. The impact of receiving all 3 agents during the course of the disease was also evaluated.	oxaliplatin-based combination therapy were considered; 1991 patients)		The analysis was based on single arm level data. Preliminary analysis showed heterogeneity between monotherapy and combination therapy, indicating a survival advantage with combination therapy.	• patients receiving all three active agents during the course of their disease.

Abbreviations: ACR American College of Rheumatology; DAS disease activity score; DMARD disease modifying antirheumatic drug; ETA etanercept; EULAR European League Against Rheumatism; HAQ Health assessment questionnaire; FU-LV fluorouracil-leucovorin; MTC Mixed Treatment Comparisons; MTX methotrexate; RA rheumatoid arthritis; RCT Randomised controlled trial; RTX rituximab; TNF tumour necrosis factor.

* Nixon, R. A., N. Bansback, et al. (2007). "Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis." Statistics in Medicine 26: 1237–1254.

** Jansen, J. P., B. Crawford, et al. (2008). "Bayesian Meta-Analysis of Multiple Treatment Comparisons: An Introduction to Mixed Treatment Comparisons." Value in Health 11(5): 956-964.

Table C1.2: Summary of included studies using subgroup-analysis

Study	Interventions (patient population) Specific biological DMARDs evaluated	Overall statistical ES Methods	Subgroup analyses conducted that relate to treatment sequencing	Comments
Singh, 2009 (Cochrane umbrella review)	1st and subsequent line biologic agents (comparison of biologics in patients with or without inadequate response to previous treatment, including conventional DMARD, in same analysis) *INF, ADA, ETA, ANA, ABA, RTX:* as individual agents and class 'all biologics' (plus TNF-inhibitors vs other biologics in subgroup analysis)	Cochrane overview of reviews conducted to evaluate the efficacy (benefit and harm) of 'biologics' as a single group and as individual drugs (vs placebo). Planned stratified analysis conducted of the efficacy of 'biologics' (as a single group vs placebo) across different factors (n=7). Indirect treatment comparison of efficacy between biologics also conducted. A mixed-effects logistic regression using an arm-based random effect model within an empirical Bayes framework was performed (generalised linear mixed model). The stratified meta-analysis allowed for heterogeneity because of study and study x drug interaction. Comparison between biologics (ignoring prior treatment failed) based on the Bucher et al, 1997 method, by subtracting log transformed effect estimates derived from hierarchical models.	Biologics evaluated a class ('all biologics') in a series of subgroup analysis to evaluate effect of 7 factors including: • Previous treatment failed: conventional DMARD; biologic; none [DMARD vs biologic, and DMARD vs none] • Previous treatment failure with a TNF-inhibitor: yes vs no • Biologic is a TNF-inhibitor: yes vs no (other biologic) • Mean duration of disease: early <2 vs established 6–10 vs late >10 years	*Heterogeneity in placebo group due to concomitant DMARD, and MTX vs DMARD vs biologic in some trials. Subgroup analysis also included concomitant MTX (vs no MTX).* *Analyses susceptible to type II error due to small number of studies in some subgroups.*
Lloyd, 2011	2nd and/or 3rd line TNF-inhibitor (patients who had withdrawn from 1-2 TNF-inhibitors, but not all three; switch to a TNF-inhibitors not previously used) *INF, ADA, ETA:* TNF-inhibitors as a class and individual agents	Conventional pairwise MA using random effects model of uncontrolled studies based on change from baseline (percentage of patient responders). Comparative effect of 'sequential' TNF-inhibitor (2nd and 3rd line) vs 1st line also evaluated in a separate MA (based on 4 comparative observational studies). *Also included univariate random-effects meta-regression analyses to test interaction between treatment effect and the following covariates: number of previous DMARD, duration of previous biologic, disease duration (plus 3 others).*	• Sequence of biologics: type of TNF-inhibitor switched from (sequence) and to (TNF-inhibitor received) • Reason for switching: intolerance due to adverse events; primary inefficiency; secondary inefficacy; or inefficacy, primary or secondary No results reported, only statements that no significant difference identified in estimates by sub-groups for each outcome measure.	Only relevant observational studies identified (n=20); most uncontrolled. 4 studies made comparisons with other cohorts of patients taking a first TNF-inhibitor and may or may not include the group that subsequently switched. Considerable heterogeneity present not explained by meta-regression and subgroup analyses
Suarez-Almazor, 2007 (CADTH, Canada)	1st or 2nd line TNF-inhibitors (the use of TNF-inhibitors after failure of other treatments; and switching between INF and ETA)	Placebo controlled RCT pooled using conventional pair-wise meta-analysis; both fixed effect and random effects models	• Previous treatment failed (before starting TNF-inhibitors): MTX vs MTX-naïve [to *evaluate clinical*	Only observational studies (n=11; most uncontrolled) were available for switching between INF and ETA.

Study	Interventions (patient population) Specific biological DMARDs evaluated	Overall statistical ES Methods	Subgroup analyses conducted that relate to treatment sequencing	Comments
	INF, ETA: analysis restricted to the following comparisons: INF+MTX vs placebo+MTX; ETA+MTX vs placebo+MTX; and ETA vs MTX	used. Observational studies were not pooled using meta-analysis. Results of subgroups compared using informal indirect analyses, and statistical significance established by examining the point estimates of the treatment effect and whether the 95% CIs overlapped. No overlap was considered to be statistically significant at the 0.05 level.	*impact of introducing biologic as initial therapy vs after failure of other drugs*] • Duration of disease: <2 years (early) vs ≥2 years (established and late) No quantitative syntheses conducted of switching between the two TNF-inhibitors (2nd-line treatment) due to clinical heterogeneity; the observational studies used different strategies to analyse efficacy after switching.	
Rendas-Baum, 2011 (Pfizer, UK)	[1st,] 2nd, 3rd or 4th line TNF-inhibitor / Biologics (patients with inadequate response to one or more TNF-inhibitor; some included studies reported response to biologics used for the first time) *INF, ADA, ETA, GOL, CZP, ABA, ANA, RTX, TOC:* Evaluated by class: 'any biologic' (and as TNF-inhibitors vs other biologics in subgroup analysis)	Efficacy rates (clinical response to any biological agent) were estimated for groups of patients who differed according to no. of previous TNF-inhibitors. Pooled weighted averages, based on sample size (of each group), calculated to give an average response rate per no. of previous TNF-inhibitors (1, 2, 2+, 3, 4) for each outcome measure (7 binary measures). Various other subgroups also evaluated, including efficacy estimates stratified by: Type of biologic ('*TNF*-inhibitors' *vs 'other' biologics*); Study design (*RCT vs observational*); Reason for discontinuation of a 1st TNF; Duration of follow-up (based on 1 study). The results were presented as bar-graphs for visual comparison of trends, according to number of previous TNF-inhibitors. Pooled data based on single cohorts	Main analyses based on comparison of response rates (to any biologic) according to the no. of previous TNF-inhibitors used: 1, 2, 2+, 3, or 4 (Some studies did not report results disaggregated by no. of previous TNF-inhibitors, only results for biologic under evaluation in patient with inadequate response to at least one TNF-inhibitor; recorded as 2+. Some studies reported multiple switches for the same group of patients; contributing to more than one response rate) Further subgroup analyses included: • Type of biologic drug: TNF-inhibitor; other biologics • Response rates for 2nd-line TNF stratified by reason for discontinuing 1st TNF-inhibitor: intolerance; lack of efficacy; loss of efficacy *Results*: The response rates declined with increasing number of previous treatments; trend persisted with TNF-inhibitors or other biologics; response rates higher for patients who discontinued for safety reasons; there were exception to these trends for some outcomes, which was not always explained by study characteristics.	Analysis based on individual patient data taken from both RCTs (n=7; randomisation broken; only biological treatment arm used) and observational studies (n=21). (*Compared findings from RCTs and observational studies in sub-group analysis; response rates lower for RCTs*) Biologic agent switched to and from ignored and data pooled across different agents. Subgroup analysis for TNF-inhibitors vs other biologics did not consider other factors such as reason for discontinuation, at the same time. The results were based on subjective evaluation of patterns in bar graphs. The actual pooled response rates were not reported, and estimates of variances not calculated. The findings therefore highlight potential trends only. No statistical comparison of groups conducted and degree of heterogeneity not summarised quantitatively. The review highlights the problem of poor reporting by primary studies; s*ome studies did not state the actual drug used and presented results aggregated over TNF-*inhibitors *(ADA, ETA, INF). The scope of the available data did not allow the comparison of patients who switched to a 2nd biologic that was not a TNF-inhibitor. Only a few studies reported response rates according to subgroups relating to reason for discontinuation.*

Study	Interventions (patient population) Specific biological DMARDs evaluated	Overall statistical ES Methods	Subgroup analyses conducted that relate to treatment sequencing	Comments
Salliot, 2011	1st or 2nd and subsequent line biologic agent (Comparison of biologics in 2 clinical situations: after inadequate response to MTX or TNF inhibitors). *INF, ADA, ETA, CZP*, ABA, GOL, RTX, TOC: TNF-inhibitors (as a group) compared with non-TNF inhibitors (as individual agents and single group) in patients with inadequate response to MTX Only non TNF-inhibitors [ABA, GOL, RTX, TOC] included in RCTs of patients with inadequate response to TNF inhibitors – evaluated as single agents.	Frequentist adjusted indirect treatment comparison using Bucher method (initially pooled using random effects model).	• Previous treatment failed: DMARD-IR and anti-TNF-IR (separate evaluation of biologic as 1st and 2nd -line) The main patient characteristics were similar in the evaluation of 1st-line biologics (9 RCTs, making 16 comparisons). In the analysis of biologics after inadequate response to TNF-inhibitors (5 RCTs), there were differences in the number of TNF-inhibitor failures (only stated in 2 studies; ranged 58% to 92%), and reason for failure (proportion who withdrew due to lack of efficacy ranged from 1.5% to 44%). The inability to ascertain precise numbers was acknowledged as a limitation of the study; as was potential differences in duration of TNF-inhibitor administration and dose before considering inadequate response.	*Biologics with concomitant DMARD (MTX or other). All patients received concomitant MTX for all but one of the newer biologics (CZP*, GOL, RTX, TOC*); in the ABA trial 76% to 82% received concomitant MTX with the remaining patients receiving other synthetic DMARDs.*
Schoels, 2012	2nd, 3rd or 4th line biologic (new biologic used after trying 1-3rd lines of TNF-inhibitors: *INF, ADA, ETA*) 'New biologics' (ABA, GOL, RTX, TOC) *combined with DMARD*	Frequentist adjusted indirect treatment comparison using Bucher method	• No. of previous TNF-inhibitors: 1, 2, or 3 (efficacy [biologic vs placebo] after 1 previous TNF inhibitor compared with outcomes after multiple TNF-inhibitor failures for individual drugs; and comparison of response rates for TOC vs GOL after 1, 2 and 3 previous TNF-inhibitors at baseline. (Data only available for GOL and TOC) *Results:* *Efficacy rates did not differ significantly after multiple treatments (actual data not reported). Response rates for GOL vs TOC were similar after 1, 2 or 3 TNF-inhibitors, with a trend towards significance after 3.*	Only 1 relevant RCT identified for each new biologic. Proportion discontinuing previous biologic due to inadequate response only reported in GOL (58%) and TOC (95%) trials. Subgroup analysis of no. of previous TNF-inhibitor failures also only available for these trials. Long-term extension data excluded patients who had received rescue treatment after the primary endpoint of 14-16 weeks

Abbreviations: ABA abatacept; ADA adalimumab; ANA anakinra; CZP certolizumab pegol; DMARD disease-modifying anti-rheumatic drugs; ETA etanercept; GOL golimumab; INF infliximab; MTX methotrexate; RCT randomised controlled trial; RTX rituximab; TOC tocilizumab; TNF tumour necrosis factor.

D. APPENDIX FOR CHAPTER 6: SUMMARY OF INCLUDED MODELLING STUDIES AND SIMPLIFYING ASSUMPTIONS USED TO REPRESENT TREATMENT SEQUENCING EFFECTS

INCLUDES:

Appendix D1: Description of included modelling studies

Table D1.1: Summary of included non-rheumatology modelling studies

Table D1.2: Summary of included rheumatology modelling studies

APPENDIX D1: DESCRIPTION OF INCLUDED MODELLING STUDIES

Table D1.1: Summary of included non-rheumatology modelling studies (*Studies ordered alphabetically by condition and then author*)

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Cameron, 2008 UK	**Cancer** Women with HR+ metastatic breast cancer (BC) whose disease has progressed or relapsed after previous treatment with oestrogen therapy.	Adding fulvestrant (FULV), which is a hormonal agent, to a sequence of 4 therapies containing: a non-steroidal aromatase inhibitor (AI; one of 2 drugs), then exemastane (EXT; and AI), followed by two chemotherapy treatments (docetaxel then capacitabine). FULV was added as the 2nd or 3rd treatment.	SR and MA of RCTs and other experimental trials conducted for each treatment by line of therapy. No information was given on prior treatments. Time to progression (TTP) data for chemotherapies were obtained from individual studies that presented sufficient data. Pooled estimates for median TTP were obtained for: FUL used as 2nd-line (based on 2 studies) FUL used as 3rd-line (based on 1 study) NSAI as 1st-line (based on 2 studies for one drug, and 3 studies for another) EXT as 1st-line (3 studies) EXT as 2nd-line (3 studies; same studies as 1st-line but different estimate obtained for each line of treatment) EXT as 3rd-line (2 studies) Estimates of the proportion of patients skipping treatments were derived from clinician survey, as they were not available in the literature.	Treatment effects were assumed to be dependent on the number of previous treatments used, but independent of the type of treatments used. Sensitivity analysis included: - Assuming same efficacy for FULV and EXT at 2nd and 3rd –line. - To reflect the multiple potential chemotherapy options in advanced BC, the use of different chemotherapies in the sequence was tested. - To reflect uncertainty around chemotherapy effectiveness and use in later lines of therapy and increase median TPP of 6.5 mths was assumed for docetaxel.
Dranitsaris, 2011 Malaysia	**Cancer** Metastatic colorectal cancer.	Chemotherapy sequence with and without bevacizumab (BEV) added to the first line treatment. Sequence included FOLFOX [oxaliplatin + 5-fluorouracil (5FU), leucovorin] +/- BEV followed by FOLORI [irinotecan, 5-FU, leucovorin]. Treatment switched due to toxicity or disease progression. (	Data on safety and efficacy based on 2 RCTs: - RCT of BEV in combination with oxaliplatin-based chemotherapy (FOLFOX or a clinically similar regimen XELOX [oxaliplatin and capecitabine]) vs FOLFOX / XELOX as 1st-line therapy. - Sequencing RCT (GERCOR) of FOLORI - FOLFOX vs reverse sequence.	Treatment effects were assumed to be dependent on the number of previous treatments used, but independent of the type of treatments used. The same data for 2nd-line FOLORI (following 1st-line FOLFOX) used from the RCT for both sequences with and without BEV as first-line. No sensitivity analysis were conducted.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Heeg, 2015 The Netherlands	**Cancer** Multiple myeloma (MM) ineligible for stem cell transplantation (SCT).	The study aimed at developing an analytic framework for comparing overall survival (OS) of different treatment sequences in MM. MM is incurable, so main aim of consecutive treatment is to increase survival. In clinical practise, complete response is used as a short-term marker for treatment success, as it is a proven predictor of OS in MM. Compared 17 fixed chemotherapy sequences, of 4 treatment lines (modelled as 1st, 2nd, 3rd and 'further' lines). 5 different treatments (MPR+R, MP, MPT, MPV, MPVT) were used as 1st-line, one (MPR+R) was only used in one sequence, and the remaining 4 were the first treatment in 4 sequences each. Treatments considered: MP melphalan/prednisone, D dexamethasone, T thalidomide, V bortezomib, and R lenalidomide. The aim of the study was to compare 10-12 sequences relevant to clinical practice. However, only treatments for which RCTs were available were included. (*Some treatments used in clinical practice, e.g. TD and VD, were therefore excluded as they were only investigated in single arm studies*.) Only one treatment (bortezomib), for which data on efficacy of re-treatment was available, was allowed as a re-treatment within a sequence. *2nd line treatments included: RD, V; and 3rd-line: D, RD, V. Two treatments (T, R) represented continuous treatments.*	*A SR was conducted of treatments compared in RCTs of newly diagnosed patients and refractory/relapsed patients.* *Data on response rates* (complete response CR; partial response PR; and non-response NR) were obtained from network meta-analyses (NMA) [*NR = 1- CR+PR*]. *As the response outcome was multinomial, rather than binomial, the NMA was extended to an ordered logistic NMA, which means that two logistic models were combined: one of overall response (OR) versus NR and one of CR versus PR (the latter representing the contribution of CR to OR).* Separate NMAs were conducted for each treatment line. The NMA for newly diagnosed patients (1st-line) included 16 RCTs, and the comparison of 9 treatments (1 monotherapy: D; and 8 combination therapies: MPT, MP, MPR+R, MPV, MPVT, TD, RD, VPT). The NMA of relapsed/refractory patients (3rd-line) included 3 RCTs and the comparison of 3 treatments (2 monotherapies: V, D; and 1 combination therapy: RD). Due to a lack of specific 2nd-line RCTs, the response for 2nd-line treatments (*RD, V*) were based on the results of the NMA for 3rd-line (relapsed) combined with subgroup analysis reported in the 3 RCTs on the treatment effect of 2nd-line patients vs later line patients. *Time to next treatment (TTNT) and mortality (OS) data* were taken from the results of the bortezomib arm in three selected clinical trials. The three trials were selected to match the patient population, and the same data used for all treatments (bortezomib was selected for consistency). For 1st-line treatments data were obtained from a single study of newly diagnosed patients (using a Weibull model). Mortality and transition probabilities for 2nd and 3rd-line treatments were obtained from a second clinical trial of relapsed patients (using the published Kaplan Meier curves). Mortality rates for all 'later lines' were obtained from a third trial, which included patients who had had many treatments before entering the trial.	Response to a specific treatment was assumed to be independent of response to previous treatments. Treatment switching was assumed to be dependent on type of response and line of treatment, but independent of the specific treatment used. The same duration of response was assumed across different treatments. In the Markov model: - response rates (CR, PR, NR) were specific to the treatment itself and the line of treatment; - probability of switching treatment were specific to response category (not treatment) and line of treatment; - and the mortality probability was specific to response (not treatment) and line of treatment. Transition probabilities (treatment switching and mortality) for 2nd- and 3rd-line treatments were assumed to be constant over time. Median TTNT and OS were converted to monthly transition probabilities using the exponential survival curve. The impact of using TTNT instead of time to progression (TTP) was tested using scenario analysis, where it was assumed that patients on continuous treatments (T, R) switch treatment immediately after progression. *Sensitivity analysis did not incorporate sequencing effects, but external validation conducted:* the predicted median OS were compared with the reported median OS values of the 1st-line trials that were include in the NMA.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Hind, 2008 UK (NICE TA93)	**Cancer** Metastatic colorectal cancer.	Evaluated the cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for advanced colorectal cancer. Economic model implemented using data from 2 trials that planned treatment sequences. The resulting model included 7 chemotherapy sequences, 5 from the FOCUS trial (I) and 2 from GERCOR trial (II). Both trials included 2 lines of therapy and the FOCUS trial included subsequent salvage chemotherapy (SCT).	Treatment effects were derived from 2 sequencing trials. The aim of the FOCUS trial was to determine whether there was an advantage associated with the use of combination chemotherapy for 5-FU/FA plus oxaliplatin or irinotecan compared with the standard approach of sequential single agent therapy (5-FU/FA followed on progression by irinotecan), and to determine whether combination therapy was best used in 1st-line management or reserved for planned 2nd-line management following progression on 1st-line single-agent 5-FU/FA. The aim of the GERCOR trial was to determine whether 5-FU/FA in combination with irinotecan followed on progression by 5-FU/FA in combination with oxaliplatin, or the reverse sequence, was optimal.	Not applicable - restricted inclusion to treatment sequencing trials. Only the FOCUS trial included QoL assessment. Data on overall survival and PFS was available from both trials for 1st and 2nd-line treatment, but at the time PFS data for 2nd-line were only available for the GERCOR trial (II).
Lee, 2013 South Korea	**Cancer** Women with platinum-sensitive ovarian cancer (*relapsed >6 months since receiving platinum-based therapy). 'Guidelines recommend taxane+platinum as fist-line'*.	Comparison of two chemotherapy sequences starting with either polyethylene glycolated liposomal doxorubicin (PLD)+carboplatin or paclitaxel+ carboplatin. The remaining sequence was identical in both groups: topotecan or belotecan – docetaxel – etoposide – carboplatin or cisplatin.	Pivotal RCT of PLD+carboplatin vs paclitaxel+ carboplatin used as 2nd line therapy in platinum-sensitive patients (CALYPSO study). In the sensitivity analysis estimates were based on the weighted average of all relevant studies of PLD+carboplatin or paclitaxel+ carboplatin vs other comparators (*not clear if based on unadjusted arm level data*). Efficacy of the remaining 4 lines of therapy based on a SR and MA of RCTs and other experimental studies. Data for 3rd-line topotecan based on 2 studies of topotecan in patients who had failed 1st line platinum+paclitaxel; 4th -line chemo from 1 study in patients with recurrent cancer; 5th-line from 1 study in patients who were 'platinum resistant' or 'platinum-sensitive']; and 6th-line from 3 studies of carboplatin in 'platinum-sensitive', 'carboplatin- or cisplatin-sensitive', or recurrent cancer. Adverse effects data were obtained from the clinical expert group.	Matching evidence used for 2nd line treatments, but overall survival (OS) will have been affected by subsequent chemotherapies. Treatment effects of subsequent treatments assumed to be independent of positioning in sequence. The line of therapy of included trials were not reported, but the patient population indicated in the reference details included recurrent cancer or 2nd-line treatment in most cases. Two chemotherapy agents were available as 3rd and 6th line, but only one summary effect estimate reported for each line. Varying the source of clinical parameters of 2nd-line treatment had the biggest impact on results (almost doubling the ICER). One way sensitivity analysis also included changing the 'topotecan – docetaxel – etoposide – carboplatin or cisplatin' sequence to: 'docetaxel – topotecan – etoposide – carboplatin or cisplatin', and 'docetaxel – topotecan – carboplatin or cisplatin – etoposide'.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Lux, 2009 Germany	**Cancer** Hormone receptor-positive (HR+) postmenopausal women with advanced breast cancer. *Patients were assumed to have previously received tamoxifen or* aromatase inhibitor (AI) *as adjuvant anti-hormonal therapies for at least 5 years and have persistent endocrine resistance.*	Adding fulvestrant (FULV), as 2nd -line treatment, to a sequence of 4 therapies. The sequences were based on the concept of two AIs followed by 2 chemotherapy lines (polytherapy then monotherapy). Therapies included in sequence without FULV: anastrozole / letrozole – exemestane – epirubicin+cyclophosphamide or docetaxel+capecitabine –capecitabine or docetaxel.	SR of RCTs and other experimental studies. MAs conducted for each treatment given at different lines of therapy. As 1st line: anastrozole (2 studies), exemestane (3 studies), and letrozole (3 studies). As 2nd line: anastrozole (5 studies), exemestane (3 studies), and FULV (3 studies). As 3rd line: exemestane (2 studies) Monochemotherapy as 3rd/4th/5th: docetaxel (1 study), capecitabine (1 study) polychemotherapy as 3rd/4th/5th based on expert panel estimation. Estimated proportion of patients dying per line of therapy based on expert panel estimation. Adverse effects also identified from review	Treatment effect was assumed to be dependent on the number of previous treatment used, but independent of the type of treatments used. Reference details for included studies did not match the line of therapy in some cases. The same studies were referenced for exemestane used as 1st, 2nd and 3rd –line; not clear why 1st line included or how MA of 2nd line anastrozole was used. Polychemotherapy therapy was placed as 3rd/4th and monotherapy 4th/5th in the illustrated sequence. Exemestane will have been displaced to 3rd-line by adding FULV, this is not considered in the evidence. Sensitivity analyses included changing chemotherapy positioning (monotherapy followed by polychemotherapy) and altering the chemotherapy regimens used (type and member of agents used); with little effect on ICER.
NICE CG81, 2009 UK	**Cancer** Patients with advanced breast cancer who have received anthracycline therapy.	Comparing 17 predefined sequences: to identify optimal sequencing of 6 chemo drugs (up to 3 lines of therapy); four :Docetaxel (DOC), Cyclophosphemide (CPH), Capacitabine (CAP), Gemcitabine (GEM)] as 1st line (including taxanes), two CAP, Vincristine (VIN) (non taxanes) as 2nd or 3rd (only one in all three lines).	Bayesian ITC of RCTs conducted for 1st-line only. For 2nd and 3rd line - only one RCT (VIN) of mixed population (previous treatment lines) identified. Some uncontrolled trials and observational studies available for subsequent lines. [*Survival for CAP as 2nd line was high, even higher than 1st line]*	A single study selected to inform the treatment effect of each drug used as 2nd line: VIN – RCT; CAP – non-randomised trial (study selection informed by clinical guideline group or reporting appropriate outcomes). Assumed that 3rd line would be same as 2nd line (justified by fact that VIN informed by mixed population including patients receiving 3rd line treatment). Assumptions made about 'no chemotherapy treatment' used as 2nd and 3rd line. 3rd-line treatment effect reduced by one third in deterministic sensitivity analysis [scenario analyses], resulting in no change in strategies that were dominated or ranking of strategies [*CAP time without progressive disease also reduced by third; strategy with CAP still cost effective*].
NICE CG131, 2011 UK	**Cancer** Metastatic colorectal cancer	Comparing 10 predefined sequences to identify the optimal sequencing of 5 chemo regimens (up to 3 lines of therapy)	Separate Bayesian NMA conducted for 1st and 2nd line (Progression free survival (PFS) and response rates). Only prospectively sequenced studies considered for survival outcomes, but only 3 identified (one quasi-sequencing). These providing evidence for 3/10 sequences, which did not form a connected network. Sequences were therefore grouped based on the assumption that FOLFOX [oxaliplatin + 5-fluorouracil, leucovorin] and XELOX [oxaliplatin and capecitabine] are equivalent (based on a separate quasi-sequencing study), providing a connected network of 4 fixed sequences.	Not applicable - restricted inclusion to treatment sequencing trials.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Soini, 2012 Finland	**Cancer** Follicular non-Hodgkin lymphoma (FL). All patients assumed to be receiving induction treatment of rituximab with cyclophosphamide, doxorubicin, vincristine, and prednisone (RCHOP).	Comparison of 4 immunochemotherapy sequences containing rituximab and bendamustine. Patients responding to RCHOP continued with or without first-line rituximab maintenance treatment. In the case of 1st-line treatment failure patients received bendamustine or COP (cyclophosphamide, vincristine, and prednisone). RCHOP-R – RCOP-R+bendamustine RCHOP-R – RCOP-R+COP RCHOP – RCOP-R+bendamustine RCHOP – RCOP-R+COP	3 RCTs. Progression free survival (PFS) for 2nd-line bendamustine estimated indirectly according to an RCT of bendamustine+ rituximab vs RCHOP as 1st line treatment. 1st- and 2nd-line PFS for other treatments derived from the randomised maintenance phase of the PRIMA and EORTC20981 trials, respectively. PRIMA evaluated maintenance treatment in previously untreated patients responding to one of 3 non-randomised rituximab containing induction chemotherapies (74% had RCHOP); *those achieving response were randomised to rituximab maintenance or observation.* Patients were then considered ineligible for 2nd line induction with RCOP. EORTC20981 evaluated maintenance treatment in patients with relapsed or resistant FL who were randomised to induction with CHOP or RCHOP, *with those in remission randomised to rituximab maintenance or observation*. *Probability of dying in PF1 state based on age- and gender-dependent Finish background Mortality data, and for PF2 or disease progression states based on EORTC20981 trial or Finish data, which ever was highest.*	The efficacy of bendamustine based on the anchored indirect comparison of the results of a trial of 1st-line use of bendamustine and the EORTC20981 trial (2nd-line RCHOP). It was assumed that 2nd-line rituximab induction and rituximab maintenance was not affected by previous treatments for patients with long remission after first-line treatment. Due to lack of data, it was assumed that the treatment effect of 2nd-line COP was the same as CHOP. In the model, it was assumed that all patients received 1st-line RCHOP induction. The induction treatment varied in the PRIMA trial, where 74% of patients received RCHOP. *In the model patients could progress from the states: progression free survival on 1st line therapy (PF1) and 2nd line therapy (PF2), to the disease progression (DP) state. After PF2, patients were assumed to be similar in terms of sequence and survival times in the PD state.* Because of the inherent uncertainty related to bendamustine's 2nd-line efficacy, a direct comparison between the sequences without bendamustine were conducted as part of sensitivity analyses. A 5-year scenario was also conducted, where most time is spent in the PF1 state.
Wong, 2009 US	**Cancer** Patients with newly diagnosed metastatic colorectal cancer	Comparison of 9 predefined sequences of various chemotherapy and/or monoclonal antibodies used in up to three lines of therapy. 2 sequences only considered one (active) treatment line: 5-fluorouracil (5-FU), leucovorin or FOLFOX [oxaliplatin + 5-FU), leucovorin]. The remainder were: FOLIRI [irinotecan (Iri), 5-FU, leucovorin] – FOLFOX FOLFOX+ bevacizumab (Bev) – Irin FOLFOX – Iri – cetuximab (Cet) FOLFOX+Bev – Iri – Cet FOLIRI+Bev – FOLFOX – Cet FOLIRI+Bev – FOLFOX – Cet+Iri FOLFOX+Bev – Iri – Cet+Iri]	Rates and progression based on phase II and phase III clinical trials. When >1 study available, the one with the largest sample was chosen. 1st-line treatments: FOLFOX or FOLIRI informed by unpublished trials with the 'added' effect of Bev obtained from RCT of Bev+IFL vs IFL [Iri, bolus 5-FU, leucovorin]. 2nd line treatments (Iri or FOLFOX) based on 2 RCTs: Iri vs 5-FU (in patients with 5-FU failure); FOLFOX vs individual agents (after failure with FOLIRI) 3rd line treatment (Cet +/- Iri) informed by 1 RCT: Cet vs cet+Iri (in patients refractory to Iri)	Used evidence matching line of therapy and drug sequence where feasible, but not all studies for 2nd-line matched the sequences being assessed (prior treatments) and 3rd line treatments informed by RCTs of 2nd-line treatment.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
NICE CG152, 2012 UK	**Crohn's disease** *Treatment induction model included patients experiencing an acute exacerbation of Crohn's disease. The maintenance model included patients in remission, with those who relapsed entering the acute induction sequence.*	Separate analyses conducted for the introduction and maintenance of remission. Analysis of induction based on treatment sequences, whilst maintenance based on the comparison of 5 monotherapies (with induction treatment sequences embedded in the economic model) Induction of remission included the comparison of 9 pre-predefined treatment sequences made up of 7 treatment regimens: sulfasalazine (SUZ), mesalazine (MEZ), glucocorticosteroid (GCS), budesonide (BUD), azathioprine (AZA)+GCS, methotrexate (MTX)+GCS, and biologic. SUZ - GCS - AZA+GCS - biologic SUZ - GCS - MTX+GCS - biologic MEZ - GCS - AZA+GCS - biologic MEZ - GCS - MTX+GCS - biologic GCS - AZA+GCS - biologic GCS - MTX+GCS - biologic BUD - GCS - AZA+GCS - biologic BUD - GCS - AZA+GCS - biologic GCS - biologic Predefined sequences were chosen by the guideline development group (GDG), based on consensus view on clinical practice and recommendations in NICE TA187. Induction sequence used in the maintenance model was: GCS - AZA+GCS – biologic - surgery (most cost-effective sequence)	Treatment effects in the induction model were based on probabilities of remission and withdrawal. These were derived from two separate Bayesian NMA of 1st- and 2nd-line treatments (did not include biologics). The NMA of 1st-line treatments included the 4 monotherapies (and placebo), and the NMA of 2nd-line treatments included the 2 combination therapies (and glycocosteroind monotherapy). The combination therapies included the use of glucocorticosteroid in patients who had failed glucocorticosteroid. Treatment effect estimates for the maintenance model were based on relapse and withdrawal rates.	Evidence matching line of therapy available for most 1st- and 2nd-line treatments; remaining treatments assumed to be independent of position in sequence. The same treatment effect estimate was used for GCS monotherapy (taken from the NMA of 1st-line treatments) irrespective of whether it was used as the 1st- or 2nd-line treatment, or type of previous monotherapy used. The same estimate for each combination therapy was also used irrespective of whether they were modelled as 2nd- or 3rd-line treatment, or the specific previous treatment used. The same generic treatment effect was used for 'biological' treatment irrespective of whether it was modelled as 2nd, 3rd- or 4th-line. In the analysis comparing treatment sequences for the induction of remission it was assumed that once remission was successfully induced, people did not relapse. The treatment effect of glycocorticosteroid following budesonide failure was adjusted in sensitivity analysis. The GDG reasoned that it would be less effective, and appropriate to multiply the efficacy by an adjustment factor between 0-1 (probability of remission in base case = 75%, and value in sensitivity analysis = 50-100%) [The evidence tables indicated that the effect estimate in the NMA was 66%] It was assumed (based on trial data and GDG opinion) that the trials from which efficacy was based were of sufficient duration such that remission or withdrawal would occur by a certain time or not at all. This was explored in sensitivity analysis.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Greenhalgh, 2005 UK (NICE TA59)	**Depression** Patients suffering a major depressive disorder (MDD) requiring hospitalisation.	Evaluated the optimum positioning of electroconvulsive therapy (ECT). Opinion differed on whether ECT should be undertaken as a final option when all else has failed or provided higher up the treatment hierarchy. ECT was evaluated as either a 1st-, 2nd-, or 3rd-line treatment. The comparative treatments for ECT used as 1st- or 2nd-line were three classes of antidepressant drugs (Tricyclic antidepressants TCAs; Selective Serotonin Reuptake Inhibitors SSRIs; and Serotonin and Norepinephrine Reuptake Inhibitors SNRIs), and for 3rd-line (defined as treatment resistant) lithium augmentation (Lith). Following successful ECT therapy, maintenance (M) ECT can be used. The comparative maintenance / continuation therapy that were used in the model included TCA, lithium, and no therapy. The maintenance therapies used following pharmacological interventions were SSRI or no therapy. 8 predefined sequences were modelled: SNRI – SSRI – Lith – M: SSRI for all ECT – SSRI – Lith – M: ECT(for ECT), and SSRIs ECT – SSRI – Lith – M: Lith+TCA (for ECT), and SSRIs SNRI – ECT – Lith – M: Lith+TCA (for ECT),SSRI for all ECT – SSRI – Lith – M: SSRI for all SNRI – SSRI – ECT – M: Lith+TCA (for ECT), and SSRIs SNRI – ECT – Lith – M: ECT (for ECT), and SSRIs SNRI – SSRI – ECT – M: ECT (for ECT), and SSRIs	The model required two probability estimates for each treatment: i) successful treatment and ii) leaving the treatment early, due to an adverse event or not responding to treatment. The treatment success rate for ECT was taken from an RCT comparing the use of ECT treatment in patients who were defined as treatment resistant and those that were not. The failure to complete treatment rate was taken from a separate study. The successful treatment rates and failure to complete treatment (dropout) rates for the different classes of antidepressant drugs were taken from three RCTs, which were all in turn based on a published meta-analysis comparing TCAs, SSRIs and SNRIs. (RCTs undertaken within an inpatient setting, with severely depressed patients; but line of therapy or previous treatments not reported). The treatment effect of lithium augmentation (usually with an SSRI) in patients who were 'treatment-resistant' was taken from a published meta-analysis of placebo controlled RCTs. The failure to compete treatment rates were assumed to be the same as those for an SSRI intervention. Duration of treatment was based on a generic estimates of: 6 weeks for pharmacological treatments, with dropouts averaging 2 weeks of treatment; 4 weeks for ECT treatment, dropouts averaging 1 week of treatment. Relapse rates at 48 weeks for each maintenance therapy were obtained from Kaplan-Meier survival curves of selected publications	It was assumed that treatment effects were independent of previous depressive episodes and previous treatments received. It was assumed that treatment given as 2nd-line therapy had the same success rate as those given as first-line therapy. Patients requiring 3rd-line therapies were deemed to be 'treatment resistant' (and lithium augmentation assumed to be the preferred pharmacological therapy). It was assumed that each treatment's 'failure to complete treatment' rate was independent of the line of therapy. The value for lithium augmentation was assumed to be the same as that used for the SSRI intervention. Longer term treatment psychotherapy treatment was assumed to be an 8-week treatment in which patients were assumed to make a moderate improvement. No sensitivity analysis were conducted regarding sequencing effects.
McEwan, 2010 UK	**Diabetes** Newly diagnosed patients with type 2 diabetes (*initialising their 1st treatment on entering the model*).	4 predefined sequences representing (oral) therapy escalation, starting with the same monotherapy (metformin MF) followed by the addition of a new therapy (one of 3: sulphonylureas SU, thiazolidinediones TZD, or dipeptidyl peptidase DPP-4), followed by the addition of a third therapy (based on one of two triple therapies); providing up to three lines.	Monotherapy: SR for MF; Combination therapy: 2 RCTs of individual drugs used in combination with MF in patients with IR to MF. RCT of sitagliptin (a DPP-4 inhibitor) vs rosiglitazone (a TZD) and an RCT of sitagliptin (a DPP-4 inhibitor) vs glipizide (a sulfonylurea). MF+ DPP-4 informed by the latter RCT only. Triple therapies: RCT of glargine (insulin) vs rosiglitazone (a TZD) in patients with IR to SU+MF [MF+SU+TZD]; and an RCT of sitagliptin (DPP-4 inhibitor) in patients with IR to glimepiride (SU) alone or glimepiride+MF [MF+SU+DPP-4].	The predefined sequences were based on add on therapy rather than switching to a new treatment, which meant that RCTs of combination therapy and triple therapy included a patient population who had an inadequate response to previous treatments matching the model sequences. But these were based on individual treatments, which were assumed to represent the class effect

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Erhadt, 2012 Germany	**Diabetes** Patients with type 2 diabetes (*all patients in the model had received 1st-line treatment with metformin. The duration of diabetes for the model cohort was 5.4 years, matching the population from an RCT of saxaglibitin + metformin in patients inadequately controlled on metformin alone*).	The comparison of 2 drugs used as second line treatment within a predefined sequence starting with the metformin (MF) monotherapy followed by the addition of a new drug (saxaglibitin vs sulfonylurea) as combination therapy, followed by the same 3rd line treatment (metformin plus insulin). Sequences matched published German guidelines.	Efficacy and safety data obtained from: Published SR for MF monotherapy (1st-line); a RCT of saxaglibitin+MF vs sulfonylurea+MF in patients with IR to MF; and a published MA comparing different types of insulin (line of therapy not stated). Baseline Characteristics of model cohort based on patients in the saxaglibitin vs sulfonylurea RCT. Prespecified insulin (HbA1c) threshold value used to invoke treatment switching to second and third line, which was varied in the scenario analysis.	Evidence matching sequence available for 1st and 2nd line; but not considered for 3rd line. Patient population of interest were those receiving 2nd line treatment; using a base case analyses where all patients received 1st-line MF. One way sensitivity analyses included line of therapy (1st vs 2nd). The scenario where patients entering the model received combination therapy as their 1st treatment, switching to MF+insulin on reaching relevant threshold was a key driver of results; ICER for MF+saxaglibitin compared to MF+sulfonylurea were lower.
Connock, 2006 UK (NICE TA79)	**Epilepsy** Newly diagnosed children with partial epilepsy.	Adding one of 3 'newer' anti-epileptic drugs (AEDs) to a sequences of older drugs; 2 used as the first-choice add-on therapy (after two older 'monotherapy' failure) and one used as 1st-line monotherapy or add-on therapy providing 7 predefined sequences, with up to 4 lines if therapy.	SR identified placebo controlled RCTs of each new AEDs used in patients with refractory epilepsy, and as 1st-line monotherapy for one AED; effect estimates for this latter drug used at as anchor points and to develop reduction factor.	The RCT data of AED as add-on therapy assumed to be reasonably representative of 4th-line therapy Treatment effect of newer AEDs reduced by a set amount for different lines of therapy based on a reduction factor of 0.4.
Knoester, 2007 Netherlands	Epilepsy Newly diagnosed patients with epilepsy.	Six predefined sequences of 3 antiepileptic drugs used in up to 2 lines of therapy.	3 RCTs of 1st line treatment and 3 controlled trials of 2nd line treatment (1 only used for adverse effects data, 2 included patients refractory to a relevant 1st line drug). *No data were found on probabilities for 2nd-line VPA or CBZ after failure of LTG.*	Used evidence matching line of therapy and drug sequence where feasible. Where matching evidence not available for 2nd line, treatment effects assumed comparable to an alternative treatment used in relevant position in sequence.

NICE CG137, 2012	**Epilepsy** Newly diagnosed patients with focal epilepsy	Separate economic evolutions were conducted for the comparison of antiepileptic drugs AEDs as monotherapy in newly diagnosed patients; and the comparison of AEDs as adjuvant therapies in refractory patients. The same model structure was used for both, based on 1st-line monotherapy (for newly diagnosed patients), 2nd-line monotherapy (for refractory patients), and 3rd-line adjuvant therapy. The comparison of AEDs for newly diagnosed patients included: carbamazepine (CBZ), gabapentin (GBP), lamotrigine (LTG), levetiracetam (LEV), oxcarbazepine (OXC), sodium valproate (VPA) and topiramate (TPM). The same 2nd-line treatment assumed across all comparators: CBZ (and LTG in sensitivity analysis). A common adjucat therapy also used: tiagabine (TGB) The comparison of AEDs as adjuvant therapies for refractory patients included: eslicarbazepine acetate (ESL), gabapentin (GBP), lacosamide (LAC), lamotrigine (LTG), levetiracetam (LEV), oxcarbazepine (OXC), pregabalin (PGB), tiagabine (TGB), topiramate (TPM), vigabatrin (VGB) and zonisamide (ZON). These were used as adjunct to sodium valproate and carbamazepine.	*Initial treatment response for monotherapies* (treatment failure and remission) were based on a network meta-analysis of individual patient level data from RCTs (including SAND trial) with up to 3-yrs follow-up, Data on the time to treatment failure, broken down by cause, for 1st-line monotherapies were taken from the SAND trial (reported as 100-day intervals). The data showed that withdrawal for intolerable side effects is largely limited to the early post-randomisation period, whereas treatment failure due to lack of effect seems to occur later, once a well-tolerated dose has been reached. *Initial response to 2nd monotherapy* An observational study of newly diagnosed patients (Kwan and Brodie, 2000) showed that response to initial drug treatment was a powerful prognostic factor, with non-responders more likely to have more refractory epilepsy and less likely to achieve seizure freedom with any AED in the future. If, the failure of 1st-line monotherapy was due to lack of efficacy, then the likelihood of achieving seizure freedom with a second-line monotherapy was much lower (11% vs 47%). However, if failure of treatment was due to intolerable side effects of some type of idiosyncratic reaction, probability of response to second-line monotherapy treatment was similar to that in treatment naïve patients (45% vs 47%). The data from Kwan and Brodie was not directly usable in the model, but used to inform treatment sequencing assumptions. *The probability of treatment failure subsequent to achieving 12-mth remission* were taken from Wilby, 2005 (NICE TA76), who interpolated the data from a published graph of an observational study (long term follow-up data from the National General Practice Study of Epilepsy) for monotherapy. This data was not specific to the drug under consideration, but indicated that the probability of failure, and thus discontinuation, declined over time for patients who successfully completed the first cycle on any given therapy. *Initial treatment response to adjuvant therapy* (seizure freedom [100% reduction], partial seizure control [50-99% reduction] without withdrawal, and treatment failures due to adverse effects or lack of seizure control) for adjuvant therapy were based on a meta-analysis of placebo-controlled trials with up to 6 mths follow-up. *Continued response to adjunct therapy* The probability of treatment failure subsequent to the first model cycle were interpolated from a published graph	Patients failing 1st-line due to inadequate seizure control were assumed to be 75% less likely (risk ratio 0.25) to achieve remission with 2nd line monotherapy (informed by observational study). One way sensitivity analysis were conducted varying this figure to 0.5, 0.75, and 1. For patients who failed 1st-line due to intolerable side effects it was assumed that response to the second line monotherapy was independent of response to 1st line AED. The probability of changing treatment after achieving remission were assumed to be the same across all treatments. Patients who have not achieved remission, but also not classified as having failed treatment, were assumed to persist with 1st line monotherapy for 2 years, at which point the patient was classified as having failed due to inadequate seizure control and moved on to 2nd line treatment. Sensitivity analysis were conducted using 18 and 36 months trial period before switching. For the PSA, uncertainty in the probabilities of treatment discontinuation was accounted for by using beta distributions.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
			presented by Wilby, 2005 (NICE TA76), of the data from an open-label follow-up study of tiagabine. This data indicated that the probability of failure, and thus discontinuation, declined over time for patients who successfully completed the first cycle on any given therapy.	
Wilby, 2005 UK (NICE TA76)	**Epilepsy** Newly diagnosed patients with epilepsy. Separate analyses conducted for generalized seizures and partial seizures.	Included fixed treatment sequences for the comparison of various antiepileptic drugs (AEDs) used as monotherapy for newly diagnosed patients (1st-line), monotherapy for refractory patients (2nd-line), and combination therapy (3rd-line). For the evaluation of generalised seizures 3 older AEDs (CBZ, VPA, PHT) and 1 newer AED (LTG) were considered as monotherapy, and 2 newer AEDs as adjuvant therapy (LTG, TPM). For partial seizures 5 AEDs were evaluated for newly diagnosed patients (CBZ, VPA, LTG, OXO, TPM); 3 for refractory patients (CBZ, VPA, LTG); and 7 as adjuvant therapy (PLA, LTG, TPM, GBP, LEV, OXC, TGB). For generalised seizures, 5 AEDs were evaluated for newly diagnosed patients (PLA, GBP, TGB, LEV, TPM); and 2 as adjuvant therapy (PLA, TPM). Monotherapy for refractory patients was not considered as no clinical trial data were available for this setting. For the evaluation of monotherapy for newly diagnosed patients with partial seizures, CBZ and adjuvant GBP was used as the 2nd- and 3rd- line treatment, respectively; with other treatments considered in sensitivity analysis. For the evaluation of monotherapy of newly diagnosed patients with generalised seizures, TPM was assumed to the adjuvant treatment.	Bayesian meta-analysis of RCTs (with typical FU of 6mths) was conducted for each of the three indications: monotherapy for newly diagnosed patients, monotherapy for refractory patients (2nd-line), and combination therapy. Observational data showed that rate of withdrawal varied with time, with patients becoming less likely to withdraw from drug as time progressed. The withdrawal rates used in model (probability of remaining on treatment after the first 6-mth cycle) were based on observational data, which was not specific to the drug under consideration. This included a longitudinal cohort study (the National General Practice Study of Epilepsy) for monotherapy, and an open label trial of tiagabine for combination therapy. The uncertainty in the probability of treatment failure (for the PSA) were incorporated using beta distribution parameterised using the observed data.	Treatment effects assumed to be independent of position in sequence. The probability of changing treatment after achieving remission were assumed to be the same across all treatments.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Beard, 2011 US	**Fibromyalgia** Patients with fibromyalgia for at least 3 months and eligible for pharmacotherapy.	Adding duloxetine (DUL) to a standard treatment sequence of 5 treatments. DUL added as 1st, 2nd, 3rd, 4th, 5th, and 6th line treatment. Base-case sequence: TCA-SNRI-ANTI-TRAM-PRAM	SR and ITC of placebo RCTs, with data for each treatment based on: DUL: 3 RCTs; TCA (triyclic antidepressant): 1 RCT of amitriptyline; SNRI (serotonin-norepinepherine reuptake inhibitor): 3 RCTs of milnacipran; ANTI (anticonvulsant): 3 RCTs of pregabalin; TRAM (tramadol): 1 RCT; and PRAM (pramipexole): 1 RCT. None of the RCTs reported data for patient subgroups explicitly defined by prior treatment failure. *Model allowed for a proportion of patients to drop-out of current treatment, which would then be lost to subsequent treatments. An annualised rate developed based on 2 RCT extensions for DUL. The same percentage (25%) was used for all active treatments, and explored across a 20-30% in sensitivity analyses.*	The model assumed that response rates were independent of placement in the treatment sequence. Assumption made that all active treatments had the same expected level of long term adherence. Scenario analyses showed the results were sensitive to the proportion of long term drop out patients who were assumed to be lost to any subsequent treatment. It was assumed that pain control was maintained whilst on treatment. *The model and treatment responses were based on individual drugs which were assumed to represent the class effect.*
Denis, 2008 France	**Glaucoma** Newly diagnosed patients with glaucoma or ocular hypertension.	Comparison of two predefined sequences of intraocular pressure lowering agents, starting with either latanoprost or travoprost, as monotherapy, followed by the addition of timolol (combination therapy). The model was used to estimate the probability of starting 3rd-line treatment	2 RCTs: Travoprost vs latanoprost (patients were not newly diagnosed); and latanoprost+timolol vs travoprost+timolol (patients had not experienced first line prostaglandin failure). (*Patients underwent washout treatment prior to enrolment in both trials*).	Treatment effects assumed to be independent of position in sequence.
Orme, 2012 UK	**Glaucoma** Newly diagnosed patients with glaucoma or ocular hypertension.	Comparison of predefined sequences starting with one of 3 topical hypotensives (prostaglandin analogues) used as monotherapy: latanoprost, travoprost or bimatoprost followed by the progressive addition of timolol (combination therapy), then dorzolamide (triple therapy). If 1st drug discontinued due to poor tolerance, then timolol monotherapy followed by the addition of dorzolamide used. *Sequences included 11 different treatments, three 1st-line, and four 2nd-line and 3rd-line treatments (4th if poor tolerance).*	Treatment switching was based on 3 triggers: intolerance, none controlled intraocular pressure (IOP); and glaucoma progression. Data on the first 2 triggers based on Bayesian NMA of RCTs of prostaglandin analogues. The NMA for trigger 1 included 9 different active treatments (72 RCTs), 8 of which were included in the sequences being investigated (3rd-line combination therapies were missing). The NMA for trigger 2 included 8 treatments (18 RCTs), 6 of which were included in the sequences. Data modelled for triggers 1 and 2 for treatments not available in the NMA were based on matching the nearest drug in class or combination. Meta-regression (73 RCTs) also used to predict on-treatment IOP over time. Data for trigger 3 was based on 3 long term trials of glaucoma and ocular hypertension	Sequencing effects appear to have been ignored with data on individual treatment effects taken from the same NMA comparing all treatments simultaneously. Treatment effects for some drugs were based on the nearest drug in class or combination. In the base case analysis data for intolerance was based on the incidence of hyperaemia, as a proxy). In sensitivity analysis data on intolerance was obtained from expert opinion.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Holmes, 2006 South Africa	**HIV** Women with HIV who have been exposed to nevirapine (NVP). Single dose NVP is used to prevent mother-to-child transmission of HIV, but can lead to NVP resistance in the mothers.	Comparison of 5 antiretroviral therapy (ART) strategies, 2 of which included sequencing: HIV care without ART; NVP based ART; lopinavir-ritonavir (LPV/r) based ART; NVP based ART followed by LPV/r based ART; LPV/r based ART followed by NVP based ART. Both NVP and LPV/r base regimens also included two nucleoside reverse transcriptase inhibitors (NRTIs), making up the triple therapy.	Efficacy and adherence based on 2 studies. NVP based regimen: observational cohort study - follow-up data for women who received a NVP postpartum after participating in the Perinatal HIV Prevention Trial 2 (PHPT-2) [*placebo RCT of adding single-dose perinatal NVP plus standard zidovudine for preventing mother-to-child transmission of HIV*]. LPV/r based regimen: RCT of LPV/r versus nelfinavir for initial treatment of HIV. (*Efficacy of NVP based ART were analysed according to two subgroups, with and without NPV resistance*) Mortality based on published studies of HIV cohorts; adverse effects not considered.	Efficacy data based on 1st line treatment with decrement in efficacy assumed for 2nd-line: 2nd-line regimen efficacy was estimated to be 90% of its efficacy as an initial therapy, because of NRTIs resistance resulting from 1st-line treatment failure. The reference given for this was a published cost-effectiveness analysis of resistance testing for treatment-naïve HIV-infected patients by the same authors. The NVP and LPV/r base regimens included differing NRTIs, but their efficacy were assumed to be equal. Sensitivity analyses did not consider structural uncertainty.
Tebas, 2001 US	**HIV** HIV-infected patients in which therapy is started immediately vs progressively at a rate of 5, 10. 15, 20 or 30% of the original population each year. *Population model based on virologic, rather than clinical, outcomes.*	Timing of initiating antiviral therapy evaluated using a sequence of up to three possible regimens (before the development of multidrug-resistant virus).	Major antiviral trials used for data on probability and duration of response on 1st (10 and 3 trials, receptively), 2nd (2 and 1 trial), and 3rd regimen (no trials). Where insufficient evidence, 'best estimate' based on consensus determination by authors used.	Response to 3rd regimen and long-term durability of response to the 2nd regimen were based on 'best estimate' according to consensus determination by the authors. One-way and two-way sensitivity analysis only considered response to first regimen only.
Frankum, 2005 US	**Onychomycosis** Patients seeking treatment for onychomycosis.	Optimal sequencing of 4 treatments used in up to three lines of treatments. 12 sequences presented in ascending order according to their average cost per responder. Sensitivity analyses conducted on top 4 sequences	Safety and efficacy based on MA of clinical trials (uncontrolled and RCTs) of each treatment. Line of therapy not specified, and effects based on unadjusted arm level data. One treatment was only informed by non-comparative or non-RCTs. All study designs were treated on an equal bases.	Sequencing effects were not considered. It was assumed that uncontrolled clinical trials (and open-label RCTs) provide an un-biased estimate of treatment effect. *Recurrence initially treated with the same initial successful treatment.* No sensitivity analysis conducted relating to simplifying assumption for sequencing effects. The results were most affected by response rates for 2 drugs used 1st line that were informed mainly by non-RCTs.
Smith, 2007 US	**Postheretic neuralgia** 70 year old patients with established postherpetic neuralgia (PHN). Side-effects of tricyclics can be life-threatening in patients with coronary artery disease (CAD). Two scenarios were considered, patients with and without CAD, to account for differences in tricyclic use between these groups.	Comparison of treatment sequences, for established PHN in 70 year old patients, based on 6 drugs (tricyclic, gabapentin, pregablin, tramadol, opioid, topical lidocaine patches). Treatment switches occurs due to inadequate pain relief or intolerance. A set of procedures were used to produce a manageable number of sequences (n=25). The listed sequences included oxycodone as the opioid, but the tricyclic was not specified.	Main data source was a published SR (with MAs) of RCTs, with values from other reviews and more recent studies examined in sensitivity analysis. Results quantified as numbers needed to treat (pain relief) and numbers needed to harm. Median duration of PHN (90 days) was based on natural history of PHN. (*Beneficial effect and adverse effects of medication were assumed to occur during first month; with treatment switching, if necessary, occurring in next monthly cycle. Once beneficial medication found, it was assumed to have sustained benefit and continued until PHN resolved or death.*) Death rates based on US mortality tables.	Failure to respond to one medication was assumed to have no effect on the likelihood of response to other medications. This was not explored in sensitivity analyses. PSA were performed to estimate the likelihood that each medication were favoured early in treatment sequence (as 1st, 2nd or 3rd drug) under different cost-effectiveness thresholds. It was assumed that therapy reduced PHN symptoms but did not decrease PHN duration.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Anis, 2011 US	**Psoriasis** Patients with moderate-to-severe psoriasis.	Optimal sequencing of systemic treatments using 5 biologics, including 3 Tumour necrosis factor (TNF)-inhibitors, (up to 6 lines) identified by ranking individual treatments in order of their cost-effectiveness (benefit and cost) compared with no treatment (placebo). *The SR* (*Bansback, 2009*) *considered any treatment for patients with moderate-to-severe psoriasis who had an inadequate response to topical treatments alone and had received (or candidates) for systematic therapy or phototherapy. The NMA included 5 biologics and 2 conventional Disease-modifying anti-rheumatic drugs (Adalimumab, Etanercept, Infliximab, alefacept, efalizumab, Methotrexate, and Cyclosporin). It did not include other forms of therapy, such as retinoids, phototherapy and combination therapy, as they did not have a link to placebo that would allow them to be included in the treatment network.*	Same as Sizto, 2009 Short-term efficacy (vs placebo) taken from Bayesian NMA of 22 RCTs (*Bansback, 2009*), based on relative probabilities of archiving Psoriasis Area and Severity Index (PASI) response (50/75/90). (*Each level of PSAI response related to a different change in health utility*). Long-term efficacy based on published evidence and assumptions. Assumed that responders remained on therapy, maintaining the response rate achieved at end of trial period. *A separate model was run for each treatment, which considered the initial 'trial period' (to see if treatment works) and subsequent 'treatment period' (in patients in whom the treatment does work) separately.*	Treatment effects assumed to be independent of positioning in sequence. Assumed that treatment does not alter disease progression. Treatments only provide benefit whilst they are being administered. (*Used same modelling approach as Woolacott, 2006*)
Sawyer, 2013 UK (NICE CG153)	**Psoriasis** Patients with psoriasis for whom topical therapy is expected to be practical, effective and safe in the long term. Separate analyses were conducted for psoriasis of trunk and limb, and scalp	Initial topical therapy followed by referral to specialist care for more intensive treatments. Initial therapy included up to 3 lines, made up of 8 different topical therapies, which could be used as 1st-, 2nd- or 3rd- line, and a further 3 available as 3rd-line. A number of restrictions were used on some resulting sequencing to ensure safe and logical use of treatments; 118 sequences were evaluated for trunk and limb psoriasis, and 169 for psoriasis of the scalp.	Bayesian NMA of RCTs for probability of response (at up to 4 weeks) for each treatment included 39 RCTs for trunk/ limb, and 13 RCTs for scalp. Data within a drug class were pooled. Probability of response was broken down into early (4 weeks) and late responders (8 weeks) based on clinical opinion and data from 2 RCTs for trunk/limbs and 3 RCTs for scalp. Risk of relapse taken from 3 RCTs that reported average relapse rates at 4 weeks. Assumed that relapse could occur at any time after response.	Treatment effects assumed to be independent of position in sequence. Treatment effects taken from NMAs that compared all treatments simultaneously, including those that were only considered in the sequences as 3rd-line. As the evidence to support treatment-dependent relapse rates was limited, the same relapse rate was assumed for all treatments. Also assumed that relapse could occur at any time after response. Scenario sensitivity analyses conducted, excluding sequences that included consecutive use of corticosteroids.
Sizto, 2009 Canada [*see also Anis, 2011*]	**Psoriasis** Patients with moderate-to-severe psoriasis. The accompanying SR of effectiveness was published separately (*Bansback, 2009*) and also used by Anis, 2011 (*see above for details*).	Optimal sequencing of 6 Disease-modifying anti-rheumatic drugs (4 biologics and 2 non-biologics; identified by ranking each treatment in order of their cost-effectiveness (benefit and cost) compared with no treatment (placebo). *Evidence for other forms of therapy, such as retinoids, phototherapy, and combination therapy, was not included in the analysis because the other forms did not have a link to placebo that would allow them to be incorporated in the evidence synthesis.*	*Same model as Anis, 2011.* Short-term efficacy (vs placebo) taken from Bayesian NMA of 22 RCTs (*Bansback, 2009*), based on relative probabilities of archiving Psoriasis Area and Severity Index (PASI) response (50/75/90). (*Each level of PSAI response related to a different change in health utility*). Long-term efficacy based on published evidence and assumptions. Assumed that responders remained on therapy, maintaining the response rate achieved at end of trial period. *A separate model was run for each treatment, which considered the initial 'trial period' (to see if treatment works) and subsequent 'treatment period' (in patients in whom the treatment does work) separately.*	Treatment effects assumed to be independent of positioning in sequence. Assumed that treatment does not alter disease progression. Treatments only provide benefit whilst they are being administered. (*Used same modelling approach as Woolacott, 2006*)

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Woolacott 2006 UK (NICE TA103)	**Psoriasis** Patients with moderate to severe psoriasis. *Hypothetical cohort entering model was not specifically described, but the inclusion criteria for the SR noted that theses patients are usually defined as having an inadequate response to topical treatments alone and have either received prior systemic therapy or phototherapy or are candidates for such therapy.*	Optimal sequencing identified by ranking treatments in order of their cost-effectiveness (net-benefit per unit time) compared with supportive care. Primary analysis included 2 biological agents, one was a tumour necrosis factor (TNF)-inhibitor (Etanercept) administered as three different strategies. Secondary analysis included a further 4 agents, 1 TNF-inhibitor (Infliximab) and 3 conventional Disease-modifying anti-rheumatic drugs. *7 further systemic treatments included in the SR of clinical effectiveness were not included the NMA (or economic evaluation) as they did not link into the network of evidence.*	Initial treatment response (PASI response 50/75/90) taken from a Bayesian NMA of 16 RCTs. Treatment duration based on observational studies. *The Markov model consider the initial 'trial' period (the interval during which a new treatment is used to see whether it works or not) and the subsequent 'treatment' duration (for responders) separately. The 'trial' period for each treatment was based on the period over which treatment response was assessed in the efficacy trials for each treatment option and 'expert opinion'. The mean duration of response was estimated based on an assumed annual dropout rate for responding patients and a maximum assumed treatment period based on published guidelines if appropriate. The impact of patient attrition rate was explored in sensitivity scenario analysis.*	Treatment effects assumed to be independent of positioning in sequence. Assumed that treatment does not alter disease progression. Treatments only provide benefit whilst they are being administered. The estimated 'trial' and 'treatment' periods were entered into the Markov model as fixed values. The probability of treatment failure was assumed to be constant (allowing the model to be implanted as a Markov chain).
Davies, 2008 UK	**Schizophrenia** Patients with stable schizophrenia. The model accounted for the relapsing nature of schizophrenia and the differing tolerability profiles of atypical antipsychotics (*whilst on treatment patients could relapse, discontinue or experience adverse effects*).	12 alternative treatment sequences containing 3 atypical antipsychotics. The first 2 successive atypicals were selected from the following 4: aripiprazole (ARI), olanzapine (OLZ), quetiapine (QTP), and risperidone (RSP). The third atypical was clozapine (CLZ).	Clinical data based predominantly on the US Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) study: a pragmatic trial designed in 3 phases incorporating treatment switching with re-randomisation to next treatment. The CATIE Phase 1 RCT provided data of 1st-line a-typical antipsychotics OLZ, RSP, and QTP [vs typical antipsychotic, perphenazine]; Phase 2 included re-randomisation of patients who had failed to respond to their first a-typical antipsychotic to CLZ, OLZ, QTP or RSP. Data for ARI based on an RCT of ARI vs OLZ (this was an optional extension of a previous RCT of ARI vs placebo in stable schizophrenia; patients who had completed initial treatment or relapsed entered into extension study). Study did not report relapse rates.	Data for CLZ based on 2nd line use (prior antipsychotic not specified). Discontinuation rates for OLZ, QTP, RSP based on 1st and 2nd-line use of antipsychotics, but did not take into account the actual sequence used (prior antipsychotic not specified for second line use). For ARI, same effect estimate used for 1st and 2nd line. Data on relapse and adverse effects based on 1st -line use only for all antipsychotics, and relapse for ARI assumed to be same as RSP. Deterministic sensitivity analyses included substituting second-line discontinuation rates with 1st-line discontinuation, which had negligible effect on results.
Heeg, 2008 Netherlands	**Schizophrenia** Patients with chronic schizophrenia requiring antipsychotic treatment.	The comparison of atypical versus conventional antipsychotics as first-line treatment. Treatment sequences were developed as part of the model. First-line treatment included the selection of one of 6 atypical or one of 6 conventional antipsychotics. The subsequent three treatments were the same in both sequences being compared (a-typical vs conventional), and could include one of 12 conventional atypicals for the 2nd and 3rd-line treatments, and clozapine for the 4th-line treatment. The choice of specific drug used as 1st, 2nd, and 3rd-line treatment were selected based on UK market share data, in order to approximate treatment selection in clinical practice.	The atypicals were evaluated as a class. This was necessary as relevant comparator studies between atypicals were not available. Positive and Negative symptom score (PANNS) during and between relapses for connectional drugs, and a-typical drugs were obtained form 3 published systematic reviews (2 investigating risperidone). The incidence of side effects were obtained from relevant Cochrane reviews of various antipsychotics.	Treatment independence. Individual drugs within the two groups 'atypcial' and 'conventional antipsychotic' were assumed to have identical treatment effect.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Fitzsimmons, 2014 UK	**Sciatica** Patients presenting to the GP with sciatica.	Three treatment pathways were compared with 'inactive control', representing: primary care management (initial treatments only, n=5); stepped approach (initial treatments n=5, followed by intermediate n=6, and then invasive treatments (epidural or disk surgery); and immediate referral to surgery following initial treatment in primary care.	NMA of RCTs and observational studies, providing average effect for 18 different treatment types. More invasive treatments tended to be used after failure of conservative treatments. Intensive treatments/surgery tended to be evaluated in RCTs as last line – this effect estimate used in model for 'surgery' early and last line	Treatments were applied in succession until successful resolution of symptoms achieved. Assumptions made that the treatment effects were additive and that the effect of individual treatments were independent of position in sequence (derived from the same NMAs). Sensitivity analysis included applying a reduction factor to treatments used later therapies, and allowing for the subsequent effect of a proportion of patients being non-responsive at each stage (treatment resistant subgroup).
Bensmail, 2009 France	**Spasticity** Patients with disabling spasticity.	Compared the use of intrathecal baclofen (ITB) as the 1st-line strategy with current treatment options. Two separate models were developed. The model representing current treatment pattern started with physical treatment only, followed by oral treatment then one of three strategies: i) neurosurgery – nursing; ii) ITB; or iii) focal treatment – neurosurgery. The final option in the strategy was nursing. The ITB as first-line strategy started with ITB followed by one of three strategies: i) oral treatments – neurosurgery - focal treatment or nursing; ii) neurosurgery – nursing; iii) nursing – neurosurgery; or iv) focal treatment - nursing	Due to the paucity of clinical trials of ITB vs alternative strategies, only descriptive data were used as parameter estimates. Transition probabilities were based on an analysis of a retrospective survey of patient databases performed in a single hospital specialising in rehabilitation activity. These were expressed as a range of values validated by expert opinion (to track extreme values). The data were not presented in the publication.	Sequencing effects were not considered.

Abbreviations: ICER incremental cost effectiveness ratio; ITC indirect treatment comparison; MA meta-analysis; mths, months; NMA network meta-analysis; PSA probabilistic sensitivity analysis; RCT randomised controlled trial; SR systematic review.

Table D1.2: Summary of included rheumatology modelling studies (*Studies ordered alphabetically by condition and then author*)

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Albert, 2000 US	**Rheumatoid arthritis (RA)** Patients with early RA and Disease-modifying anti-rheumatic drugs (DMARD) naïve.	Starting with least toxic (pyramid strategy) vs most effective drug first (inverted pyramid). *Four different sequences of three conventional DMARDs: methotrexate (MTX), hydroxychloroquine (HCQ), gold.*	Compared 3 sources: expert rheumatologists, published MA of second-line agents, and a survey of American Rheumatologists.	Treatment effects assumed to be independent of position in sequence.
Bansback 2005 Sweden	**RA** Patients with an inadequate response to at least 2 conventional DMARDs	Comparison of TNF-inhibitors [Adalimumab, Etanercept, Infliximab] (or no biologics) in patients who had failed at least two conventional DMARD. Biologics added to a fixed sequence of 3 conventional DMARDs starting with leflunomide (LEF) and the latter considered as palliative care.	TNF-inhibitors: ITC of placebo RCTs (n=5) and observational study (for treatment withdrawal). LEF: Swedish observational study (registry data). Patients treated with LEF had failed on average 4 previous conventional DMARDs.	Used evidence matching line of therapy for initial TNF-inhibitors. Treatment effects assumed to be independent of position in sequence for remaining treatments. Same effect estimate used for conventional DMARDs. *This did not account for previous TNF-inhibitors, and based on population who had received a greater number of previous DMARDs than in TNF trials.*
Barton 2004 UK (NICE TA36)	**RA** Patients with early RA and DMARD naive	A predefined sequence of 9 conventional DMARDs with and without a TNF-inhibitor (Etanercept (ETA) vs Infliximab (INF)) added as the 3rd and last drug. *All Sequences included the same initial 2 conventional DMARDs (*sulfasalazine (SSZ) *and* methotrexate (MTX)*).*	TNF-inhibitors: placebo RCT for ETA (in patients with IR to previous DMARDs) and personal communication for INF. Initial DMARDs (*SSZ and MTX*): RCT of LEF vs SSZ vs placebo in SSZ naive patients; and RCT of LEF vs MTX vs placebo in MTX naïve patients). Subsequent DMARDs: Same RCTs as above used for LEF. Substitute data from SSZ used for remaining 6 DMARDs. *LEF and remaining 6 DMARDs used in model after both SSZ and MTX.*	Treatment effect of TNF-inhibitor used as 3rd drug assumed to be dependent on number of previous treatments but independent of type of treatments. Assumed that treatment effect of INF is same as ETA. Data for 2nd conventional DMARD based on RCT of 1st time use. Treatment effects for TNF as last drug and conventional DMARDs after the first 2 assumed to be independent of position in sequence. Same effect estimate used for both TNFs irrespective of point in sequence. Substitution effect used for all conventional DMARDs except LEF (assuming exchangeability of effects). LEF based on use in early RA.
Beresniak, 2011 Spain	**RA** Patients with inadequate response to one TNF-inhibitor	4 predefined sequences of 3 biologics, used for comparing 2 different non-TNF-inhibitors as 2nd-line and a non-TNF inhibitor vs a TNF-inhibitor as 3rd-line (after inadequate response to 2 TNF-inhibitors). All sequences started with the same TNF inhibitor. *etanercept (ETA)-abatacept (ABA)-adalimubab (ADA)* *ETA- rituximab (RTX)-ADA* *ETA-ADA-ABA* *ETA-ADA- infliximab (INF)*	ABA and RTX (non-TNF inhibitors) as 2nd-line: Placebo RCT plus extension study for each drug. ABA as 3rd-line: subgroup analysis from the RCT, of patients with inadequate response to 2 TNF-inhibitors. INF (TNF-inhibitor) as 3rd-line: uncontrolled trial [ReAct] of an alternative TNF (ADA) in patients who had previously received INF +/or ETA. Same study also used to inform 2nd and 3rd line use of ADA, which was the remaining drug making up all 4 sequences.	Used evidence matching line of therapy from RCTs and uncontrolled trial for 2nd-line biologics and subgroup data and uncontrolled trial for 3rd-line. Comparable effects assumed for TNF-inhibitors used as 3rd-line. Assumed 100% inadequate response to 1st-TNF-inhibitor in the model for all 4 sequences.
Beresniak, 2013 Germany	**RA** Patients with inadequate response to one TNF-inhibitor	4 predefined sequences of 3 biologics, used for comparing 2 different non-TNF-inhibitors as 2nd-line and a non-TNF inhibitor vs a TNF-inhibitor as 3rd-line (after inadequate response to 2 TNF-inhibitors). All sequences started with the same TNF inhibitor. *ADA-ABA-ETA* *ADA-RTX-ETA* *ADA-ETA-ABA* *ADA-ETA-INF*	ABA and RTX (non-TNF inhibitors) as 2nd-line: Placebo RCT plus extension study for each drug. ABA as 3rd-line: subgroup analysis from the RCT, of patients with inadequate response to 2 TNF-inhibitors. INF (TNF-inhibitor) as 3rd-line: uncontrolled trial [ReAct] of an alternative TNF (ADA) in patients who had previously received INF +/or ETA. Same study also used as a substitute effect for remaining drug making up all 4 sequences - 2nd and 3rd-line ETA.	Used evidence matching line of therapy from RCTs for 2nd-line non-TNFs and substitute data from uncontrolled trial for 2nd-line TNF. Subgroup data and substitute data from uncontrolled trial used for 3rd-line. Comparable effects assumed for TNF-inhibitors used as 2nd or 3rd-line, but the prior TNFs failed in uncontrolled study included the TNF being modelled. Assumed 100% inadequate response to 1st-TNF-inhibitor in the model for all 4 sequences.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Brennan, 2004 UK (Industry submission for NICE TA36)	**RA** Patients who have failed two conventional DMARDs (Methotrexate (MTX) and Sulfasalazine (SSZ))	A sequence of 3 conventional DMARDs (cDMARDs) with and without the TNF-inhibitor, etanercept (ETA) (added to the start of the sequence). [(ETA) – gold (A) – leflunomide (LEF) (B) – cyclosporin (CyC)+MTX (C)]	Placebo RCT of ETA (in patients who had inadequate response to cDMARDs). Data for DMARDs derived from published studies selected based on duration of disease matching ETA trial (LEF: Swedish observational study; CyC+MTX: placebo RCT). Comparable data not available for Gold; initial response taken from meta-analysis of numerous DMARDs of patients with disease duration >10 years. Progression for patients receiving 3rd and 4th line conventional DMARD scarce (A-B). Proxy used for long term disease progression, which did not account for previous failing biologic due to lack of data.	Used evidence matching line of therapy for first-line biologic. Subsequent treatment effects assumed to be independent of position in sequence. Different response rates used for each DMARD, but did not account for positioning in sequence or previous biologic failure. Generic effect estimate for cDMARDs used for gold. *Scenario analyses included:* *Changing the order of DMARDs (B-C-A); and Increasing number of DMARDs in the sequence to 6 (A-B repeated). But response rates were not altered according to change in sequencing position.* Progression of disease assumed over time; greater for ETA than cDMARD.
Brennan, 2007 UK (BSRBR submission for NICE TA30)	**RA** Patients with an inadequate response to at least 2 conventional DMARDs	TNF-inhibitors (as a class) vs conventional DMARDs (cDMARDs) in patients with inadequate response to at least 2 conventional DMARDs. Sequential TNF-inhibitors also considered in subgroup analyses.	Individual patient level data from register (BSRBR); regression analysis adjusted for no. of previous cDMARDs. No statistically significant correlation was identified between 1st and 2nd TNF inhibitors. Generalised weighted average effect estimate used for both TNF-inhibitors and cDMARD, irrespective of specific drug used.	Used evidence matching line of therapy for initial TNF-inhibitor. Given the absence of correlation, response to 2nd TNF-inhibitor was assumed to be independent of response to 1st.
Chen, 2006 UK (NICE TA130)	**RA** Patients with early RA and DMARD naive	A predefined sequence of conventional DMARDs with and without 1-2 consecutive TNF antagonists (ETA, adalimumab (ADA), or Infliximab (INF)) added as the 1st, 3rd and last active drug(s) (representing the introduction of TNFs in early and late stage disease).	1st TNF-inhibitor: when used as the 1st drug, studies of early RA used (MTX naive patients); as the 3rd drug, data sets of both early and later RA used; and as the last active drug, studies of late RA used. Duration of treatment from observational study of ETA, INF and LEF (Geborek, 2002); ADA assumed to be same as INF. Conventional DMARDs (cDMARDs): various studies, some RCTs of early RA; previous treatments not considered. Where data not available assumed to be same as another cDMARD. 2nd TNF-inhibitor: No RCTs were identified of TNF inhibitors in patients with IR to TNF-inhibitors	Used evidence matching line of therapy for 1st TNF-inhibitor. Treatment effects of second 2nd TNF-inhibitor assumed to be same as 1st. Treatment effects assumed to be dependent on disease duration and substitute effect used for some conventional DMARDs. Treatment effect of cDMARDs in early RA halved for use in late RA after TNF inhibitor (*reduced by set amount*); identical for all sequences. If no data available assumed to be same as another cDMARD. A patient's condition was assumed to decline over time.
Cimmino 2011 Italy	**RA** Patients with inadequate response to one TNF-inhibitor	Same 4 predefined sequences as Beresniak, 2011, evaluating the use of a new non-TNF-inhibitor biologic after inadequate response to one and two TNF-inhibitors. *ETA-ABA-ADA* *ETA-rituximab (RTX)-ADA* *ETA-ADA-ABA* *ETA-ADA-INF*	ABA and RTX (non-TNF inhibitors) as 2nd-line: Placebo RCT plus extension study for each drug. ABA as 3rd-line: subgroup analysis from the RCT, of patients with inadequate response to 2 TNF-inhibitors. INF (TNF-inhibitor) as 3rd-line: uncontrolled trial [ReAct] of an alternative TNF (ADA) in patients who had previously received INF +/or ETA. Same study also used to inform 2nd and 3rd line use of ADA, which was the remaining drug making up all 4 sequences.	Used evidence matching line of therapy from RCTs and uncontrolled trial for 2nd-line biologics and subgroup data and uncontrolled trial for 3rd-line. Comparable effects assumed for TNF-inhibitors used as 3rd-line (ADA and INF). Assumed 100% inadequate response to 1st-TNF-inhibitor in the model for all 4 sequences.
Clark, 2004 UK (NICE TA72)	**RA** Patients with early RA and DMARD naive	Comparison of predefined sequences (of conventional DMARDs plus 1-2 consecutive TNF-inhibitors) with and without the additional of a new biologic (non TNF-	Data on ANA from SR of 5 studies (mean no. of previous DMARDs ~ 2 in 3 RCTs, not stated in 2).	Treatment effects assumed to be independent of position in sequence.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
		inhibitor Anakinra (ANA)) One sequence had 3 and the other 5 conventional DMARDs (cDMARDs) before the TNF-inhibitors. New biologic added to the sequence following the TNF inhibitor(s) or as the last active drug.	HAQ improvement on starting treatment for TNFs taken from Jobanputra, 2002; and conventional DMARDs from various selected studies (Including a study of early disease for Gold. *No reliable data for azathioprine (AZA), assumed to be lower end of effectiveness*). Time spent on treatment based on Swedish observational study for TNFs, a published MA of cDMARD termination rates [Maetzel, 2000] for 5 DMARDs, and various studies for remaining 4.	Treatment effect reduced by set amount in sensitivity analysis (*limitations*). Constant disease progression assumed.
Coyle, 2006 Canada	**RA** Patient receiving their primary treatment for aggressive RA.	Adding a TNF-inhibitor (one of 2) to a predefined sequence of conventional DMARDs at two different places (before or after gold). All sequence started with methotrexate (MTX), and if no toxicity, followed by 2 consecutive MTX combination therapies, then gold; *straight to gold if toxicity* [*5 sequences*]	TNFs – single RCTs (and extension studies) for each TNF in patients with IR to previous DMARD including MTX. Gold – simulation of data developed from a published RCT of gold vs auranofin.	Treatment effects assumed to be independent of sequence. *TNF-inhibitors used before or after gold. The d*ecision point, or comparison, was TNF vs gold in different populations, who had either previously received an MTX sequence, or MTX followed by TNF/gold sequence (MTX sequence included 1 or 3 treatment lines). The same effect estimates used at both decision points irrespective of variation in prior sequence.
Davies, 2009 US	**RA** Patients with early RA.	Comparing a TNF-inhibitor (one of 3) vs MTX used at the start of a fixed sequence of 3 conventional DMARDs. A further TNF-inhibitor (etanercept (ETA)) then added to the most cost-effective strategy (adalimumab (ADA) as the 1st-line); *ETA selected based on next best option.* [5 sequences]	TNF inhibitors: ITC of RCTs (each TNF vs MTX in MTX naïve patients with early RA). Subsequent conventional DMARDs: Swedish registry study of 2 TNF-inhibitors (ETA, infliximab (INF)) and one conventional DMARD (leflunomide (LEF)) in established RA (Geborek, 2002)	Used matching evidence to line of therapy for 1st TNF-inhibitor. Treatment effects of 2nd-TNF-inhibitor assumed to be same as 1st; and all conventional DMARDs assumed to have same treatment effect. Treatment response assumed to deteriorate with time.
Diamantpoulus, 2014 UK	**RA** Patients who have had an inadequate response to at least 1 conventional DMARD.	Comparing the current treatment sequence of biological agents with and without tocilizumab (TOC) added as the 1st- or 2nd-drug. Standard care sequence for MTX-contraindicated population: certolizumab pegol (CZP) - ETA - ADA; and for MTX tolerant population: CZP - rituximab (RTX) - ETA - abatacept (ABA) - ADA - INF. (*CZP, ETA, and INF are TNF-inhibitors*)	Treatment response and safety profile for all biologics, except RTX, taken from Bayesian NMA comparing monotherapy vs combination therapy. This included 22 RCTs of patients with inadequate response to conventional DMARDs. The response data for all TNF-inhibitors were pooled; treated as a class. Treatment response for RTX based on clinical trial of RTX in patient with inadequate response to a TNF-inhibitor. *Treatment discontinuation based on Swiss registry data (SCQM-RA), excluding RTX sample.*	Treatment effect is independent of positioning in treatment sequence. Treatment effects for TNF-inhibitors, ABA and TOC based on RCTs of 1st-line biologic, and RTX as 2nd-line biologic. This did not always match positioning in sequence. Same treatment effects used irrespective of whether TOC was added to the sequence or not [displacement], and same treatment effect used for TOC irrespective of whether it was used as 1st or 2nd-line. Class effect used for TNF-inhibitors, assuming exchangeability of the efficacy profile across all agents. No sensitivity analysis conducted relating to treatment sequences.
Diamantpoulus, 2012 Italy	**RA** Patients who have had an inadequate response to previous conventional DMARD.	Incorporating a new biologic (TOC) to an established sequence of 4 biologics with 2 TNF inhibitors followed by two non-TNF biologics. TOC either added to the start of a baseline sequence starting with ETA, or replacing one of 3 alternative 1st line TNF-inhibitors (ADA, INF, ETA). Baseline sequence: ETA-ADA-RTX-abatacept (ABA) *[7 predefined sequences]*	Response rates for TOC and ETA taken from Bayesian NMA of placebo RCTs of 1st-line biologics (ABA, RTX, ETA, INF, ADA, TOC) in patients with IR to 2 conventional DMARDs. Due to lack of evidence about efficacy of TNF-inhibitor after ETA or TOC, response rates for ADA reduced by 30%. NICE TA130 2007 referenced as the data source for this. Response rates for RTX and ABA taken from Bayesian NMA of biologics in patients with IR to TNFs.	Used matching evidence to line of therapy for ETA vs TOC as 1st-line drug. Response rates for ADA reduced to correspond to its 2nd line position in baseline sequence, but ETA also considered as the 2nd TNF in alternative sequences, and it was not stated that the effect of ETA was reduced. The same effect estimate appears to have been used for all TNF-inhibitors in alternative comparator sequences; assuming same class effect.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
			Treatment withdrawal rate based on the average rate for ETA and INF taken from a Swedish registry study of patients with established RA (Geborek, 2002)	Treatment effects for RTX and ABA (last 2 consecutive non-TNF-inhibitors) taken from the same NMA accounting for IR to previous TNF-inhibitor but not another biologic for ABA. Same treatment effects used irrespective of whether TOC was added to the sequence or not No disease progression assumed during treatment; deterioration assumed once patient enters palliative care.
Finckh, 2009 US	**RA** Patients with early rheumatoid arthritis; a disease duration of 3 mths and no previous DMARD use.	Comparison of 3 different management strategies: pyramid strategy (early pain management followed by conventional DMARD then TNF-inhibitors); early DMARD (start with conventional DMARD followed by TNF-inhibitors; and early therapy with biologics (start with TNF-inhibitors followed by conventional DMARDs). The two treatment groups, TNF-inhibitors and conventional DMARDs, included 3 consecutive drugs. NSAIDs - 3 DMARDs – 3 TNFs (*pyramid*) 3 DMARDs – 3 TNFs 3 TNFs - 3 DMARDs	Treatments considered as 3 groups: early pain management; conventional DMARD; and TNF-inhibitors. Patients' response to treatment was categorised as excellent, good, moderate, and none (presented as non-overlapping percentages). Excellent response corresponded to treatment induced remission. Good and moderate response were based on ACR50 and ACR20, respectively. Remission rates (excellent response) for *early* treatment was taken from a sequencing trial (BeSt study) in early RA. The estimate for TNF-inhibitors was based on the infliximab arm. Moderate/good response rates for *early* treatment were taken from a published MA of TNF-inhibitor RCTs in *early RA*. The estimate for conventional DMARDs was based on the MTX arms. The ability to induce remission or achieve moderate/good response was estimated to decline over time. The probabilities for treatment response six months after therapy were estimated using a multivariate relationship based on a number of covariates, which included, among others, baseline HAQ, disease duration, and number of previous DMARDs (based on the analysis of NDB registry data by Wailoo, 2006; the model used was based on the Wailoo model). *The probability of remission (excellent response) and response (good, moderate) were incorporated in the model using a dirichlet distribution to account for the correlation between response types (e.g. cannot be a good responder without being a moderate responder)*. Duration of treatment depended on type of response, which was based on estimates from NDB registry.	Treatment effects of individual treatments were assumed to be independent of position in sequence. Treatment effects for conventional DMARDs, and TNFs were based on an unspecified (generic) treatment effect, with the same estimate used for each of the three consecutive treatments used within each class. It was assumed that response to TNF-inhibitors would be the same whether the patient had previously responded to a TNF-inhibitor or not. Treatment response adjusted for disease duration, which was included as a covariate in the meta-regression analysis. Duration of response accounted for the fact that better responders are more likely to continue treatment. Assumed that patients maintained their response for the duration they remain on treatment.
Hallinen, 2010 Finland	**RA** Patients with inadequate response (IR) to one TNF-inhibitor.	Fixed sequence of 3 conventional DMARDs (best supportive care) with and without the addition of one of 5 biologics (at the start) [*2nd-line biologic*]. Further biologics were then added to the most cost-effective strategy in a stepwise manner [*3rd and 4th -line biologic*]. Biologic (n=5)-best supportive care (BSC) Rituximab (RTX)-biologic (n=4)-BSC RTX-infliximab (INF)-biologic (n=3)-BSC	Biologics: placebo RCT of each biologic; RCTs of TNFs (n=3) included TNF naïve patients; and non-TNF biologics (n=2) included patients with IR to TNFs. *[patients with IR to TNFs entering model]* Duration of treatment for TNFs derived from single study.	Treatment effects assumed to be independent of position in sequence. Same effect estimate used irrespective of point in sequence. Disease progression assumed whilst on treatment; differed for conventional DMARD and biologics).

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Jobanputra, 2002 UK (NICE TA36)	**RA** Patient with early RA (model decision point included patients with inadequate response to 2 conventional DMARDs)	Fixed sequence of 9 conventional DMARDs with and without a TNF-inhibitor (etanercept (ETA) vs infliximab (INF)) added as the 3rd or last drug. *All sequences included the same initial 2 conventional DMARDs (sulfasalazine (SSZ) and methotrexate (MTX)).*	TNF-inhibitors: pooled data from placebo RCTs for ETA (in patients with IR to previous DMARDs) and personal communication for INF (exact same estimate used for both). Initial DMARDs (*SSZ and MTX*): RCT of leflunomide (LEF) vs SSZ vs placebo in SSZ naive patients; and RCT of LEF vs MTX vs placebo in MTX naïve patients). Subsequent DMARDs: Same RCTs as above used for LEF. Substitute data from SSZ used for remaining 6 DMARDs. *LEF and remaining 6 DMARDs used in model after both SSZ and MTX.* Duration of treatment for TNFs derived from single study	Treatment effect of TNF-inhibitor used as 3rd drug assumed to be dependent on number of previous treatments but independent of type of treatments. Assumed that treatment effect of INF is same as ETA. Data for 2nd conventional DMARD based on RCT of 1st time use. Treatment effects for TNF as last drug and conventional DMARDs after the first 2 assumed to be independent of position in sequence. Same effect estimate used for both TNFs irrespective of point in sequence. Substitution effect used for all conventional DMARDs except LEF (assuming exchangeability of effects). LEF based on use in early RA. *Treatment effect of each DMARD modelled as a constant increase in QALY per unit time, with fixed reduction when starting and ending treatment.*
Kielhorn, 2008 UK	**RA** Patients who have had an inadequate response to two conventional DMARDs (including MTX and SSZ) and one TNF inhibitor (etanercept (ETA)). *Baseline characteristics of patients entering model based on RCT of rituximab (RTX) (REFLEX).*	A fixed sequence of 3 conventional DMARDs (cDMARDs), to which a sequence of two TNF-inhibitors were added to the start [adalimumab (ADA)-infliximab (INF)], followed by RTX (a non TNF-inhibitor biological agent). Leflunomide (LEF) - gold - cyclosporin (CyC) RTX - LEF - gold – CyC ADA - INF - LEF - gold - CyC RTX - ADA - INF - LEF - gold - CyC	Treatment response based on adjusted indirect treatment comparison of key RCTs for each biologic (biologic+MTX vs placebo+MTX) and LEF (LEF vs MTX). Average MTX treatment effect used as the reference placebo response rate to develop adjusted response rates that accounted for variation in baseline differences (disease severity or previous DMARDs). RCTs had differing patient populations, with patients in the RTX trial (REFLEX) refractory to TNF-inhibitors. Due to the lack of comparable studies for Gold and CyC, they were given the same response as MTX in the REFLEX trial. Time on treatment for ADA, LEF, gold, and CyC derived from published economic model (Barton, 2004 [HTA]), and for INF published observational study based Swedish registry data. Mortality based on data from normal life table adjusted with RA risk multiplier.	Sequencing effect was assumed to be independent of positioning in treatment sequence. Same effect estimate used for each biologic irrespective of whether it was used as the 1st, 2nd or the 3rd biologic. Treatment effect of the last two conventional DMARDs were assumed to be the same as MTX (from RCT that included patients with an inadequate response to MTX). The same effect used irrespective of prior treatment added.
Kobelt, 2011 Sweden	**RA** Patients with early RA	Comparison of a conventional DMARD (MTX) vs a TNF-inhibitor (ETA). Patients could then switch to a biologic (1st or 2nd) or standard DMARD therapy (ST).	All data sets were available at patient level. Treatment response and discontinuation rates: RCT of ETA+MTX vs MTX+placebo in MTX naïve patients. Data on subsequent biologic from Swedish registry of ETA, INF and LEF (923 patients receiving their first biologic and 125 received a second biologic, excluding those who previously received ETA) and ST from observational study.	Used matching evidence to line of therapy. Assumed that 50% of patients who discontinued their first treatment switched to a subsequent biologic; this was varied in sensitivity analysis. Treatment effect of unspecified biologic taken from registry data. Switching to another biologic not considered due to lack of data. Underlying disease progression assumed.
Launois, 2008 France	**RA** Patients with an inadequate response to a TNF-inhibitor	*Budget impact model: comparing the costs of patients with RA before and after the introduction of rituximab (RTX) used after failure of one or more TNF-inhibitors.*	NMA of RCTs (patients with IR to one or more TNFs or newly diagnose patients)	Treatment effects assumed to be independent of position in sequence. Sensitivity analysis conducted to account for patients in 3rd and subsequent line treatments consuming more resources.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
Lindgren, 2009 Sweden	**RA** Patients with an inadequate response to a TNF-inhibitor	Comparison of a sequence starting with RTX followed by 2 TNF-inhibitors with a sequence of 3 TNF-inhibitors.	TNF-inhibitors (Non-specified) - individual patient level data from Swedish registry (data on up to 3 treatment lines) analysed using Cox-regression with treatment line as one of the covariates. (Rituximab) RTX – placebo RCT (in patients with IR to one or more TNFs). Data for patients who switch to TNF after IR to RTX based on event data for 2nd line TNF.	Assumed that RTX did not influence magnitude of effect of subsequent TNF. Patients in registry had not received prior RTX. Assumed that treatment effect of 4th line TNF same 3rd line. Patients entering model had an IR to their 1st TNF, and in one of the modelled sequences received a further 3 TNFs. Registry only included up to 3 lines. Assumed same class effect for TNFs. Differentiated by line of treatment but not agent.
Maetzel, 2002 Canada	**RA** Patients with early RA and eligible for methotrexate (MTX)	A sequence of conventional DMARDs with and without leflunomide (LEF), added after a series of MTX containing regimens (n=3).	LEF: RCT of LEF vs MTX vs placebo in MTX naïve patient population. Remaining DMARDs: SR and MA of RCTs and observational studies of.	Treatment effects assumed to be independent of position in sequence.
Malottki, 2011 UK (NICE TA195)	**RA** Patients with an inadequate response to a TNF-inhibitor.	Predefined sequence of 4 conventional DMARDs (cDMARDs) with and without the addition of one of 5 biologics (at the start).	TNFs: Observational study of etanercept (ETA) after IR to previous infliximab (INF); and observational study of adalimumab (ADA) in patients with history of ETA or INF. Non-TNFs: placebo RCT plus extension study for each biologic (in patients with IR to first TNF). Conventional DMARDs: only available evidence were trials of early RA (taken from Chen, 2006 (HTA); estimates were halved to represent later RA. *Time to quitting treatment taken from studies of patient registries (study of BSRBR for TNFs and GPRD for cDMARDs).*	Used matching evidence to line of therapy, where feasible. Treatment effects assumed comparable to an alternative treatment used in relevant position in sequence (INF assumed to be same as ETA) Treatment effect reduced by a set amount (Treatment effect of cDMARDs in early RA halved for use in late RA after TNF inhibitor; identical for all sequences). Time on treatment assumed to be same for TNFs. Also assumed to be same for cDMARD and does not change with position in sequence. Sensitivity analysis conducted to test halving efficacy of conventional DAMRD after biologic therapy, and using equal time on treatment for TNFs.
Merkesdal, 2010 Germany	**RA** Patients with an inadequate response to one TNF-inhibitor (etanercept ETA). Baseline characteristics matching patient population in REFLEX RCT	A sequence of two TNF-inhibitors followed by 2 conventional DMARDs with and without rituximab (RTX) added to the start of the sequence.	The ITC included RCT of RTX (patients with IR to one or more TNF) and RCTs of TNF-inhibitors (patients with IR to MTX). Source of effect estimates for cDMARDs not stated. Average time on treatment from a German registry. *(Sensitivity analysis included making this equivalent to RTX for all TNFs and all drugs).*	Treatment effects assumed to be independent of position in sequence. Same generic effect estimate used for both conventional DAMRDs, which did not differ according to sequence. Underlying disease progression assumed.
Puolakka, 2012 Finland	**RA** Patients with inadequate response to one TNF-inhibitor.	6 predefined sequences of 3 biologics, used for comparing two new non TNF-inhibitors (ABA and RTX) as the 2nd drug, with the remaining drugs being different TNF-inhibitors (ADA, ETA, INF).	Abatacept (ABA) and RTX as 2nd-line: Placebo RCT plus extension study for each drug. All TNF-inhibitors as 3rd-line: uncontrolled trial [ReAct] of ADA (a TNF-inhibitor) in patients who had previously received conventional DMARDs or TNF-inhibitors (INF +/or ETA).	All sequences assumed inadequate response to 1st-TNF-inhibitor. Used evidence matching line of therapy for non-TNF-inhibitors. Comparable effects were assumed for all TNF-inhibitors used as 3rd-line, with treatment effect taken form uncontrolled trial.
Rodgers, 2011 UK (NICE TA199)	**Psoriatic arteritis (PsA)** patients who have failed at least two non-biologic DMARDs	Comparison of 3 TNF-inhibitors with the inclusion of a subsequent TNF-inhibitor (2nd line) evaluated as part of a sensitivity analysis only.	Treatment effect for 1st TNF: Bayesian ITC of RCTs (in patients with IR to 2 or more conventional DMARDs). Treatment withdrawal rate for 1st line TNF from MA of registry studies; assumed same for all TNFs as data not considered reliable enough to differentiate between drugs.	Treatment effect for 1st line TNF reduced by a set amount in order to represent its use as 2nd-line. Same reduction factor used for all TNFs. Different reduction factor used if TNF discontinued due to adverse effects or inefficacy. Similar reduction factor applied to risk of discontinuing 2nd-line TNF. The hazard ratio (HR) for failing 2nd-line treatment compared to 1st assumed to be the same for all TNFs.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
			No RCT for 2nd-line TNF. Reduction factor and withdrawl rates for 2nd TNF compared to 1st taken from observational study (UK register); differentiated between reasons for discontinuation.	Underlying disease progression assumed.
Russell, 2009 Canada	**RA** Patients who have failed at least two non-biologic DMARDs	3 predefined sequences of 3 biologics, used for evaluating the use of a new non-TNF-inhibitor (abatacept (ABA)) as the 1st or 2nd drug (remaining biologics were TNF-inhibitors) compared with a sequence of three TNF-inhibitors.	ABA: 2 RCTs and extension studies. One included patients with IR to TNF-inhibitors, and one included patients with IR to previous conventional DMARDs (TNF-naive) 1st TNF-inhibitor (etanercept (ETA)): RCT plus extension study for ETA in patients with an IR to cDMARDs (TNF-naive). 2nd and 3rd TNF-inhibitors (ETA, infliximab (INF) or adalimumab (ADA)): taken from RCT of ABA (in patients with IR to TNF) and assuming 10% reduction in efficacy with switching. Assumption that switching TNFs is associated with lower efficacy based on clinical experts and published observational/registry studies.	Treatment effect for TNF-inhibitors used as 2nd or 3rd line based on the effect of a non TNF-inhibitor using the assumption of a 10% reduction in effectiveness after each switch. The same effect estimate appears to have been used irrespective of whether the prior treatment was TNF or ABA. Comparable effects assumed for all TNF-inhibitors.
Saraux, 2010 France	**RA** Patients with inadequate response to one TNF-inhibitor.	4 predefined sequences of 3 biologics, used for comparing 2 different non-TNF-inhibitors as 2nd-line and a non-TNF inhibitor vs a TNF-inhibitor as 3rd-line (after inadequate response to 2 TNF-inhibitors). All sequences started with ETA and were the same as Beresniak, 2011	ABA and RTX (non-TNF inhibitors) as 2nd-line: Placebo RCT plus extension study for each drug. ABA as 3rd-line: subgroup analysis from the RCT, of patients with inadequate response to 2 TNF-inhibitors. INF (TNF-inhibitor) as 3rd-line: uncontrolled trial [ReAct] of an alternative TNF (ADA) in patients who had previously received INF +/or ETA. Same study also used to inform 2nd and 3rd line use of ADA, which was the remaining drug making up all 4 sequences.	Used evidence matching line of therapy from RCTs and uncontrolled trial for 2nd-line biologics and subgroup data and uncontrolled trial for 3rd-line. Comparable effects assumed for TNF-inhibitors used as 3rd-line. Assumed 100% inadequate response to 1st-TNF-inhibitor in the model for all 4 sequences.
Schadlich, 2005 Germany	**RA** Study conducted in two stages. Stage 1 evaluated 4 sequences in DMARD naive patients. Stage 2 considered DMARD-naïve patients and those in whom a given DMARD had to be changed due to loss of effectiveness or adverse effects. *Leflunomide (LEF) was not considered as an initial option in DMARD-naive patients in healthcare regulations in Germany.*	Fixed sequence of conventional DMARDs with and without the addition of LEF, as the 2nd (in 4 sequences of 5 drugs) or 1st drug (in 4 sequences of 1 or 4 drugs). Sequences selected by rheumatologists. Alternative sequences based on decreasing effectiveness (relative to MTX) evaluated in sensitivity analysis.	LEF vs Methotrexate (MTX): 2 RCTs, 56-67% of patients in these trials received LEF or MTX as their 2nd or subsequent DMARD. The model derives clinical response of other DMARDs by means of their relative effectiveness compared with MTX. These relative effects were derived from published MA of DMARDs used as second line. Decreasing probability of remaining on treatment with DMARDs over time extracted from observational studies (n=8) with at least 3 years follow-up; 5 studies reported termination rates of MTX. withdrawal rate for LEF was estimated by applying the ratio of withdrawal under MTX in the 2 RCTs of LEF vs MTX to the termination of MTX in the observational studies	The model assumed a decreasing effectiveness and probability of remaining on treatment with each DMARD over time. Relative efficacy of the remaining DMARDs (*not LEF and MTX*) were based on the comparison with MTX. Treatment 'response years' for each DMARD was derived by multiplying the interval rate for MTX by the DMARD specific relative effect estimate. However, these relative effect estimates were for DMARDs used as 2nd line, therefore subsequent sequencing effects were not taken into account. Sensitivity analysis included reducing effectiveness by 25% and retention rates by 20% where the DMARDs were used as 2nd or subsequent line. Disease duration not accounted for.
Schipper, 2011 Netherlands	**RA** Patients with early RA	A sequence containing two consecutive conventional DMARDs, 2 consecutive TNFs, and RTX (*5 lines of therapy*) was used to compare 3 different strategies starting with either the 1st conventional DMARD, the 2nd conventional DMARD (add-on therapy), or the first TNF.	Effectiveness based on individual patient level data taken from 2 clinical practice cohorts. Methotrexate (MTX) and Leflunomide (LEF): regional cohort of patients with RA for <1yr and no prior DMARD. The numbers were low for MTX+LEF as data from clinic practice.	Assumed that treatment effect of TNFs in DMARD-naïve patients comparable to that of patients who had at least two DMARDs. Impact assessed using scenario analysis: using estimate of 30% (from an RCT) instead of 20% (from clinical practice data).

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
		MTX - MTX+LEF - TNF I - TNF II - RTX MTX+LEF - TNF I - TNF II - RTX - MTX TNF I - TNF II - RTX - MTX - MTX+LEF TNF = adalimumab (ADA) or etanercept (ETA)	TNFs and RTX: Dutch register of patients starting their first TNF antagonist (received at least two prior DMARDs). Patients starting with MTX, MTX+LEF or TNF were matched using DAS-28 score at baseline. *Scenario analysis based on RCT of TNF (ETA+MTX vs MTX) in patients who were DMARD naïve.*	ETA and ADA selected as the consecutive TNFs, but generic effect estimates used based on patients receiving their 1st or 2nd TNF-inhibitor.
Tanno, 2006 Japan	**RA** Patients who had failed a previous course of the DMARD bucillamine. patient population were modelled after the characteristics of patients in the ETA RCT	Conventional DMARD (cDMARD) sequence with and without the TNF-inhibitor, etanercept (ETA) added to the start. ([ETA] - MTX - SSZ - MTX+ sulfasalazine (SSZ))	Placebo RCT of ETA and open label extension study (patients with IR to cDMARDs; 90% patients had previous MTX). MTX and SSZ: observational study with methods or prior treatment not stated; treatment effect of SSZ+MTX assumed to be same as SSZ.	Used matching evidence to line of therapy. ETA was compared to MTX, which was the first drug in the conventional DMARD sequence. Most patients in the RCT had failed previous MTX. In the ETA sequence MTX was the second drug, and the same treatment effect estimate used for MTX in both sequences.
Tran-Duy, 2014 Netherlands	**RA** Patients who were newly diagnosed. *Baseline characteristics based on registry data (Nijmegen Inception Cohort).* This included patients who were DMARD naive and had a disease duration <1 yr.	Comparison of treatment sequences with and without biological agents. *Strategy 1* (baseline) included 8 available conventional DMARDs (cDMARDs): MTX followed randomly by SSZ or leflunomide (LEF), followed by azathioprine (AZA), cyclosporin (CYC), hydroxychloroquine (HCQ), and gold in random order. *Strategy 2* included same 8 cDMARDs plus 4 biologics: 2 TNFs used after initial 2 cDMARDs, followed by 2 non-TNF inhibitors, then the reaming 5 cDMARDs. The 2 TNF-inhibitors randomly chosen from ETA, adalimumab (ADA), Infliximab (INF), golimumab (GOL), certolizumab pegol (CZP); and 2 non-TNFs from rituximab (RTX), abatacept (ABA), tocilizumab (TOC).	Treatment response and time to events based on individual patient level data taken from 2 clinical practice cohorts: the local registry, Nijmegen Inception Cohort and the Dutch register of patients starting their first TNF-inhibitor [Dutch RhEumatology Arthritis Monitoring (DREAM)]. Biologics: DREAM (included patients who had not responded to ≥1 cDMARD, including MTX). Conventional DMARDs: Nijmegen Inception Cohort (included patients who were cDMARD naïve). Observations for some drugs were very small. All DMARDs received after MTX and SSZ (or LEF) were therefore analysed as one class ('other DMARDs'); all non-TNF biologics used at the same position in the treatment sequence were grouped (class effect); and specific pairs of anti-TNFs were grouped (ETA/CZP and INF/GOL). The same biologic was found to have different effectiveness between first and second administration, therefore treatment effect (absolute changes in DAS-28 score) were sampled for each drug or drug class, distinguishing the first and second biologic, using a linear regression with DAS-28 at the start of the treatment as an explanatory variable.	Used matching evidence to line of therapy. Whether a biologic was used as the 1st or 2nd TNF or non-TNF was taken into account, but the specific drugs previously used were not accounted for. Treatment effect of cDMARDs used later in the sequence were based on a registry of DMARD naïve patients, and assumed to be independent of the strategy used (with and without prior biologics). Treatment effects were grouped, assuming exchangeable efficacy profile within treatment class (for other cDMARDs and non-TNF biologics) or TNF pairs. It was assumed that the effectiveness of a specific drug was independent of the (identity and) cause of failure of the drugs that had been given previously. No sensitivity analysis conducted relating to treatment sequences. Times to end of DAS-28 decrease and loss of response followed exponential curves, with similar values in the distributions among cDMARDs, TNFs, and non-TNF biologics. Times to events were therefore sampled for different classes of drugs instead of individual drugs.
Tran-Duy, 2011 Netherlands	**Ankylosing spondylitis (AkS)** Patients with ankylosing spondylitis with axial involvement	The addition of two TNF-inhibitors (used in sequence) to a sequence of 5 non-steroidal anti-inflammatory drugs (NSAIDs). Biologics were given after failure of 2 NDAIDs. *Sequence with and without biologics*	Drug efficacy and time to events based on various published studies (*no details of studies provided, just references*). Single estimate presented for initial response to NSAIDs, subcutaneous TNF-inhibitor (*referenced studies = ETA*), and intravenous TNF-inhibitor (*referenced studies* = INF). The efficacy of the 2nd TNF was considered to be a fraction of the efficacy of the 1st, with the reduction factor based on data obtained from a registry study and expert opinion.	Treatment effect of 2nd TNF reduced by a set amount. Efficacy of NSAIDs, toxicity and time on treatment were assumed to be independent of treatment history. Class effect used for NSAIDs.

Author, year Country	Patient population entering model	Treatment sequencing include in model	Source of effectiveness	Assumptions relating to sequencing effect estimates
			Owing to a lack of data, the effect of a drug on disease measures in model were independent of baseline measures such as previous drugs.	
Wailoo, 2006 USA (AHRQ)	**RA** Patients treated with a biologic, and for whom treatment with a biologic had not previously failed.	Strategies including TNF-inhibitors, adalimumab (ADA) and etanercept (ETA), and the IL-1 antagonist anakinra (ANA) compared with infliximab (INF). Initial analysis based on single strategies (followed by conventional DMARDs). Subsequent analysis (of the 3 TNFs only) included treatment sequencing, which evaluated using a 2nd or 3rd TNF compared with a single TNF (INF) followed by conventional DMARDs (cDMARDs).	The probabilities of response to treatment were estimated using two separate analysis of clinical effectiveness, one based on a NMA and met-regression of 13 RCTs, and one based on the analysis of a patient registry data (National Data Bank for Rheumatic Diseases NDB). The NMA was extended to incorporate a meta-regression, which included the covariates disease duration, and baseline HAQ. The analysis of patient registry data was based on a meta-regression, with covariates representing (among others): age of patients, disease duration, number of DMARDs failed, and HAQ at start of treatments. *The treatment effects of subsequent cDMARDs were not considered; instead an average deterioration in HAQ over time was applied after withdrawal from biologic.*	It was assumed that the position of a TNF-inhibitor in a sequence of treatments does not affect the probability of response i.e. a patient that has already failed one TNF-inhibitor is as likely to respond to a second TNF-inhibitor as a patient that has not failed a TNF-inhibitor, allowing for the fact that several covariates will have changed (age of patients, disease duration, number of DMARDs failed, HAQ at start of treatments)
Welsing, 2005 Netherlands	**RA** Patients with an inadequate response to at least 2 conventional DMARDs including MTX	Usual treatment with and without the addition of prior leflunomide (LEF) or TNF-inhibitor. The addition of subsequent TNF-inhibitor or LEF, respectively, also considered.	TNF: Pooled data from 2 placebo RCTs of etanercept (ETA) (in patients with IR to 1-4 previous DMARDs). LEF: RCT of LEF vs methotrexate (MTX) (in MTX-naïve patients) Usual treatment: open follow-up study of patients with early RA. Patients who had discontinued sulfasalazine (SSZ) or MTX due to insufficient effect or toxicity were selected from the study. Mean no. of previous treatments at baseline: usual treatment ≥2, TNF 3, LEF 0.8.	Sequencing effects not considered. Single data sources for effect estimates when both LEF and generic TNF were evaluated as 1st and 2nd line. Effect estimate for LEF was reduced by 25% 'because patient population in trial did not match indication in the economic evaluation'. *LEF used as both 1st and 2nd line (after TNF-inhibitor) treatment.*
Wu, 2012 China	**RA** Patients with IR to at least 2 conventional DMARDs including MTX	Fixed sequence of conventional DMARD with and without the addition of one of 3 TNF-inhibitor as the first drug. Rituximab (RTX) then added as the second biologic to each sequence. [*7 sequences*]	Efficacy and for TNFs from ITC of placebo RCTs (patients with IR to previous conventional DMARD including MTX), and withdrawal rates from registry data. RTX from RCT (patients with IR to TNF-inhibitors); and DMARDs used a 3rd and 4th line from Pharmacoeconomic reports [*Brennan, 2004 was referenced for this*]	Used matching evidence to line of therapy. Response rate for conventional DMARDs did not account for positioning in sequence of previous biological failure.

Abbreviations: AHRQ Agency for Healthcare Research and Quality; BSRBR British Society for Rheumatology Biologics Registry; DAS-28 Disease Activity Score 28 joints; DMARD/cDMARD disease-modifying anti-rheumatic drug/conventional DMARD; DREAM Dutch Rheumatoid Arthritis Monitoring; GPRD General Practice Research Database; HAQ health assessment questionnaire; IR inadequate response; ITC indirect treatment comparison; MA meta-analysis; NDB National data bank for rheumatic diseases; NMA network meta-analysis; RA rheumatoid arthritis; SCQM-RA Swiss Clinical Quality Management in Rheumatoid Arthritis; SSATG Southern Swedish Arthritis Treatment Group; TNF tumour necrosis factor.

Geborek, 2002: Geborek, P., Crnkic, M., Petersson, I. F., & Saxne, T. (2002). Etanercept, infliximab, and leflunomide in established rheumatoid arthritis: clinical experience using a structured follow up programme in southernSweden. Annals of the Rheumatic Diseases, 61(9), 793-798.

E. APPENDIX FOR CHAPTER 7: MODELLING APPROACHES FOR ASSESSING TREATMENT SEQUENCES

INCLUDES:

Appendix E1: Cohort models

Table E1.1: Cohort decision tree models used in studies of any condition

Table E1.2: State transition cohort models used in rheumatology studies

Table E1.3: State transition cohort models used in non-rheumatology studies

Appendix E2: Individual sampling models

Table E2.1: State transition individual patient simulation models in rheumatology studies

Table E2.2: Individual patient simulation models in non-rheumatology studies

Table E2.3: Discrete event simulation models used by rheumatology studies

APPENDIX E1: COHORT MODELS

Table E1.1: Cohort decision tree models used in studies of any condition

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
Partitioned survival within a decision tree framework					
NICE CG131 model (*NICE CG131, 2011*) CUA	lifetime	***Cancer (mCRC)*** Pts receiving their 1st chemotherapy treatment for metastatic CRC.	10 sequences of up to 2 lines of treatment. A choice of 4 treatments was available as 1st-line treatment, and 5 treatments were available for 2nd-line.	Quality of life (QoL) and progression free survival. (Disutility due to AEs also considered).	A generic decision tree was implemented for each sequence. The tree started with the receipt of 1st-line treatment. Following disease progression on 1st-line (1st-line **PFS**), the model allowed for a proportion of pts to discontinue treatment, and would receive no further treatment (disease progression until death). The remaining proportion of pts went on to receive 2nd-line treatments (2nd-line PFS). In time they would experience disease progression (until death). While receiving chemotherapy, and prior to the onset of progressive disease, pts were assumed to be in a stable disease state. Following the point of disease progression, pts were assumed to be in a progressive disease state with a lower overall QoL. The main effectiveness outcome in the model was QALYs. Survival time was partitioned in the model using the progression free survival (PFS) and overall survival (OS) results from NMAs. 1st-line PFS was taken from a NMA of 23 RCTs. Overall survival and 2nd-line PFS were taken from a NMA restricted to prospective sequencing studies (n=3). Interventions were grouped by mode of action in order to develop a complete network. Survival time was quality adjusted in the cost-effectiveness analysis using utility weights obtained from published sources. 60% of pts who went on to receive 2nd-line treatment (reported in 15 studies). It was not possible to obtain separate OS curves for the subgroup of pts who only received one line of treatment and the subgroup that received two, therefore the QALY calculations were based on a weighted average of quality-adjusted survival across the combined pt populations and not as a separate absolute estimate for each subgroup.

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
					QALY calculation: For pts who only received one line: (PFS1 x utility in stable) + ((OS-PFS1) x utility in progression); For pts who received 2 lines: (PFS1 x utility in stable) (PFS2 x utility in stable) + ((OS-PFS1-PFS2) x utility in progression). The model did not explore survival conditional on best response to treatment due to insufficient detail reported in the literature to facilitate survival analysis dependent on tumour response. The impact of treatment-related toxicities was accounted for in the model both in terms of disutility due to the patient (based on mean rates for each treatment obtained from the literature) and cost associated with management. **Implementation:** The model was constructed in WinBUGS.
Decision tree models					
NICE CG81 model (*NICE CG81, 2009*) CUA	lifetime	***Cancer (aBC)*** Pts who have received anthracycline therapy.	Evaluated 17 fixed treatment sequences, of up to 3 lines of chemotherapy treatments. This included a choice of 4 different 1st-line treatments. There were 2 active treatments available for 2nd-line and 3rd-line treatments, depending on what was used previously. It was assumed that a chemotherapy agent would not be used twice in the same sequence. Treatment was administered for a fixed period; (inevitable) progression, could occur before or after this. Some of the 17 fixed sequences included the use of 'no cancer therapy' for 2nd or 3rd-line treatment, in order to account for the fact that some patients would have been ineligible for subsequent-lines of treatment, as they had already received them previously.	Probability of response (CR or PR), stabilisation, or non-response Overall survival (OS) (AEs and also considered, in terms of toxic death, and treatment discontinuation due to AEs)	The initial decision node included the choice between1st-line treatments. It was assumed that a pt would receive one cycle of the 1st-line treatment, at which point there was a possibility that the pt might die of toxic death. This could only occur after the 1st cycle of treatment. Those who survived, received 2 more cycles. The pt then faced another chance event of experiencing major toxicity that would lead to the discontinuation of 1st-line treatment or no toxicity, and continue treatment. Those who discontinued then faced another decision node regarding which 2nd-line treatment to use. There was a time-lag of 1 mth between discontinuing 1st-line and starting 2nd-line treatment. Those who continued 1st-line treatment faced the probability of responding to treatment (defined as complete or partial response), having a stable disease, or not responding (progressive disease or non-assessable). Responders and stable pts went on to receive additional cycles of treatment, receiving 6 in total. Non-responders did not receive further 1st-line treatment. Regardless of whether the pt responded to 1st-line treatment or not, progression is an inevitable outcome, but the time to progression will differ. (*Progression occurred on all three branches – response, stable, non-response*.) Once a pt experienced progressive disease, they faced the probability of dying from progressive disease. (*All three branches had a chance node leading to survive or die*.) Death only resulted from progressive disease or toxicity; the possibility of death from other causes was not considered relevant due to the poor prognosis of the pts. If the pt survived, they continued 2nd-line treatment. The pt then experienced the same chance events as with 1st-line treatment (*toxic death, experiencing toxicity leading to discontinuation, responding to 2nd-line therapy*). Once 2nd-line treatment was discontinued or progression reached after completing the full course of 2nd-line treatment, the pt continued on to 3rd-line treatment, which, may include palliative care. If the 3rd-line was active treatment, the same chance events as with 1st and 2nd-line treatment could occur, whilst palliative care was a terminal branch. OS calculation: Progression free survival (PFS for 1st-line +PFS from 2nd-line +PFS from 3rd line + period from progression to death (assumed to be 5 mths, and fixed regardless of prior treatment). PFS for 1st-line taken from NMA of 5 RCTs; PFS for 2nd taken form published studies for individual treatments (n=2); and PFS for 3rd line assumed to be same as 2nd-line (reduced by varying degree in sensitivity analysis). **Implementation:** The model was constructed in Excel and later rebuilt in TreeAge.

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
Dranitsaris model (*Dranitsaris, 2011*) CUA (*value based pricing scheme*)	lifetime	***Cancer (mCRC)*** Pts receiving their 1st chemotherapy treatment for metastatic CRC.	*The model was used to develop a value based price for bevacizumab* Stranded chemotherapy sequence, with and without a 'new drug' (bevacizumab) which is a targeted VEGF. The model started at cycle one of first-line chemotherapy until death. The model compared 2 simple fixed treatment sequences: FOLFOX (+/- bevacizumab) – FOLFIRI	Treatment success (CR, PR, or SD) (AEs and overall survival also considered)	All outcomes of the model resulted in eventual death, which was the ultimate consequence of the population. The decision tree composed of 16 branches or value nodes (8 representing treatments starting with FOLFOX+bevacizumab, and 8 FOLFOX alone). Each chance node was associated with a fixed time period (in mths), representing the time until the event, as well as the probability of the event occurring (depicted as percentage of pts). This provided a fixed time until death (branch termination) for each branch, and an estimation of the total time spent in 'health' state. The model began at the decision node where a choice between 1st-line treatments would be made (FOLFOX+bevacizumab or FOLFOX alone). During the first 2 cycles (*2 mths*) of chemotherapy, pts would be assessed for intolerable toxicity. In cases of severe toxicity, pts switched to 2nd-line treatment, after which they could either experience clinical benefit and continued treatment, disease progression (*5 mths later*), or treatment-related death (*2 mths later*). When progression occurred, best supportive care was offered until death. Pts who continued treatment then went on to experience disease progression (*until death 22 mths later*) or death (*2 mths later*). Pts who did not experience toxicity, would either continue treatment, experience disease progression and switch to 2nd-line treatment until death, or died within the first 2 cycles. Pts who continued treatment could experience disease progression (and switched to 2nd-line treatment) or death. **Implementation:** The model was developed with DATA software (TreeAge Software). *The model, was also implemented, using the same effectiveness data, but different cost and utility data from various cancer centres in different countries [Malaysia, Spain, Canada, India, and South Africa] in order to develop a global pricing index.*
Sciatica model (*Fitzsimmons, 2014*) CUA	1 year	***Sciatica*** Pts presenting to the GP with sciatica.	3 different pathways: i) *primary care management:* initial treatments only (n=5); ii) *a stepped approach*: one of the initial treatments (n=5) followed by an intermediate treatment (n=6) and then an invasive treatment(s) (epidural, or epidural followed by disk surgery); and iii) *immediate referral to surgery* following initial treatment (n=5). This provided 100+ individual treatment strategies, which were compared to 'inactive control'.	Treatment success/failure (global assessment - composite scale)	All the potential treatment sequences were implemented in a single DT. The model began at the decision node represented a choice between 5 initial treatments, with each treatment branch leading to a chance node: success or failure. Treatment success was a terminal branch (ending in a value node), whilst treatment failure then branched to a decision node representing the choice of intermediate treatments, or a value node for success or failure. The successive use of treatments were assumed to have an additive effect. The probability of success for individual treatments, compared to inactive control, were obtained from a single network meta-analysis. Sensitivity analyses included the potential reductions in effectiveness of intermediate therapies and/or surgery (using relative reduction: 10%) and the subsequent effects of non-responders at each stage of the pathway (estimated at 5-10%). Utility values, for treatment success and failure (0.83 and 0.37, respectively) were obtained from a published study. The total utility for each treatment regime was calculated by multiplying the number of successful outcomes of each treatment (for 1000 pts) by the utility of success or failure. It was assumed that there was no reduction in utility for previous unsuccessful treatments, and successful outcome had a utility of 0.83 regardless of how many interventions were required to achieve this (tested in sensitivity analysis). **Implementation:** Software used not stated.
Frankum model (*Frankum, 2005*)	1-3 years	***Onychomycosis***	All possible sequential patterns of 4 treatments used in up to 3	Clinical response (global assessment or >50%	A sequential treatment pathway framework, depicted as a decision tree, was developed for the analysis. The framework potrayed the various series of health states, or possible outcomes and any resulting treatment

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
CEA (budgetary effect of 3 lines of treatment)		Pts within a health plan who are seeking treatment for toenail onychomycosis.	lines of treatment (12 fixed treatment sequences)	reduction of acetated area); treatment failure (AEs or lack of response)	switching, which patients could experience if they followed a sequential treatment pathway. Rather than starting with a treatment decision node, the framework assumed that all pts were initially given one of the designated 1st-line treatments. Pts would then either experience a positive clinical response, or treatment failure (due to an AE requiring discontinuation of treatment during first prescription, or lack of response after a full course of treatment). Those who had a positive response could experience a relapse, and received another course of the 1st-line treatment. Whilst the remainder required no more treatment. The probability of failure on the second course of treatment was then considered, with pts who failed switching to 2nd-line treatment, and again treatment response or failure considered. Pts who initially failed 1st-line treatment switched directly a 2nd-line agent, and again the model considered response rates and relapse rates. Those who failed the 2nd-line agent switched to the 3rd-line treatment. **Implementation:** Software used not stated.
Knoester model (*Knoester, 2007*) CEA (complete success)	1 year (2 x 6-mths periods)	***Epilepsy*** Newly diagnosed pts with epilepsy. A decision tree model was used to depict the potential clinical pathways and outcomes within the first year of treatment.	This included all possible variations of two lines of treatment with 3 antiepileptic drugs (6 fixed treatment sequences)	Complete success (seizure free), partial success (reduction of 50% compared to baseline), or failure (inadequate seizure control or occurrence of unacceptable AEs)	The analysis of the decision tree model provided probabilities of a theoretical patient ending up in one of three outcome groups: complete success, partial success, or failure (*health states*), which were referred to as path probabilities. The decision tree depicted the series of health states, or outcomes, that the patients could follow, and the resulting treatment choice. The model began at the decision node representing the choice between six fixed treatment sequences. The effectiveness of the 1st antiepileptic drug (monotherapy) was assessed after 6 mths. The 2nd antiepileptic drug was used as monotherapy for pts who experienced unacceptable AEs**,** and adjuvant therapy for those with inadequate seizure control (with the first drug being withdrawn after 2 mths). Pts who were seizure free at 6 mths, and did not experience AEs, remained on the 1st antiepileptic drug for the remaining 6 mths. At the end of the year it was assumed that everyone was in one of three outcome groups: complete success, partial success, or failure. This resulted in 6 branches for each treatment sequence. **Implementation:** The model was developed with DATA software (TreeAge Software).
NICE CG152 model (*NICE CG152, 2012*) CUA	*Induction model*: 30 wks; *longest sequence* (treatment cycles) *Maintenance model:* 2 yrs (2 mths)	***Crohn's disease*** 2 models implemented: i) Induction model, which included patients experiencing an acute exacerbation of Crohn's disease (Crohn's Disease Activity Index [CADI] score >150) ii) Maintenance model (used to capture longer term costs and effects of the most cost-effective induction sequence) included patients in remission, with people who relapsed entering the acute induction sequence.	*Induction model* used for comparing 9 pre-defined sequences containing 7 treatments (4 monotherapies, 2 combination therapies, and a generic 'biologic') used in up 4 treatment lines. All sequences started with a monotherapy, and Biologic were used as last therapy. *Biologics are only recommended for people with Crohn's (TA187); it was assumed that pts whose exacerbation failed to response to 2 lines of treatment would be eligible.*	*Induction model:* Treatment specific probability of withdrawal due to adverse events, and treatment-specific probability of achieving remission (CDAI score of < 150) conditional on no withdrawal. (*Response and treatment withdrawal were modelled as mutually exclusive; to obtain treatment effects of remission conditional on non-withdrawal, the number of withdrawals*	*Model used or comparing treatment sequences for the induction of remission:* A decision tree model was used to implement each sequence of 4 treatments. People who withdrew from treatment due to an adverse event or did not respond to treatment moved on to the next line of treatment. It was assumed that people in whom remission was successfully induced, did not relapse, and remission would be maintained until the end of the model. [*Treatment withdrawals was used as a proxy for adverse events; it was considered that costs and disutility pertaining to adverse events for each treatment would be captured by both the additional cost of further treatment, and by patients still having the utility weight associated with active disease*.] The same treatment-specific probability estimates of response and treatment withdrawal were used for each treatment irrespective of its positioning in the sequence. However, the probabilities for monotherapies were obtained from a NMA of 1st-line treatments, and the probabilities for combination therapies were obtained from a NMA of 2nd-line treatments with glucocorticosteroid having failed glucocorticosteroid. Timing of events was based on treatment cycles (duration of treatment trial). Each treatment was assumed to continue to the end of the treatment cycle regardless of whether the patients entered remission, and remission was assumed to have occurred half way through the treatment cycle. Treatment duration was based on clinical practice, which was 8 weeks for all treatments except biologics, which was 6 weeks. The overall QALY was estimated based on both the probabilities of inducing remission for each individual treatment, and time spent in remission over the course of the model for a given treatment strategy.

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
			Patients who were not in remission by the end of the time horizon for the induction model were assumed to undergo surgery. *Maintenance model* used for comparing 5 monotherapies, with the most cost-effective induction treatment sequence embedded within it: glucocorticosteroid (GCS) - azathioprine (AZA)+GCS - biologic - surgery. A three-line sequence used for AZA maintenance treatment (GCS – biologic – surgery), which was explored in sensitivity analysis.	*were removed from the denominators for remission in the NMAs*) *Maintenance model:* Treatment specific probability of withdrawal; relapse; and relapse+withdrawal. Due to the way withdrawals were reported in RCTs two separate analyses were conducted for the clinical review, a non-conservative analysis where only the 'relapse' outcome was analysed, and conservative analysis where 'relapse + withdrawals' was analysed.	*Comparisons of maintenance treatments:* In this analysis only the maintenance treatments (n=5) were varied between comparisons and not the induction sequence. Each treatment was modelled separately. The maintenance model was based on a Markov cohort model with a 2-yr time-horizon (10 yrs explored in sensitivity analysis). A cycle length of 2 mths was used to reflect the duration of induction treatments. It had 8 health states: remission - maintenance treatment; remission - no maintenance treatment; active disease - 1st-line induction; active disease - 2nd-line induction; active disease - 3rd-line induction (biologic); active disease - surgery; remission – on biologic; remission (after biologic) - no maintenance People entered the model in remission states, and those who relapse entered the acute induction treatment sequence. People in whom remission is successfully re-induced went back to their initial maintenance treatment. Patients who fail induction on biologic underwent surgery. If remission is successfully induced on biologics, patients either: i) stayed on biologic until treatment failure leading to dose escalation (could then respond and put back on maintenance, or not-respond and have surgery), ii) completion of 12 mth treatment and re-assessed (could then be in remission and put back on maintenance, or if not have dose escalation). [*Dose-escalation equivalent to re-induction using biologic*]. A series of one way sensitivity analysis were conducted. PSA were also conducted using Monte Carlo simulation, based in 10,000 runs. Sensitivity analyses of the induction model included exploring the effects of including drug-related adverse events for glucocorticosteroid monotherapy; observational data was used to conduct this analysis. They also included varying the treatment durations, based on the average length of the clinical trials; and treatment effect of glycocorticosteroid following budesonide failure. **Implementation:** The Induction model was conducted in Microsoft Excel.
Beresniak / Advanced simulation model (*Russell, 2009; Saraux, 2010; Beresniak, 2011; Beresniak, 2013; Cimmino, 2011; Puolakka, 2012*) CEA	2 yrs (4 x 6-mths periods)	***Rheumatoid arthritis*** All pts entering the model assumed to have an inadequate response to the first Tumour necrosis factor inhibitor (TNF), or the first 2 TNFs, with decision population being partially developed within the model.	Compared fixed sequences of 3 biologic Disease-modifying anti-rheumatic drugs (bDMARDs) (no. of sequences ranged from3 to 6, depending on the study). Treatment sequencing was used for comparing specific treatments at a single point, predominantly 2nd or 3rd-line.	Treatment success/no success Based on thresholds for remission state (RS), and low disease activity state (LDAS) using DAS28 scores (RS <2.6; LDAS ≤3.2)	The decision tree model depicted the pathway of treatments the pt could follow, with each treatment resulting in a probability of success or failure. Two separate dichotomous clinical endpoints, based on levels of disease activity, were used to define success: achieve remission or low disease activity. A separate model was implemented for each treatment sequence and for each outcome measure (2 endpoints). The model was run over 2 yrs using four 6-mth periods. An 100% inadequate response was assumed for the first treatment (or first 2 treatments, deepening on the sequence being modelled), which was followed by a switch to one of 2 treatments. Pts who archived treatment success remained on their existing treatment for up to 2yrs. Pts who failed to respond adequately in 6 mths intervals would proceed to the next treatment option. The model considered the return to conventional DMARD therapy in the case of IR to all successive bDMARDs. **Implementation:** Model based on 5000 Monte Carlo simulations. Software used not stated.
Beresniak / Advanced simulation model (*Bensmail, 2009*)	2 years (4 x 6-mths periods)	**Spasticity** The population was defined as poorly functioning pts who were disabled by their spasticity and dependent for activities of daily living such as: tetraplegic pts;	A treatment sequence, representing conventional medical treatment for the management of severe spasticity, with or without Intrathecal baclofen (ITB)	Treatment success/no success (composite scale)	Two separate models were created to simulate the different treatment pathways being compared. The models started with all pts receiving their 1st treatment, which was either physical treatment or ITB, depending on the model. The model depicted the pathway of treatments the pt could follow, with each treatment resulting in a probability of success or failure. Pts who archived treatment success remained on their existing treatment for up to 2yrs. Pts who failed to respond adequately in 6 mths intervals would proceed to the next treatment option.

Model name (*related publications*) **Economic evaluation**	**Time horizon**	***Disease condition*** **Model patient population**	**Treatment sequences**	**Clinical effect measures**	**How sequences modelled**
Based on the same approach to the Beresniak model for RA CEA		very dependent pts with multiple sclerosis; traumatic brain injury, cerebral palsy, or stroke; adults with modified Barthel score <10 or children with level V Gross Motor Function Classification (GMFCS); or nonambulatory pts.	therapy used as the 1st-line treatment		The DT for conventional treatment was composed of 54 branches and 46 transition probabilities. It started with physical treatment only, moving on to oral treatment, in the case of treatment failure, then one of three strategies: i) neurosurgery followed by nursing; ii) ITB (with dose adjustment if necessary); or iii) focal treatment then neurosurgery. The final option in the strategy was nursing. The ITB as first-line strategy started with ITB, and in the case of treatment failure, was followed by ITB dose adjustment, then pump explanation. This was then followed by one of three strategies: i) oral treatments – neurosurgery - focal treatment or nursing; ii) neurosurgery – nursing; iii) nursing – neurosurgery; or iv) focal treatment – nursing. **Implementation:** The model was developed using Dscript language - Decision Pro (Vanguard, Cary).
Greenhalgh model (*Greenhalgh, 2005*) CUA	12 mths (1 week)	***Major depressive disorder (MDD)*** Patients entering the model were severely depressed and receiving acute inpatient treatment	The model was constructed to allow the evaluation of cost-effectiveness of electroconvulsive therapy (ECT) provided as either a 1st-, 2nd-, or 3rd-line treatment. Eight predefined treatment strategies were modelled, which included ECT or Serotonin and Norepinephrine Reuptake Inhibitors (SNRI) as 1st-line treatment, ECT or Selective Serotonin Reuptake Inhibitors (SSRI) as 2nd-line, and ECT or lithium augmentation as 3rd-line. The treatment strategies also included the use of maintenance/continuation therapy to help prevent relapse. This included ECT, lithium + Tricyclic antidepressants (TCA), or SSRI following ECT; and SSRI following SNRI or SSRI.	'Successful treatment' rates (50% decrease in the Hamilton Rating Scale for Depression HAM-D or other depression scoring system); and failure to complete therapy rates (due to lack of efficacy or AEs).	Decision tree model was implemented using Monte Carlo simulation. Each treatment strategy (n=8) was modelled separately. A one week time unit was used in the model, where, for each week throughout the year the model determined whether the patient was in one of 4 (depression) states: • severely depressed and receiving acute treatment; • *responders*: successfully completed acute treatment, no longer severely depressed and receiving maintenance /continued therapy; • *non responders:* receiving long-term psychotherapy [*on completing psychotherapy assumed to improve to mild depression*]; or • relapsed state following successful treatment [*pts who relapsed from maintenance therapy assumed to require treatment to maintain a quality of life (QoL) equivalent to moderate depression*]. The model attributed a QoL utility score to each state (representing severe, moderate, mild, and depression in remission) and determined the movement through these states. A decision tree model showed that three treatment-lines (referred to as treatment phases) were allowed before a final treatment of psychotherapy was used on non-responders. During each treatment phase (treatment line) there was a probability that the patient could have an adverse event/be deemed as not responding to the treatment, and so move to the next treatment phase before completing the current treatment phase. After completion of a treatment phase there was a probability that the treatment was successful and the patient was discharged. Patients who were deemed not to have responded to treatment moved to the next treatment phase. The probability of successful treatment, and leaving the treatment early due to an adverse event/not responding to treatment was related to the type of treatment received and at which phase of the process the treatment was administered (treatment-line). Following successful treatment the patients may be given continuation therapy to help prevent relapse (maintenance therapy). Duration of treatment was based on a generic estimates of: 6 weeks for pharmacological treatments, with dropouts averaging 2 weeks of treatment; 4 weeks for ECT treatment, dropouts averaging 1 week of treatment. **Implementation**: software used was not stated. The outcomes were based on 3,000 simulation runs completed per treatment strategy.

Abbreviations: AE Adverse event or effect; BC/aBC breast cancer/advanced breast cancer; CEA cost-effectiveness analysis; CR complete tumour response; CRC/mCRC colorectal cancer/metastatic colorectal cancer; CUA cost-utility analysis; DMARD/bDMARD/cDMARD disease-modifying anti-rheumatic drug/biologic DMARD/conventional DMARD; DT decision tree; mth month; NMA Network meta-analysis; OS overall survival; PFS progression-free survival; PR partial tumour response; pt, patient; QALY quality-adjusted life year; SD stable disease; yrs years.

Table E1.2: State transition cohort models used in rheumatology studies

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
Simple Markov cohort model					
York psoriatic arthritis (PsA) model (*Rodgers, 2011*) CUA	Lifetime (3-mths)	Base case model developed to compare TNF-inhibitors (ETA, INF, ADA) for the psoriatic arthritis (PsA); with the subsequent treatment being palliative care. This was extended to allow patients to switch to a second TNF-inhibitor as part of the sensitivity analysis. The model represents an update of a previous York PsA model, by also taking into account the impact of biologics on the psoriasis component of PsA.	The decision population, relating to sequential TNF-inhibitors, was developed within the model. The population represented the average characterises of participants in the RCTs, and included patients with active and progressive psoriatic arthritis (PsA). Patients entering the model were assumed to have failed at least 2 cDMARDs, but naive to biologics.	Health states for biologics: withdrawal from TNF, continue on TNF with response of arthritis but not psoriasis, continue on TNF with response to both arthritis and psoriasis, continue on TNF with response of arthritis but not psoriasis, death. Health states for biologics: no treatment, death. Events: achieving PASI 75 response or not, achieving PsARC response or not; die.	In the base case scenario patients who withdrew from biologic treatment went on to receive palliative care, but the use of a second TNF-inhibitor was considered as part of the sensitivity analyses. The impact of biologics on the arthritis component was modelled via a change in the HAQ and the impact on the psoriasis component using the Psoriasis Area of Severity Index (PASI). The structure and events in the model aimed to be consistent with licensed indications and National Guidelines. Patients entering the model either received a TNF-inhibitor (ETA, INF or ADA) or palliative care (no biological therapy). For biological treatment, initial treatment response was determined using the PsA response criteria (PsARC). Patients who had a PsARC response at 3 months (end of first cycle) continued on biologic treatment. Those who did not respond discontinued biologic treatment and switched to palliative care in the base case. It was assumed that no patient withdrew due to adverse effects within the first 3 months. Patients receiving biologics (PsARC responders and non-responders) were then assessed for PSAI 75 response or non-response. The impact of treatment on the two PsA components were then modelled. For PsARC responders, there was a HAQ gain that corresponded to a drop in HAQ score. *There was little data to inform whether patients who remain on biologics maintained initial HAQ improvement long term*. In the base care the initial gain was assumed to be maintained while the patient continued with biologic treatment. For non-responders, a slight HAQ gain was estimated during the initial (3 month) treatment assessment, after which they switched to palliative care and the associated natural history progression of HAQ. For the psoriasis component, patients who achieved PSAI 75 gained at least 75% improvement in psoriasis compared with baseline, whilst those who did not had <75% improvement. The base case model assumed an ongoing, constant rate of withdrawal from biologics after 3 months (for those with an initial PsARC response). Withdrawal could occur due to lack of continuing efficacy, AEs, or other reasons. The rate of withdrawal after 3 months was assumed to be independent of HAQ and PASI scores, and independent of whether initial response was for psoriasis and arthritis, or just arthritis. The same withdrawal rate was assumed for all TNF-inhibitors. *There was little evidence on rebound after withdrawal from biologics*. In the base case, it was assumed that rebound was equal to initial gain for HAQ and PASI scores. The model assumed no differences

					in mortality rates between treatments, or biology therapy and no treatment. For the analysis of sequential TNF-inhibitors, the reason why the first TNF-inhibitor was discontinued was considered as two subgroups (analysed separately): adverse effects, or lack of efficacy. The drug that was used first was ineligible for use as second-line. **Implementation:** R programming software.
Albert model (*Albert, 2000*) Not applicable (effectiveness only)	5 yrs (6-mths, x10)	Comparing 4 fixed sequences of 3 cDMARDs (monotherapy) A conceptual model (decision tree structure) was initially developed for comparing 2 different treatment approaches: pyramid (sequence of monotherapy) vs step-up (combination therapy). However, the implemented models were used for comparing 4 fixed treatment sequences (of monotherapies) in order to minimise the complexity of the decision problem.	Pts with early RA starting their first DMARD (pts' baseline characterises not reported) *The use of 3 different data sources to inform treatment response were compared:* *i) expert opinion (senior rheumatologists): improved / active / toxic* *ii) Meta-analysis: efficacy (composite treatment effect)/ no efficacy (no treatment effect and AEs)* *iii) Survey of US rheumatologists: effective (ratings of drugs as good or excellent; includes AEs) / not effective (ratings of drugs as poor, fair, or moderate)*	3 health states: improved; active; toxic. 18 events (branches) in an initial analysis and 26 in an expanded analysis. The branches of the Markov tree represented each health state plus instruction on the subsequent treatment, e.g. 'active, take 2nd drug', or 'toxic, take 3rd drug'. But, in the initial analysis 'improved' was a terminal branch, and in the expanded analysis included the instruction to take the same drug. 'Active, stop 3rd drug, or toxic, stop 3rd drug were also terminal branches. The branch 'active, take 1st drug' represented the initial state. (*In the conceptual model, 'toxicity' lead to discontinuation of one treatment and starting another, whilst 'ineffective' lead to adding a new treatment (rather than both 'toxic'' and 'active' leading to the same treatment switch*.)	The model was used to estimate the mean time spent in the improved state for a specific sequence, with separate models run for each sequence. The initial model was also used for comparing 3 different data sources to inform treatment response were. These estimates did not consider time on treatment. Treatment sequencing was implemented using a simple Markov tree structure. Every pt in the cohort was assumed to start on the same first drug, and then move to the 'improved', 'active (take 2nd drug)', or 'toxic (take 2nd drug)' state etc. In the initial Markov analysis the 'Improved' state was modelled as an absorbing state (no further treatments used). In the subsequent expanded Markov analysis the probabilities for continuing to take a drug and developing toxicity were varied with time, and were different for each drug. This was achieved by representing the improved state as a tunnel state, and duration of therapy as a means of terminating the improved state (using published data on duration of treatment and toxicity over time, which were presented in tables). "*Pts would have a period of time in the improved state and then be cycled back to the active or toxic states. 'Improved' was not be an absorbing state, but pts who had tried all 3 drugs and were still 'active', or experienced toxicity to all 3 medications would be in an absorbing state*." **Implementation:** Software used not stated.
Maetzel model *(Maetzel, 2002; Coyle, 2006)* CUA/CEA (ACR20RY)	5 yrs (6-mths, x10)	Evaluated adding LEF to a predefined sequence of DMARDs (at different points, depending on toxicity to MTX). LEF was added after a sequence of up to 3 treatments containing MTX, representing step-up combination therapy, and before gold. [*Coyle: adding a TNF to the sequence with LEF at different points]*	Pts with RA severe enough to require treatment with MTX (pts' baseline characterises not reported)	2 health states: 'continue same DMARD'; 'start new DMARD' 3 events: continue treatment (≥ACR20); stop due to severe AEs; stop due to lack or loss of effect (<ACR20). A sub-decision tree diagram (for each cycle) showed that pts who continued the same DMARD could experience clinical response or no response, and could also experience AEs (or none) that were minor enough to continue treatment.	Separate models were run for each sequence and compared in terms of the average time spent in the state of response. Cost-effectiveness was assessed in terms of cost per additional year of ACR20 response, and cost per additional QALY. The patient pathway followed by those who either 'continue (≥ACR20)' or 'stop treatment' (within a single cycle), in terms of depicting the type of response or AEs experienced, was implemented using a decision tree. The tree started at the decision node 'DMARD'. At the end of each cycle (terminal branches of the decision tree) pts were either in the 'continue same DMARD', or 'switching DMARD' state. In the tree the initial 'stop treatment' 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' or 'no clinical response', 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'. The model thus accounted for the fact that some pts may continue treatment despite 'no response' as per ACR20, but based on other criteria, such as X-ray. Pts could also cycle through different treatment sequences (skipping treatments) depending on toxicity (to MTX) or lack of efficacy. Treatment response and withdrawal rates were obtained using a SR and MA of observational studies and RCTs. Combined withdrawal rates were obtained using parametric regression, assuming an exponential hazard function. Withdrawal rates were converted to 6-mth treatment withdrawal probabilities for use in the decision model. **Implementation:** Maetzel - Software used not stated. Coyle – Probabilistic sensitivity analysis conducted using Monte Carlo Simulation conducted using Crystal Ball software enhanced for Microsoft Excel.
Tanno model (*Tanno, 2006*) CUA	Lifetime (6-mths)	Compared 2 fixed cDMARD sequences, with and without ETA (added to the start)	Decision population partially developed within the model. All pts entering model assumed to have an inadequate response to the first cDMARD (bucillamine) (pts' baseline characterises taken from RCT of ETA)	5-6 health states: [ETA]; SSZ; MTX+SSZ; no DMARD; death. 3 events: remain on current drug; switch to the next drug (if failed to achieve ACR20 or experienced severe AEs); die. *Probability of treatment discontinuation due to AEs incorporated (based on dropout rates in clinical trials due to AEs), but AEs did not alter the type of subsequent treatment used.*	'*Pts entered one of two treatment pathways*.' Treatment sequencing was implemented as a series health states, which pts progressed through in a fixed order. Pts could not switch to earlier treatments. At the beginning of each cycle, pts could remain on the current treatment; switch to the next treatment, if failed to respond on current treatment or experienced severe AEs; or die. Death was an absorbing state. Probability of death for RA pts in any time cycle calculated using an exponential equation incorporating age, sex, and HAQ score. Age- and sex-specific excess mortality due to RA included as an exponential function of the HAQ-score. Utility based on HAQ score, which was assumed to change by a set amount during each cycle when ACR20 achieved (factor of 0.53 used for ETA). Pts not achieving ACR20 assumed to have no change in HAQ on ETA. For the remaining treatments, HAQ score was assumed to increase at a low background rate. **Implementation:** Software used not stated.
Welsing model *(Welsing, 2005)* CUA/CEA	5 yrs (3-mths, x20)	Comparing fixed sequences of DMARDs (representing different sequences of TNF, LEF, and usual care (UC); and usual care with and without LEF or TNF) [TNF based on ETA data]	Pts who satisfy the indication for TNFs (baseline pt characterises based on the three datasets used to inform the transition probabilities for individual treatment options; the datasets were based on clinical trials).	4 health states defined by DAS28 score: remission (<1.6), low disease activity (1.6-2.4), moderate disease activity (2.4-3.7), and high disease activity (>3.7). *(The model structure was based on the fact that RA is a chronic disease with a varying disease course over time, characterised by periods of high disease activity altering with low disease activity or remission.)*	Separate models (same structure) run for each sequence. The results were compared between treatments strategies, based on expected %age of time each treatment (LEF, ETA, UC); and yrs spent in the Markov states (disease activity). Markov states represented transition through different levels of disease severity, with the simulated cohort started in the 'high disease activity' state. In the case of non-response at 3 mths, pts switch treatment. *Non-response appears to have been based on time spent on each treatment (in remission) obtained from the published literature.* Reported %age of time on treatment for each sequence: LEF-TNF-UC: LEF (51.2), TNF 41.2, UC 7.6

				Treatment specific transition probabilities between states taken from a separate (unadjusted) dataset each treatment, TNF, LEF, and UC.	TNF-LEF-UC: LEF 8.0, TNF (84.4), UC 7.6 LEF-UC: LEF (51.2), TNF 0, UC 48.8 TNF-UC: LEF 0, TNF (84.4), UC 15.6 *Utilities related to each state were derived from a clinical trial of MTX (based on EuroQoL-5D)* Implementation: The model was built using Microsoft Excel. The programme Crystal ball (version 4, Decision Engineering) was used for the probabilistic sensitivity analysis.
Builds on Welsing model *(Schipper, 2011)* CUA	5 yrs (3-mths, x20)	*Compared 3 fixed treatment sequences of 5 drugs use in different order; sequences included a sequence of 2 TNFs and RTX*. (Sequences included mono- and combination therapies, including step-up strategy with MTX)	*Pts with early RA starting cDMARD.*	4 health states defined by DAS28 score: *remission (<2.6), low disease activity (≤3.2 to >2.6), moderate disease activity (>3.2 to ≤5.1), and high disease activity (>5.1)*] Events: treatment switching depicted in Markov tree. Markov node branched to each of the 5 individual treatments. Each treatment was then attached to a decision node (sub-tree) with branches representing the 4 health states; 3 of which lead to switching to the next treatment, whilst remission lead to the same treatment; Each treatment specific remission lead to the same sub-tree.	Each sequence was modelled separately. The model was used to estimate the percentages of remission at the end of simulation (5 yrs). A less strict criterion (achieving low disease activity: DAS28≤3.2) was also used in a scenario analysis. Percentage of pts in remission (DAS <2.6 state) after each treatment (1st, 2nd, 3rd, 4th, and 5th) for all sequences were reported. (After 10 cycles (2.5 yrs), equilibrium was reached in each treatment strategy, meaning there were no more transition between Markov states. There was no difference in the number of pts who were in remission between the 3 strategies; 38% of pts had sustained remission.) Treatment sequences implemented using Markov tree structure. All pts receiving the first drug in the sequence, and then branched to the second drug or remission on the first (via a subtree depicting different levels of disease activity) etc. 'Pts were initially distributed across several disease states (Markov sates) defined by remission'. After the 1st cycle pts may be in remission (DAS28<2.6) and remain on initial treatment for next 3mths of not in remission (non-responder, DAS28≥2.6) and switch to the next treatment. Pts were assumed to sustain remission after being in remission for 2 cycles. Percentage of pts achieving remission, for each treatment based on IPD from 2 cohorts: an inception cohort and a National pt registry for TNF treatment. Pts who received one of the strategies, with complete assessment at baseline, 3 mths, and 6 mths were selected from the 2 cohorts. The occurrence of a DAS-28 Markov states after treatment was used to calculate transition probabilities (occurrence of a DAS state/total treatment group). Data for cDMARDs (LEF, MTX) taken from the inception cohort, and (TNF I-II, RTX) from the pt register (pts had received at least 2 DMARDs before a TNF). As the cohort data were likely to result in different baseline characteristics between pts starting with MTX, MTX+LEF or TNF antagonist, these were matched using DAS-28 at baseline. Clinical response to TNFs in DMARD-naïve patients assumed comparable to those who had failed previous DMARDs; tested in scenario analysis using RCT data. Scenario analysis included applying the assumption of higher effectiveness of TNF in early RA (based on RCT data).

					{Efficacy data were based matching evidence of using a first and second TNF-antagonist.) Implementation: Software used not stated.
Wu model *(Wu, 2012*) CUA/CEA	Lifetime (6-mths)	Comparing 7 fixed sequences, which included a baseline sequence of 4 cDMARDs, with and without one of 3 TNFs used alone, or followed by an RTX. Biologics added to the start of the baseline sequence.	Pts with inadequate response to at least 2 cDMARDs including MTX, and eligible for TNFs (baseline pt characterises taken from published HTA and comparative study)	3 states: maintain treatment (ACR response); receive new treatment; death. (*It was assumed that ACR response lead to improvement in HAQ scores, and that HAQ scores in non-responders deteriorates and the disease relapses. HAQ scores were converted to utilities*.)	Pts entering the model start the first treatment. At the end of each treatment cycle, treatment response data was accessed. Those in remission (achieving any ACR response criteria 20/50/70) maintained treatment, and those who were non-responders (ACR00) or experienced intolerable AEs switched treatment. Pts could also move to death (natural mortality) at the completion of each cycle. Used normal life tables and adjusted mortality risk for pts with RA (1.33 per unit HAQ). **Implementation:** Software used not stated.
Partitioned survival (area under the curve) model					
Schadlich model *(Schadlich, 2005)* Model based on the unpublished 'Avara interactive model' CUA/CEA (ACR20RY; ACR50RY; ACR70RY)	3 yrs (6-mths)	Adding LEF to a sequence of 5 cDMARDs, by comparing fixed sequences of cDMARDs with and without LEF (at different points).	Evaluation implemented in two stages: 1st limited to DMARD-naïve pts, and the 2nd considering DMARD-naïve pts and those on their 1st cDMARD (characteristics of baseline cohort were based on 2 RCTs of LEF vs MTX, where previous MTX was excluded in one, and previous cDMARDs were allowed in the other)	6 states: DMARD 1-6 ??ACR response used to differentiate health states. Treatment sequences compared in terms of the time the patient benefited from treatment. Effectiveness was based on response years (RYs) gained according to ACR20/50/70 criteria, and by QALYs gained, which were quantified as area under the curve (AUC) calculations.	Simulation model, based on the migration over time (in 6 mth intervals) of a patient cohort treated with a fixed sequence of 6 DMARDs Each ACR response category (and QoL) were modelled separately. ACR response years (ACR20/50/70RYs) were quantified using the AUC for the proportion of pts with ACR response (ACR20/50/70) within a given interval. Data on efficacy (ACR response rates) were taken from 2 RCTs (for LEF and MTX) and a MA for remaining DMARDs (relative to MTX), and termination rates of all DMARDs taken from observational studies with a minimum of 3 years follow-up. The AUC survival functions were determined graphically. A survival plot, depicting the decrease in the proportion of patients remaining on treatment, for each DMARD (n=6) over 3 years, was presented (*providing the proportion of pts remaining on treatment, for each DMARD, at the beginning of each successive 6-month interval*). Treatment sequencing was implemented as a series health states, which pts progressed through in a fixed order. The model was able to account for the decreasing probability of remaining on treatment with a given DMARD. Pts switched to the next DMARD due loss of effectiveness or AEs at the start of each interval. The migration of pts over 3 yrs with each predefined sequence was based on the DMARD specific proportions of pts remaining on treatment at the beginning of each treatment interval (as depicted in the survival plot). At each interval, patients were either treated *initially* with a given DMARD or received *follow-up* treatment with the same DMARD as the proceeding interval. This differentiation allowed the application of quantified costs and effectiveness parameters. The respective values relating to the first 6 mths were applied to pts assigned to initial treatment within an interval, and the values relating to the following 6 mths were applied to pts assigned to follow-up treatment within the interval. (The same values were used for each DMARD irrespective of the treatment sequence and positioning used.) The interval related representation was necessary for proper discounting of costs and effects in the second and third year of treatment.

					Implementation: Software used not stated.

Abbreviations: ACR20/50/70 American College of Rheumatology response criteria; ADA Adalimumab; AE adverse event or effect; cDMARD conventional Disease-modifying anti-rheumatic drugs; CEA cost-effectiveness analysis; CUA cost-utility analysis; DAS28 Disease Activity Score 28 joints; ETA Etanercept; HAQ Health Assessment Questionnaire; INF Infliximab; LEF Leflunomide; mths months; MTX Methotrexate; pt patient; RA, rheumatoid arthritis; RTX Rituximab; SSZ Sulfasalazine; TNF Tumour necrosis factor inhibitor; UC usual care; yrs years

Table E1.3: State transition cohort models used in non-rheumatology studies

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
Simple Markov cohort used for modelling predefined sequences					
Beard model (*Beard, 2011*) CUA/CEA (additional symptom control mths)	2 yrs (3-mths)	Evaluating optimal positioning duloxetine (DUL) within a standard treatment sequence of 5 treatments (DUL added at different points (1st-6th treatment), resulting in 7 fixed sequences, with 5-6 treatment lines). Treatment was used for 3 mths if successful and no AEs.	***Fibromyalgia (FM)*** Adults eligible for pharmacotherapy for FM, which they would have had for at least 3 mths.	5 health states: full response (≥50% improvement in pain severity) with AEs; full response without AEs; partial response (30-49% improvement) with AEs; partial response without AEs; and inadequate response (<30% improvement). Events: pts with inadequate response, or experienced intolerable AEs, switched treatment. Pts who achieved full response or partial response with no AEs continued treatment (for 3 mths). The clinical effectiveness outcome was additional symptom control mths, defined as the amount of time spent at a response level of ≥30%	The model structure was designed to represent the movement of pts through a series of treatment lines. Markov health-state transition methodology was used to track pts through a series of health states, delineated by predefined levels of improvement in baseline severity score. The model included five clinically important health states, which considered 3 discrete-pain response health states and the possibility of experiencing intolerable adverse effects (AEs). Where pts failed to achieve at least partial response, or experienced intolerable AEs, they were then switched to the next treatment in the sequence. Pts who achieved full response or partial response with no AEs continued treatment (for 3 mths). The model assumed average pain improvement applied to the first 3-mth treatment period, for full responders (70%) and partial responders (38%). Starting with an average baseline score of 6.5, the model assumed the majority of improvements would be achieved within the first two weeks of treatment, reaching a plateau at three mths. Beyond the 3-mth point, the model assumed that patients could maintain levels of pain response during the two-year time horizon, provided they remained on active treatment. The model assumed that patients who moved through the full treatment sequence and remained in an inadequate response health state continued to experience their baseline pain severity. The model allowed for a proportion of patients to drop-out of current treatment, which would then be lost to subsequent treatments. An annualised rate developed based on 2 RCT extensions for DUL. The same percentage (25%) was used for all active treatments, and explored across a 20-30% in sensitivity analyses. **Implementation:** The model was developed in Microsoft Excel.
Cameron model (*Cameron, 2008; Lux, 2009*) CUA	10-yrs/ lifetime (28 days)	Adding fulvestrant (FULV) to a sequence of 4 treatments (aa two different points, 2nd and 3rd-line) [*Lux: FULV added as 2nd line only*]	***Advanced breast cancer (BC)*** Hormone receptor-positive postmenopausal women with advanced breast cancer whose disease has progressed or relapsed after previous treatment with oestrogen therapy.	7-8 states (depending on sequence): initial treatment state, 4-5 treatment (lines) states, best supportive care (BSC), and death. Events: remain in same state; experience disease progression and move to another line of treatment; or die. Disease progression was characterised by time to progression (TTP).	The model consisted of health states representing each line of treatment and death. The model was set up to compare 2 cohorts of identical pts receiving different sequences of treatments, where Cohort A included the option of fulvestrant, whilst Cohort B did not include fulvestrant. At the end of each cycle, pts could either remain on their current line of treatment, experience disease progression and move to another line of treatment, or die. The model allowed one or more treatments to be skipped, as well as direct transition from every treatment to BSC or death. **Implementation:** The model was developed in Microsoft Excel by Cameron *et al*., and in Microsoft Excel and Microsoft Visual Basic by Lux *et al*.
Davies model (*Davies, 2008*) CUA	10-yrs (18 wks)	Treatment sequences contained 2 of 4 antipsychotics followed by clozapine (providing 12 alternative sequences).	***Schizophrenia*** Pts with stable schizophrenia	10 heath sates: stable, relapse, and 'experience AEs' (which included one of 3 predefined AEs or diabetes) associated with each of three treatment lines, and death. The number of states were not stated. The schematic showed that patients could transit	The model reflected the fact that on each treatment pts may relapse, discontinue, experience predefined AEs (n=3), or develop diabetes. It also included a series of three treatment options (treatment #1, treatment #2, and clozapine), which pts would cycle through. It was assumed that pts who became refractory to clozapine would have an atypical added to clozapine after the predicted time of clozapine discontinuation. With the exception of clozapine, which increases mortality due to agranulocytosis (particularly in

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
				between the stable, relapse, and experience AEs states. Treatment switching only resulted from the stable and relapse states. Both these states were also linked to death.	the first 18 weeks), no treatments were considered to directly increase the risk of mortality. **Implementation:** Software used to develop model was not stated.
Heeg (cancer) model (*Heeg, 20015*) Not applicable (Modelling overall survival)	Lifetime (1 mth)	Multiple myeloma (MM) is incurable, so main aim of consecutive treatment is to increase survival. In clinical practise, complete response is used as a short-term marker for treatment success, as it is a proven predictor of OS in MM. The study compared 17 fixed chemotherapy sequences, which included 4 lines of treatment. Only treatments for which RCT data were available were included. (A SR of RCTs comparing treatments for newly diagnosed and relapsed patients was conducted.) Only one treatment (bortezomib), for which data on efficacy of re-treatment was available, was allowed as a re-treatment within a sequence.	***Multiple myeloma*** Newly diagnosed patients with multiple myeloma, who were ineligible for stem cell transplantation (SCT). Median survival after diagnosis ranges form 36-60mths	11 health states: representing three levels of response (complete CR, partial PR, and none NR) associated with each of the first three lines of treatment; 4th-line treatment (representing later lines of therapy) [*represented by a single health state*]; and death. Non response data was needed for the 4th-line treatment as it was assumed that patients remained there until they die.	The primary outcome for comparing the different sequences was median overall survival (OS). Secondary outcomes included life expectancy (i.e. mean OS), and time on 1st-, 2nd-, and 3rd-line treatment. Patients entering the model transited to one of three health states, representing the different levels of response to first-line treatments. At the start of the model patients were distributed over the response categories in 1st-line based on the treatment they receive in first line. In each following monthly cycle, the members of the imaginary cohort progressed through the model, i.e. they could remain in the current response state, switch treatment, or die. In the model, response rates (specific to the treatment itself and the line of treatment) were combined with the probability of switching treatment (specific to response category and line of treatment) and the mortality probability (specific to response and line of treatment). Patients who switched treatment were then redistributed over CR, PR and NR health states in 2nd-line, where they could again remain on treatment, switch treatment or die. This calculation process was repeated until the cohort of patients enters the "later lines of treatment" state where they remained until they died. *Treatment specific* probabilities of response (CR, PR, and NR) on 1st-line treatment were based on data obtained from a network meta-analysis (NMA) of newly diagnosed patients. The *response and line specific* probability of transition to the next treatment or death were obtained from a single trial (VISTA study) of newly diagnosed patients (using a Weibull model). The *treatment specific* probabilities of response on 2nd- or 3rd- line treatments were obtained from a second NMA of relapsed/refractory patients. The *response and line specific* probability of transiting to the next treatment (3rd-line or 'further lines', respectively) or death were obtained from a published study (APEX trial) 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 lines' health state were obtained from a third trial (SUMMIT trial), which included patients who had already received many treatment before entering the trial. For consistency, the treatment switch and mortality probabilities obtained from the selected single trials were derived from the results of the bortezomib arm.
Lee model (*Lee, 2013*) CUA	10-yrs/ lifetime (9 wks)	Comparing two chemotherapy sequences starting with different regimens. The remaining treatments were identical, and included 4 lines of treatment,	***Ovarian cancer*** Women with platinum-sensitive ovarian cancer (failed 1st-line treatment).	Study described as having 4 health states: responsive, progressive, clinical remission, and death. The diagram depicting the structure of the model also included three treatment states: 1st-treatment (as an initial state); 3rd-6th treatment; and BSC.	The model compared 2 cohorts of identical pts receiving a sequence of different chemotherapies and best supportive care (BSC). The treatment period, if the pts responded, was 18 weeks. If pts did not progress or show any serious AEs in this time they would enter a clinical remission state, withdrawing from the drug. If pts progressed on treatment they would enter the next line of treatment. 'Progress' was a tunnel state, which pts had to pass through to enter the subsequent line of treatment or BSC, and

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
		with a choice of 2 drugs as the 2nd and 4th treatment.		*Transition probabilities (TPs) to the next line of treatment and death were drawn from treatment specific median TTP or PFS and overall survival data obtained from clinical trials. Same TPs applied in both cohorts for treatments 2-4, but no. of pts progressing from the1st treatment will have differed.*	could not revert to an earlier line of treatment. At any given time (cycle), a patient could remain on a current treatment or make a transition to the next treatment or death. The time horizon was set at 10 years, at which point 99% of the cohorts had died. The diagram depicting the structure of the Markov model showed the '1st treatment' as an initial state and 'treatments 2-5' as a recusing state. Respond, remission, BSC, and death were also recusing states, and progress was a temporary state between the respond or remission states and either 'treatments 2-5' or BSC. **Implementation:** Model was developed using Microsoft Excel spreadsheet.
Orme model (*Orme, 2012*) CUA	Lifetime (3-mths)	Comparing 3 predefined sequences starting with one of 3 topical hypotensives (prostaglandin analogues PAs), followed by the progressive addition of timolol (combination therapy), then dorzolamide (triple therapy). If 1st drug was discontinued due to poor tolerance, then the latter drugs were used as monotherapy.	***Glaucoma*** Pts with mild-to-moderate glaucoma or ocular hyper tension (OH, with no visual field loss) and eligible for long-term topical hypotensive therapy. The model considered low and high risk pts separately, in order to reflect differences in management and risk (base case used a 50:50 mix).	5 health states [and associated visual field defect (VFD) progression]: - OH: no visual field defect (VFD) [progression 1, abnormal VF]; - mild glaucoma: some VFD, vision unaffected [progression 2, worsening VF]; - moderate glaucoma: VFD manifest as loss of vision [progression 3, worsening VF]; - severe glaucoma: severe vision loss; and - death. 3 key triggers (events) for treatment switching: lack of tolerance (T1); intra ocular pressure (IOP) not meeting benchmark (T2); progression in VFD (T3). *Clinical effectiveness based on reduction in glaucoma progression, and low VF*	It was hypothesised that cost effectiveness could be optimised by minimising treatment switching. Treatment sequences were compared in terms of time spent in each line of treatment. The difference in the cost between the 3 strategies was a consequence of the difference in time spent in each line of treatment. At the start of the model is was assumed that 50% of pts in the low risk group were in the mild glaucoma state, and the remainder in the ocular hyper tension (OH) state; and that 50% of pts in the high risk group were in the mild glaucoma state, with the remainder in the moderate glaucoma state. The model structure was based around 3 triggers for switching treatment. The choice of next treatment (2nd or 3rd-line) adhered to the following rational: if PA was not tolerated then switch to treatment from another class; if intraocular pressure (IOP) response below treatment target, or progression observed then add treatment from a different class. 4 levels of glaucoma severity were used as discrete health states, which corresponded to clinically significant thresholds within the model (changes in pt's VF leading to change in visual symptoms). Utilities were applied to the 4 glaucoma states. Treatment switching was implemented using a Markov sub-tree. For mild glaucoma, the tree started at 'pt follow-up', which initially branched to good or poor tolerance (T1), good tolerance then branched to IOP meeting target or not (T2), and then all branches (poor tolerance; IOP meeting target; and IOP not meeting target) led to the chance of progression in VF defect (T3) or not. The terminal nodes represented the associated 'no change in treatment' or 'next treatment' option, the associated mild or moderate glaucoma sub-tree, and the follow-up plan (the tree would then start again at follow-up). **Implementation:** Software used to develop model was not stated.
Sawyer (*Sawyer, 2013*) CUA	1 yr (4 wks)	The comparison of fixed treatment sequences of up to 3-lines, representing initial topical treatments used in primary care prior to referral for more intensive treatments (secondary care). Sequences included 8 different topical therapies. A number of restrictions were used on some sequencing to ensure	***Psoriasis*** Pts with psoriasis for whom topical therapy is expected to be practical, effective and safe in the long run. (Psoriasis is a relapsing remitting condition, where response to treatment has no effect on natural history.)	9 health states: first-line responder; second-line responder; third-line responder; first-line non-responder; second-line non-responder; third-line non-responder; and a relapse state linked to each treatment responder state. Events: respond, not respond to treatment, or relapse. Clinical effectiveness (treatment response) based on proportion of patients who were clear	*The model structure was divided into two parts. The first part represented the use of sequential treatments used in primary care, whilst the second captured the consequences following the failure of topical therapies, and referral for more intensive treatments. Only the first part is considered here, as treatment sequencing were not considered in the second part.* The model structure was designed to represent the movement of patients through a fixed series of up to three topical treatments. Only the responder states were recursive. Pts either responded or did not respond to treatment. Responders stopped treatment and either maintained response or relapsed. Patients who relapsed resumed initial treatment,

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
		safe and logical use of treatments; 118 sequences were evaluated for trunk and limb psoriasis, and 169 for psoriasis of the scalp.	The study evaluates topical therapy, and considers treatment of the trunk and limbs, and psoriasis of the scalp, in separate analyses.	or nearly clear. All treatments were assumed to have the same relapse rate (average 4-week risk of 35.5%), which could occur at any point following response.	which they responded or did not respond to. Relapse was depicted as a temporary state which patients passed through, returning to the initial treatment. Non-response was also a temporary state linking the current treatment with the next one. Patients who did not respond after up to 8 weeks were assumed to return to their general practitioner (GP) for an alternative topical treatment. Patients tried up to three topical treatments before being referred for specialist review where second-line (intensive) treatment options were considered. **Implementation:** Software used to develop model was not stated.
Shepherd. model (*Shepherd, 2006*) CUA	Lifetime (1yr)	Compared 6 fixed sequences of antiviral drugs. All included 2-lines of active treatment, and 2 sequences also included ant-viral salvage treatment. *[The main aim of the study was to evaluate 2 new (single) treatments (pegylated-IFNα(PEG) and adefovir dipivoxil (ADV)), compared with existing treatments (standard IFNα or lamivudine LAM, respectively), or non-drug treatment strategy (BSC), but cost-effectiveness of sequential treatment scenarios were also modelled.]*	***Chronic hepatitis B (CHB)*** The cost effectiveness of treatments were considered separately for pts with hepatitis B e antigen (HBeAg) positive and HBeAg negative, but both disease variants were included in the same model. Although the structural assumptions of the model were equally applicable to both groups, they differed in terms of the distributions of age at diagnosis and transition probabilities between health states, and therefore need to be kept separate in the analysis. *(Tunnels were used to determine whether individuals had HBeAg-positive or -negative disease.)*	A natural history model for CHB was developed, which indicted that, pts with CHB (not receiving anti-viral treatments) may remain in that state; move on to more progressive stages of liver disease (such as cirrhosis or hepatocellular carcinoma); or may clear the disease spontaneously / move into remission (with normalisation of ALT and low serum DNA). Pts clear the disease, either through HBeAg seroconversion to what has been traditionally termed as the 'inactive carrier' state, or through HBsAg seroconversion, where the pt is effectively cured. 8 health states: - CHB; - HBeAg seroconversion remission (*HBeAg +ve pts)* / remission (*HBeAg –ve pts)*; - HBsAg seroconversion; - Compensated cirrhosis (CC); - Decompensated cirrhosis (DC); - Hepatocellular carcinoma (HCC); - Liver transplant (LT); - Death. *The starting state was CHB, in which pts would present for antiviral treatment. All states, other than HBsAg seroconversion ('cure') were linked to death, and all states, other than 'cure' and mortality, were recursive states.*	*Treatment sequences were implemented using tunnel states, which were used to show whether the pt was HBeAG-positive or negative, whether they were resistant to the 1st or 2nd drug, and whether they were continuing or had stopped treatment.* The Markov cycle tree included two subtrees (clones) that were attached to different locations, or nodes, in the tree. *The advantage of using cloned subtrees was that only one 'master' copy needed to be maintained rather than requiring maintenance of numerical identical trees.* The 'Progression' subtree indicated all the possible states that an individual could progress to in the next cycle. 'PreResitance' subtree showed the different management options for individuals who develop resistance, and indicated whether the pt would continue, stop, or, if other antiviral agents were available, switch treatment after experiencing treatment resistance. The 'pre-resistance' subtree was attached to each of the health states in which pts were eligible to receive antiviral treatment, including the disease progression states CC, DC, HCC, and LT. The 'progression' subtree was attached to the seroconversion states, and 'Pre-Resistance' to the remission state. Pts who did not develop treatment resistance during a cycle followed a branch called 'NoResist' and had outcomes evaluated as descried in the progressive sub-tree. Pts who developed resistance followed the pre-resistance subtree. Each terminal branch of the 'pre-resistant' subtree had a 'progress' subtree attached to it. During the 'progression' subtree, pts 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, disease progression (CC, DC, HCC, and LT), and state specific excess mortality risk. All the terminal branches, or destination states, except death were tunnel variables. Each disease state consisted of up to 12 tunnel, or temporary, states, which took into account previous treatment history (and disease variant). Developing resistance to a 2nd treatment was independent of the fact that the pt had already developed resistance to the 1st treatment. Not all of the destination states were accessible from each starting state. For example, individuals with CHB 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 transition probability for any non-allowable transition was set to zero within the tree.

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
					For the best supportive care (BSC) comparator, no antiviral drug treatment was modelled (transition probabilities based on natural history). For the evaluation of antiviral drug therapies, the natural history transition probabilities were modified to take into account the treatment effect of each treatment. **Implementation: Implementation:** The model was developed using TreeAge Pro 2009.
Smith model (*Smith, 2007*) CUA	Lifetime (4 wks)	The comparison of 25 different treatment sequences based on the use of 6 drugs.	***Postherpetic neuralgia (PHN)*** 70 year old pts with established PHN. (PHN is unpredictable in terms of severity and duration.) Two base case analyses were considered: patients with and without coronary artery disease (CAD), to account for the differences in tricyclic use between these groups.	4 health states: PHN; severe treatment adverse effects (AEs); No PHN; and dead. *Treatment was used for symptom relief, not PHN resolution.* Events: pts experienced adequate or inadequate pain relief (inadequate pain relief was based on <50% relief); severe AEs; and resolution of PHN (based on natural history).	*Treatment for PHN, when effective was assumed to decrease PHN symptoms, but not PHN duration.* Various sequences were evaluated using an identical hypothetical patient cohort starting in the PHN state. Treated pts in the PHN state had diminishing symptoms compared with untreated pts, based on the likelihood and magnitude of treatment response. Pts who experienced inadequate pain relief or intolerance were switched to the next treatment in the sequence during the proceeding cycle. Pts treated with a given medication could transit to the severe side effects state, based on the medication's AEs likelihood. Pts in the PHN or sever AEs states could transit to the no-PHN state, based on the natural history of PHN; pts in all states could transit to the dead state (no pts die as a result of treatment). *Assumptions:* PHN has a median duration of 90 days, with 25% of cases lasting > 6mths. Beneficial effects of treatment and AEs occur within the first month of a given treatment. When a beneficial treatment is found it would have a sustained effect, and continued until PHN resolved or death ensued. Failure to respond to one treatment would have no effect on likelihood to respond to others. Pts with >50% pain relief had a 50% improvement in PHN disutility. Pts with <50% relief had no change in PHN utility. Pts who experienced severe AEs had no improvement in PHN utility and an overall decrease in utility due to side effects. **Implementation:** The model was constructed using Tree-Age Pro (Tree-Age Software).
Soini model (*Soini, 2012*) CUA	Lifetime (4 wks)	Comparison of 4 fixed immunochemotherapy sequences: rituximab (RTX) vs observation as 1st-line maintenance treatment, followed by bendamustine vs chemotherapy (CTX) as 2nd-line induction for pts who have progressed within 6mths (pts who progressed after 6mths received RTX+CTX).	***Follicular non-Hodgkin lymphoma (FL)*** Pts with grade I-III FL, with complete or partial response to 1st-line induction treatment with RTX plus CTX induction (assumed to correspond with the pt population of an RCT of RTX maintenance treatment) (*FL is a long-term disease with overall survival exceeding the time frame of most trials*)	4 health states reflecting the disease status of the pts: progression free first-line treatment (PF1); progression free second-line treatment (PF2); progressive disease (PD); and death. All 4 health states were recursive, and all transitions were forward. Each state was linked to death, which was an absorbing state.	Model structure was aligned with clinical objective of placing pts into PF state for the longest period possible. Patients entered the model in a disease-free state (PF1) having successfully completed first-line induction therapy with rituximab plus chemotherapy (RTX+CHOP). Those who responded to the induction treatment either received or did not receive first-line maintenance treatment (the continuation of rituximab for 2 years), depending on the treatment sequence being modelled. After the initial treatment, at the end of each cycle patients could remain in PF1, progress to PF2, which corresponded to response to a second induction treatment, or die. Once in the PF2 health state, a patient may remain in this health state, die at the end of each cycle or move to progressive disease (PD). Patients in the progressive disease state may either remain in the same state or die at the end of each cycle. Death was an absorbing health state. In the case of PF1 failure patients received a second-line induction, which included one of two treatments depending on the sequence being modelled. In the case of PF2 failure, pts were expected to receive best supportive care (BSC).

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
					Implementation: The model was described as a 'spreadsheet-based probabilistic Markov model with a 1-mth cycle length and half-cycle correction' (actual software used was not stated).
Tebas model (*Tebas, 2001*) Not applicable (virological outcomes)	10 yrs (NS)	Evaluating the timing of initiating antiviral therapy (AVT) using a sequence of up to three possible regimens (before the development of multidrug-resistant (MDR)virus). The model essentially included one sequence with variation in timing of starting the initial treatments.	***HIV*** Hypothetical population included HIV-infected pts in which therapy was started immediately vs progressively at a rate of 5, 10. 15, 20 or 30% of the original population each year.	The main outcome was proportion of pts who were: undetectable, had developed MDR, and never required treatment, at the end of 10 yrs. 7 health states: naïve; undetectable first regimen; detectable first regime; undetectable second regimen; detectable second regimen; undetectable third regimen; multidrug resistant (MDR). ('Naïve' was an initial state, each undetectable states were recursive, each detectable states were temporary, and MDR was an absorbing state.)	This was a population model based on virologic, rather than clinical, outcomes (and did not investigate cost-effectiveness). Sequences were compared in terms of the time spent undetectable Initially each pt was treatment 'naïve' (initial state). After initiating AVT, HIV-RNA could become undetectable (with a probability based on the success rate of the regimen), or the regime could fail, with a complimentary probability. If the initial regimen succeeded, in the next cycle the pt could remain undetectable, or fail, and so on. Each of these changes had a probability that could change over time. If the pts failed 2 consecutive rescue regimens they were considered MDR, which was represented as an absorbing state. Only the undetectable states were recursive, each detectable states were temporary. **Implementation:** Software used to develop model was not stated.
Wong model (*Wong, 2009*) CEA	Lifetime (1 wk)	Comparison of 9 treatment strategies of chemotherapy and/or monoclonal antibodies (8 treatments). 7 strategies included (fixed) treatment sequences, which included 2-3 active treatment lines.	***Metastatic colorectal cancer (CRC)*** Newly diagnosed pts with metastatic CRC (median age at diagnosis is aprox 70 yrs).	5 health states: enter into model; first line treatment; toxicity; progression; change treatment to second line or best supportive care (BSC). Pts could also die in any health state. Events: continue treatment, develop toxicity, or die The CEA was based on discounted life-year gained.	*The main effectiveness estimates were the rates of progression and toxicity, and median time on each treatment. (Did not use overall survival from clinical trials.)* In the 'stylized Markov model', 'first-line treatment' was depicted as a recursive state, which also branched to 'toxicity' and 'progression'. Toxicity branched back into 'first-line treatment', or to 'change therapy'. 'Progression' appeared to be a temporary state, lying between 'first-line' and 'change therapy'. At the end of each cycle, pts could either remain on therapy at stable doses, develop toxicity, or die (from all-cause mortality). Pts who develop toxicity could die, continue therapy at a dose reduction (transition back to 1st-line), or change therapy. Pts could have up to 2 toxic events before discontinuing treatment. Pts could receive up to 3 lines of treatment. It was assumed that toxicity and progression were independent and mutually exclusive events over the course of a 1 week cycle. Progression rates from clinical trials were converted to weekly probabilities using the declining exponential approximation of life expectancy (DEALE). Grade 3 or 4 toxicity (with FOLFOX and FOLIRI regimens) were converted to probabilities using DEALE and fractioned into fatal and nonfatal outcomes. The probability of a 2nd toxic event was assumed to be 10%. Progression free survival (PFS) data used to estimate time on each treatment. **Implementation:** The model was implemented using 1000 hypothetical pts. The model was developed using TreeAge Pro.
NICE CG137 model (*NICE CG137, 2012*)	15 yrs (6 mths)	Comparing fixed treatment sequences of antiepileptic drugs (AED), which included 1st-line monotherapy 2nd-line	***Epilepsy*** Adults with focal epilepsy. An integrative model was used for the comparison of monotherapies for	12 health states: 4 treatment response outcomes associated with monotherapy for newly diagnosed pts (seizure freedom, partial seizure freedom;	The model was described as a multistate Markov model, built to reflect transitions between a set of mutually exclusive health states, defined by the outcomes of treatment.

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
Model structure based on the York epilepsy model (a semi-Markov model) CUA		monotherapy, and 3rd-line adjunctive therapy. For the comparison of different monotherapies for newly diagnosed patients, the same 2nd- and 3rd-line treatments were used for each sequence. For the analyses of adjunctive therapy for refractory epilepsy, assumptions were made about the prior treatments used.	both newly diagnosed, and adjuvant therapies for refractory epilepsy, which were analysed separately. (Hypothetical cohort assumed to be 30 yrs old for evaluation of monotherapies).	inadequate seizure control, unacceptable adverse effects); 3 response outcomes associated with monotherapy for refractory pts (seizure freedom, partial seizure freedom; treatment failure); 3 response outcomes associated with adjuvant therapy (seizure freedom, partial seizure freedom; treatment failure); maintenance therapy; and death.	Movement between the various health states was governed by transition probabilities derived from the SR of clinical effectiveness and from observational and open-label clinical trial data. It was assumed that all hypothetical patients entering the model were newly diagnosed, treatment-naive individuals with focal seizures. All patients started with monotherapy and experienced one of four outcomes: remission, or seizure freedom; a reduction in seizure frequency; or treatment failure due to either unacceptable adverse events or inadequate seizure control. In the base case, the model assumed that hypothetical patients who failed 1st-line moved to 2nd line. The use of the same drug (carbamazepine) was assumed as 2nd line across all modelled treatment arms (1st-line comparators) [sensitivity analysis conducted using another drug, lamotrigine]. Patients failing 1st-line due to inadequate seizure control were assumed to be 75% less likely (risk ratio 0.25) to achieve remission with 2nd line monotherapy (informed by observational study) [one way sensitivity analysis conducted varying this figure to 0.5, 0.75, and 1]. For patients who failed 1st-line due to intolerable side effects it was assumed that response to the second line monotherapy was independent of response to 1st line AED. Patients who had not achieved remission, but also not classified as having failed treatment, were assumed to persist with 1st line monotherapy for 2 years, at which point the patient was classified as having failed due to inadequate seizure control and moved on to 2nd line treatment [sensitivity analysis conducted using 18 and 36 months trial period before switching]. Patients who failed treatment with a second monotherapy were assumed to move on to adjunctive therapy. For the comparison of monotherapy in the treatment of newly diagnosed epilepsy a single AED (tiagabine) was chosen to be a common adjunct therapy across all comparators. Patients who started adjunctive therapy experienced one of four outcomes: seizure freedom; a reduction in seizure frequency of between 50% and 99%; a reduction of less than 50% (no response); or withdrew due to adverse events. Patients who were refractory (non-responders and those withdrawing due to adverse events) were assumed to be maintained on monotherapy with an older AED ('maintenance therapy'). A network meta-analysis of IPD from RCTs (including SAND trial) was used to inform treatment failure and remission of monotherapies in the short term (up to 3 yrs) and a meta-analysis of placebo controlled trials for adjuvant therapies (up to 6 mths). The probability that a pt remained on that treatment in subsequent cycles was based on observational or open-label clinical trial data that was not specific to the drug being considered. The uncertainty in the probability of treatment failure (for the PSA) were incorporated using beta distribution. The model accounted for epilepsy related mortality linked to whether the pts were seizure-free or not seizure-free. The Markov model was created in TreeAge pro 2008.
Simple Markov cohort used for identifying optimal sequences					

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
York psoriasis model (*Woolacott 2006; Anis, 2011; Sizto, 2009*) CUA	10 yrs (1 yr)	Seeking to identify the optimum overall ordering of treatments (n=6) (Base-case analysis based on 2 TNF-inhibitors)	***Psoriasis*** Pts with moderate to severe chronic plaque psoriasis.	2 health states: patient on treatment, and patient on supportive care. Primary outcome was pts achieving a 75% reduction in the Psoriasis Area and Severity Index score (PASI75).	Treatment sequences evaluated by developing estimates of the expected treatment period net benefit (NB) for each treatment, which were then used to estimate the optimum treatment sequence. Treatment period NB is the expected costs and health effects per unit of time for each individual treatment incurred during the entire period a patient received that treatment. It includes the weighted average of the expected net benefit incurred over the treatment 'lifetime' for patients who responded and continued treatment after the initial 'trial' period, plus the expected net benefit over the treatment 'trial' period for those who did not respond to treatment. The expected treatment period NB for each treatment was developed using a simple Markov cohort model. Patients started 'on treatment' and transitioned to 'supportive care' informed by probability of patient failing treatment, and transitioned back into the same state informed by 1-probability of failing treatment. The data for the model on duration of treatment 'trial' periods were based on the 12 weeks follow-up used in the efficacy trials and expert opinion, and the mean 'treatment' periods (annual dropout rates for responding pts and a maximum assumed treatment period) on derived from expert opinion and assumptions made on limited observational data. The model assumed no difference between treatments in terms of mortality. **Implementation:** A comprehensive decision model was conducted in WinBUGs. All decision modelling was undertaken in the programming language R.
Semi-Markov model					
York epilepsy model (*Wilby, 2005*) CUA	15 yrs (6 mths)	Comparing fixed treatment sequences of antiepileptic drugs (AED), which included 1st-line monotherapy 2nd-line monotherapy, and 3rd-line adjunctive therapy. For the comparison of different monotherapies for newly diagnosed patients, the same 2nd- and 3rd-line treatments were used for each sequence. For the analyses of adjunctive therapy for refractory epilepsy, assumptions were made about the prior treatments used.	***Epilepsy*** Adult pts with epilepsy. An integrative model was developed for the comparison of AEDs used for both newly diagnosed, and refractory epilepsy, which were analysed separately. Partial seizure-type patients and generalised-seizure type pts were analysed separately. Some of the clinical trials in the network meta-analysis of clinical effectiveness of AEDs for newly diagnosed pts included a mixture of generalised and focal epilepsy types.	8 health states: start monotherapy for newly diagnosed pts; continue monotherapy for newly diagnosed pts; start monotherapy for refractory pts; continue monotherapy for refractory pts; start combination therapy; continue combination therapy; maintenance therapy; and death. The available evidence indicated that the probability of treatment failure reduced with increased time on the treatment.	Treatment sequences were implemented as a series of health states, which also differentiated between starting and continuing each line of treatment (for 3 active treatments). Depending on the initial state, pts could move through the first seven states in sequence. Pts could move from any state to the state death. Pts only spent one cycle in the three active starting treatments, and the initial distribution within these states were established within the first cycle. Only pts who achieved response to treatment during the clinical trial went on continue treatment after the first cycle (i.e. moved to the corresponding continue state), with those who failed treatment switching to the next starting active treatment state. The probability of a patient making a given transition during a model cycle was based on the time spent in the current state, which was modelled externally (hence why the model is a semi-Markov process model). Treatment response was defined as achieving seizure freedom and reaming on drug until the end of the trial for monotherapy for newly diagnosed pts, and achieving 50% seizure freedom compared with baseline and remaining on the study drug for monotherapy for refractory pts or combination therapy. It was assumed that treatment response did not vary according to positioning in the sequence. The probability of remaining on a treatment (probability of response) during the first cycle was estimated from clinical trial data (duration generally 6 mths) specific to the drug under consideration (taken from a NMA). The probability of not remaining on treatment (probability of failure) for each drug, for subsequent cycles, was based on observational data, which was not specific to the drug under consideration. The uncertainty in the probability of treatment

Model name (*studies using model*) **Economic evaluation type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Health states and events**	**How sequences modelled**
					failure (for the PSA) were incorporated using beta distribution. The model accounted for epilepsy related mortality linked to whether the pts were seizure-free or not seizure-free. The model was implemented using R, a statistical programming language which has the ability to manipulate n-dimensional matrices. It was felt that an implementation of the model in Excel would have been extremely difficult and hard to audit.

Abbreviations: AE adverse event or effect; BSC best supportive care; CEA cost-effectiveness analysis; CUA cost-utility analysis; IPD individual patient data; mths, months; NMA network meta-analysis; OS, overall survival; PFS progression-free survival; pts patients; QALY quality adjusted life year; RCT randomised controlled trial; SR systematic review; TNF Tumour necrosis factor inhibitor; TTP time to progression; yrs years.

APPENDIX E2: INDIVIDUAL SAMPLING MODELS

Table E2.1: State transition individual patient simulation models in rheumatology studies *(the key models are highlighted in bold)*

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
Sheffield Model – ETA (*Brennan, 2004)* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	cDMARD sequence with and without ETA, added to the start. *(A subsidiary review question also considered sequential TNF-inhibitors)*	Pts with IR to at least 2 cDMARDs including MTX and SSZ (pts' baseline characteristics based on RCT of ETA, for which IPD was available)	ACR20 HAQ	The model was based on fixed treatment sequences, with a sequence of 3 cDMARDs (A-B-C) implemented as an exemplar sequence: gold, LEF and cyclosporine +MTX. Pts entering the model either started with ETA or DMARD A. After the first period a patient may be an 'initial responder' or 'non-responder'. Non-responders switched to the next treatment, and subsequently assessed for initial response. Initial responders remained on treatment for several 6-mth cycles until subsequent longer-term withdrawal owing to loss of efficacy or adverse effects. Pts could also die during any cycle (related to age, sex, and HAQ score). Clinical guidelines suggest withdrawal of TNF at 3mths if no response. For ETA, the model explicitly examined the percentage of withdrawal at 3mths and between 4-6 mths (based on patient level data from ETA trial). **Implementation:** The model was developed in Microsoft Excel.	3 states: treatment responder (ACR20); non-responder; death The model focused on the progression on HAQ disability score for the population over time. The pt's HAQ score and mortality were evaluated at the end of each cycle. Pts who initially responded, and thus remaining on treatment, were re-assessed for sustained efficacy or toxicity at the end of each cycle.
Bansback model *(Bansback 2005) [The mathematical approach used builds on both the Sheffield ETA model and BPM; the model structure was similar to Sheffield ETA model]* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	cDMARD sequence with and without a TNF, added to the start	Pts with inadequate response to at least 2 cDMARDs (pts' baseline characteristics based on clinical trials of TNFs (IPD available from RCTs of ADA)	ACR (20/50/70) HAQ	The model structure was similar to that of the Sheffield ETA model, but considered AEs as a separate state. The sequence of 3 cDMARDs (A-B-C) was implemented using a generic 'DMARD'. Pts entering the model either started with a TNF or DMARD A. At the end of each cycle both 'non-responders' and those who experienced 'severe AEs' withdrew from their current treatment and moved to the next one; whilst 'responders' continued on the same treatment. Pts who initially responded to treatment were re-assessed for sustained efficacy and toxicity at the end of each cycle. Pts classified as 'treatment success' or 'non-responders' were also assessed for the occurrence of mild to moderate AEs, as the model also incorporated the treatment of AEs (mild, moderate, or serious). **Implementation:** The model followed 10000 hypothetical pts (Software used not stated).	4 states: treatment response (based on ACR20 and ACR50 threshold; modelled separately)[2]; non-response; serious AEs; death. In clinical practice decisions to continue treatment based on DAS28. *The model assumed that ACR20 response corresponded to a moderate DAS28 response, and that ACR50 corresponded to a good DAS28 response.* *Model followed HAQ progression through treatment sequence. Initial reduction* in HAQ for each level of response based on IPD from ADA trial. (*HRQoL scores evaluated by simple linear transformation from HAQ score)* Response rates for 'DMARD' based on an observational study of ETA, INF, and LEF, in which pts on LEF had failed on average 4

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
							cDMARDs. It was assumed that all DMRADs had the same response rate as LEF, and that this was influenced negatively by disease duration (OR of 0.98 for each extra year of disease duration).
(Davies 2009) *[Model reported to be based on the same structure as Bansback model]* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	Replacing the first treatment (MTX) in a sequence of 4 cDMARD with a TNF (n=3). 5 alternative sequences were modelled, including a reference sequence without TNFs, 3 with a single TNF, and 1 with two TNFs. *(The sequence with a 2nd TNF, followed by 2 cDMARDs included as a supplementary analysis*)	Pts with early RA (pts' baseline characteristics based on based on an RCT of ADA with pts who were MTX-naïve).	ACR (ACR0-20, ACR20-50, ACR70-100) HAQ	*The model structure was the same as Bansback, but did not consider AEs as a separate state.* For modelling purposes, a maximum of 3 effective cDMARDs were assumed to follow the TNF in each treatment sequence (MTX+HCQ - LEF - gold). Pts were randomly simulated to experience several alternative sequences. Pts who did not achieve ACR50 at the end of the first cycle passed immediately to the next treatment, and subsequent withdrawal (due to AE or inefficacy) determined at 6-mth intervals. Pts who withdrew due to AEs were excluded from receiving further TNF-inhibitors. **Implementation:** 1000 pts were randomly simulated to experience the different treatment sequences (Software used not stated).	3 states: continue treatment (based on ACR50 threshold); non response (inefficacy or AEs); death. Events: Initial ACR response categorised into 4 intervals: ACR0-20, ACR20-50, and ACR70-100. Each level of response was associate with a given reduction (improvement) in the pt's HAQ score from baseline (HAQ change). Pts passed through sequences treatment, until death or the last DMARD failed. Individual pt outcomes (costs and QALYs) were sampled at 6-mth intervals. Treatment specific costs included AEs.
Sheffield Model – AHRQ (*Wailoo, 2006*) [*Model based on both the Bansback and Sheffield ETA models]* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	Initial analysis compared 4 biologics, followed by cDMARDs. Subsequent analysis included treatment sequencing (for 3 TNFs only), which evaluated using a 2nd or 3rd TNF compared with a single TNF (INF) followed by cDMARDs.	Pts treated with a biologic, and for whom treatment with a biologic had not previously failed. Baseline characteristics were based on an average National Data Bank for Rheumatic Diseases (NDB) Medicare population. (IPD available) The number of pts who had not received a biologic agent when registered on the NDB, but subsequently started ETA, INF, ANA, and ADA included 1,490, 1,403, 74, and 160, respectively.	ACR (ACR0-20, ACR20-50, ACR70-100) HAQ	The model structure was based on tracking the pt's HAQ over time, from starting biologic treatment until death. The model started by developing a representative sample of 10,000 pts. A series of regression analyses were then used to estimate the parameters used for simulating the path each individual pt would take. Separate statistical models were used to estimate: i) Type of HAQ responder (achieving <20%, 20-50%, or >50% improvement on TNF); ii) HAQ score archived at 6 mths (on TNF); iii) HAQ score at 6-month intervals (on TNF); iv) Withdrawal from TNF-inhibitor (treatment duration for TNF); v) HAQ at 6-month intervals (on cDAMRDs); vi) QALY as a function of HAQ over course of treatment strategy; vii) other costs as a function of HAQ.	Individual pts were tracked from the time of starting a TNF-inhibitor until death, with changes in important variables (including HAQ) calculated every 6 mths. The model first estimates the probability of a patient achieving: <20% improvement in HAQ, 20-50% improvement, >50%. Events: starting biologic treatment, withdrawing from treatment, death.

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
						On withdrawal from the 1st TNF, pts moved on to the next TNF, and statistical model iv was followed by i, ii, and iii for the next TNF, until the final TNF in the sequence, at which point HAQ progression was estimated using analysis v. Implementation: The model was developed in Microsoft Excel. Treatment strategies were compared by using the same 10,000 pts.	
(Finckh, 2009) *[Model based on Sheffield AHRQ Model]* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	Compared three management strategies: 'pyramid approach' starting with pain management then cDAMRDs then TNF-inhibitors; early initiation of cDAMRD, followed by TNF-inhibitors; or early initiation of biological treatment, followed by cDMARDs. *(The 'TNF-inhibitors' and 'cDMARDs', included a sequence of 3 treatments*)	Pts with very early RA (symptoms for <3 mths). Baseline characteristics were representative of the US population with RA, and based on the National Data Bank for Rheumatic Diseases (NDB) and assumptions.	Response: excellent, good, moderate, or none HAQ	*Analysis based on modified version of model presented by Wailoo, 2006.* Hypothetical cohorts were tracked through the model in 6-mth cycles from symptom onset until death. The exact route a simulated pt took depended on the treatment strategy used and the pt's disease characteristics (including, among others, type of disease progression, disease duration, HAQ, and no. of previous cDMARDs). It was assumed that pts follow one of 3 disease courses, which cannot be predicted at presentation. The course of the disease was modelled using both HAQ and radiographic evidence of structural damage. Initial HAQ improvement (at 6 mths) depended on treatment used, response, disease duration and radiographic damage. When pts withdrew form treatment, they switched to the next available treatment. This process was repeated for the patient's lifetime. The model synthesised data from RCTs, NDB a longitudinal database of patient data, and other literature. Implementation: R project statistical software was used for all decision analyses.	Events: toxicities, treatment initiation, treatment discontinuation, or death. Model schematic depicted: 3 treatment strategies: pyramid approach, early cDMARDs, early TNF-inhibitors. 3 disease courses: spontaneous (drug free) remission, slow progression, rapid progression 5 disease states: excellent response, good response, moderate response, no response, death 5 outcomes: HAQ score, eroded joints, QoL, cost, death. Excellent response based on definitions in literature of remission in RA. Moderate/good response based on ACR20 and ACR50. Initial HAQ improvement based on type of response (moderate, good, remission) adjusted for a number of patient characteristics including baseline HAQ.
Sheffield Model – BSRBR (*Brennan, 2007*) CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	cDMARD sequence with and without a TNF (evaluated as a class), added to the start. (*Also included supplementary analysis of sequential TNFs. As there was no evidence of any correlation, the response to the 2nd TNF was assumed to be independent of the 1st*)	The evaluation was based on the pt population in a registry - British Society for Rheumatology Biologics Register (BSRBR) of pts starting TNFs (IPD available) BSRBR included 3-yr FU on nearly 8000 RA pts with active disease treated	EULAR (DAS28) HAQ [EULAR response: none (DAS=0), moderate	The model ran the same pt through two arms, ie. TNF vs cDMARD. Rather than specifying a particular cDMARD at different positions in the sequence, a generalised DMARD was used in each position based on a weighted average of BSRBR pts' DMARD use. After receiving a new treatment (3 or 6 mths), EULAR response was assessed. Non-responders switched treatment and responders continued treatment until	2 states: treatment responder (EULAR: moderate or good responder); non-responder. Individual pt event histories were tracked over 6-mthly intervals. The following key events were tracked: i) initial response to treatment (in terms of EULAR response: non/moderate/good) and switching treatment;

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
				with TNFs and nearly 900 RA pts with active disease treated with cDRMARDs.	(DAS=1), or good (DAS=2)]	relapse or AE occurred. When a pt reached time to withdraw, because of either AEs or lack of response, the model moved to the 2nd treatment in the sequence, then the 3rd, and so on. The probabilities of EULAR response, magnitude of improvement and time to withdrawal were all re-adjusted to the individual pt's updated characterises. After the 6th treatment, the pt was assumed to no longer respond but would still receive maintenance treatment on cDMARD. **Implementation:** The model was developed in Microsoft Excel.	ii) Impact of initial treatment response on short-term health utility; iii) length of longer-term treatment if treatment continued; iv) impact of longer term treatment on long term utility v) utility worsening when treatment withdrawal. *Utility at 0-6 mths was measured as a function of EULAR response, treatment, and pt characterises, and utility post 6 mths was measured as a function of treatment and pt characteristics*. **Alternative shorter version** *The model tracked a large number of individual pt's event histories over 6-mthly intervals, including initial treatment response (EULAR moderate or good response) or non-response (resulting in treatment switching), and length of longer-term treatment if therapy was continued*.
Diamantpoulus model *(Diamantpoulus, 2012; Diamantpoulus, 2014)* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	Adding TOC to a sequence of bDMARD (at the start), and comparing sequences of bDMARDs (replacing 1st drug with TOC) (*TOC added to a sequence of bDMARDs as the 1st or 2nd drug 2014 version; 2 separate baseline sequences were used representing standard treatment for different pt populations: those who were tolerant to MTX, and those who were MTX contraindicated*)	Pts with inadequate response to cDMARDs; data on baseline characteristics based on RCTs of TOC. (*Characteristics of model cohort based on British Society for Rheumatology Biologics Register (BSRBR) data in the later (2014) version of the model. This also included evaluation of 2 scenarios: MTX contraindicated population and MTX tolerant*) The model assumed a homogenous group of pts at the start (average estimates for each pt characteristic used, not distributions).	ACR HAQ (and VAS pain in later model)	Pts entering the model progress through a predefined sequence of bDMARDs. Pts were assumed to all drugs in the given strategy. On starting treatment, simulated pts were allocated to one of four ACR response categories: ACR00, ACR20, ACR50, ACR70. Those with no response (ACR00) moved to the next treatment in the sequence, whilst those with a response remained on treatment until withdrawal, with significant changes in QoL (reflected by changes in HAQ scores; *and VAS pain in the 2014 model*). Pts transitioned to death based on a mortality risk adjusted for RA. The proportion of pts achieving each level of response was treatment dependent, whilst the initial HAQ benefit was assumed to be response related (not treatment). The HAQ reductions, for each level of response, were therefore applied universally to all bDMARDs.	*States: 1st bDMARD; nth bDMARD; palliative treatment; death'* Pts allocated to 4 groups: No response (ACR00); ACR20 response; ACR50 response; ACR70 response. Disease severity was represented by changes in HAQ score (and VAS pain), which is a surrogate health outcome that can be translated to utility scores and ultimately QALYs. Simulation was used to monitor HAQ changes (VAS pain also included in a later model). Initial HAQ reduction was assumed to be response related not treatment related (applied universally across treatments). The data on the relationship between ACR response and HAQ score was based on the analysis of individual pt level data from 3 TOC RCTs, which showed that the higher the response the greater the drop in HAQ. On

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
						Implementation: 2012 - 10,000 simulations were run (software used not stated); 2014 - model conducted using Excel, with aggregate results were based on 3,000 simulations of hypothetical pt pathways.	treatment withdrawal, the initial HAQ gain was assumed to be lost (100% rebound effect). *[2012 version did not consider AEs; 2014 considered disutility from serious AEs (based on data from a Cochrane review), but treatment discontinuation was response related)]*
Kielhorn model (*Kielhorn, 2008 Hallinen, 2010; (Merkesdal, 2010)* CUA	Individual sampling (IPSTM)	Lifetime (6-mths)	Sequences of DMARDs with and without a RTX (added to the start) *(Hallinen: Sequences of cDMARDs with and without a bDMARD (added to the start), Further treatments (1-2 bDMARDs) were then added to the most cost-effective strategy in a stepwise manner)*	Pts with inadequate response to their first TNF (pts' baseline characteristics matched those of pts in the pivotal RCT of RTX) *(Hallinen: baseline characteristics of pts based on a published study of the use of biologics in Finland, and baseline HAQ from RCT of RTX*)	*ACR* *HAQ*	On entering the model pts were allocated to one of two treatment sequences. At the end of each cycle pts could make a transition to the next treatment (health state) or death. On starting treatment, the response status of the pt was evaluated and non-responders were switched to the next treatment. Whilst the responders' were allocated to one of three ACR response categories (ACR20/ACR50/ACR70), after which they continued the same treatment for a predetermined time period (which was treatment dependent). After this time period, the pts were assumed to relapse and switch to the next treatment in the sequence, ending with palliative treatment (until death). **Implementation:** The model was designed in Microsoft Excel, and run using 10,000 hypothetical pts. (*Hallinen - evaluation performed using cohorts of 3000 identical pts*)	States: series of active treatments *(3-6 for Kielhorn)*; palliative care; death. (*Merkesdal: on starting treatment (ACR response, treatment failure); respond; palliative care; death)* RA progression modelled as HAQ-deterioration. Initial HAQ reduction on treatment was assumed to be response related. Pts were allocated to 4 ACR response groups: ACR00, ACR20, ACR50, or ACR70 (based on efficacy rates for individual treatments). The corresponding drop in HAQ score for each response group was based on the RCT of RTX. While on treatment, pts HAQ scores were assumed to deteriorate by +0.017 during each cycle of the model. For pts on palliative care a higher rate of increase was assumed (+0.065). Once treatment was stopped, the initial HAQ gain was assumed to be lost (100% rebound effect), and the pt proceeded to the next treatment option.
Kobelt model *(Kobelt, 2011)* CUA	Individual sampling (IPSTM)	10-yrs (6-mths)	Compared treatment strategies starting with ETA or MTX. Pts discontinuing treatment could switch to their first TNF-inhibitor or a second TNF-inhibitor. Pts in remission received a reduced TNF dose.	Pts with early RA (pts' baseline characteristics matched those of pts in an RCT of ETA vs MTX) (IPD available)	DAS28 HAQ	On entering the model pts were allocated to one of two strategies. Data for pts switching to their 1st or 2nd TNF-inhibitor (depending on sequence) were obtained from the Southern Swedish Arthritis Treatment Group (SSATG) register (IPD available). Data on cDMARDs were taken from a separate observational study. Changes in disease status (HAQ level, high/low disease activity) or treatment were modelled as transitions between the states in 6 mth intervals	5 main states based on functional capacity [where HAQ scores 0 - 2.99 = worst]: HAQ 0 < 0.6); HAQ 0.6 < 1.1), state 3 (HAQ 1.1 < 1.6), state 4 (HAQ 1.6 < 2.1), state 5 (HAQ 2.1 - 3). Each state was further divided into high or low disease activity [DAS28 scores 0 - 10 = worst] where low disease activity was defined as DAS28 ≤ 3.2. In all resulting states, pts could be on a TNF-inhibitor (1st, 2nd, or half dose), MTX, or cDMARDs.

Model name (*related publications*) **Economic evaluation**	**Model type**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
						(cycles), implemented at the start of the next treatment. **Implementation**: The model was analysed as a microsimulation using 3000,000 simulations to obtain stable results (the software used was not stated).	*The model was based on the combined effect of function and disease activity to estimate costs and utilities.* *On discontinuation of treatment pts were assumed to return to baseline HAQ, adjusted for underlying progression during the years of treatment. Pts in remission were assumed to progress at half the rate of pts on the first TNF-inhibitor in SSATG (0.005/yr). Pts on cDAMRD were assumed to progress at an average annual rate of 0.031 HAQ points.*

Abbreviations: ACR20/50/70 American College of Rheumatology response criteria; ADA Adalimumab; AE adverse event or effect; CEA cost-effectiveness analysis; CUA cost-utility analysis; DAS28 Disease Activity Score 28 joints; DMARD/bDMARD/cDMARD disease-modifying anti-rheumatic drug/biological DMARD/conventional DMARD; ETA Etanercept; EULAR European League Against Rheumatism; HAQ Health Assessment Questionnaire; HRQoL Health related quality of life; INF Infliximab; IPSTM individual patient state transition model; LEF Leflunomide; MTX Methotrexate; mth month; OR odds ratio; pts patients; RA rheumatoid arthritis; RTX Rituximab; SSZ Sulfasalazine; TNF Tumour necrosis factor inhibitor; TOC tocilizumab; UC usual care; VAS visual analogue scale; yr year.

Table E2.2: Individual patient simulation models in non-rheumatology studies *(ordered by model type then alphabetically)*

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
Individual patient simulation state transition model (IPSTM)						
Holmes model (*Holmes, 2006*) CEA (life expectancy)	Lifetime (1 month)	Comparison of 5 antiretroviral therapy (ART) strategies, 2 of which included sequencing (2 lines).	**human immunodeficiency virus (HIV)** Women with prior exposure to s*ingle dose* nevirapine (NVP) for the prevention of mother-to-child transmission of HIV during pregnancy, and an initial CD4 cell count at or near the threshold for initiating ART. In the model women were stratified by presence or absence of NPV resistance.	viral suppression, life expectancy	The model was poorly reported. It was described as a 'simulation model' that consisted of clinically relevant health states and transition probabilities between health states. Women entering the model were assigned to 1 of the 5 strategies. The probability of having or not having NPV resistance was then determined. Pts clinical progression was tracked on a monthly basis. Women had a monthly probability of viral suppression, determined by the efficacy of each regimen (which in turn affected CD4 count). They also had a monthly probability of dying from HIV-related causes (based on CD4 count) or from non-HIV related causes (based on life tables). The table of model variables also included the monthly probability of experiencing an AE requiring treatment discontinuation.	The model pathway was depicted as a decision tree with a series of three chance nodes where probabilities were estimated of being in 2-3 states: i) NPV resistance; no NPV resistance. ii) Viral load suppressed; viral load not suppressed. iii) HIV death; non-HIV death; survive. Adherence to ART was assumed to be the same as that observed in the clinical trials from which efficacy data were derived. Sensitivity analysis included reducing the efficacy of second-line regimens. **Implementation:** The model was programmed in TreeAgro software.
Discrete event simulation (DES)						
Birmingham epilepsy model (*Connock, 2006*) CUA	15 years	Comparison of fixed drug sequences (of up to 4 lines) that contained exclusively 'older' anti-epilepsy drugs (AEDs) or a combination of 'older' and 'newer' AEDs.	**Epilepsy (*in children*)** Pts with newly diagnosed partial epilepsy. The model simulated children over their childhood, from the age at diagnosis (ranged from 3 to 18) through to 18 years.	Complete/partial seizure freedom; AEs.	Model was described as an 'individual sampling model', which was not based on a fixed cycles. On entering the model personal characteristics for the individual were assigned through a process of repeated sampling from appropriate distributions for the following characteristics: gender, age, presence of learning difficulties. Pts were initially prescribed a monotherapy, with the choice defined by the fixed sequence. Pts progressed through the drug sequence, with the rate being determined by the treatment outcomes experienced. Pts could experience one of 4 treatment outcomes (main model outcome states): i) intolerable AE (lading to early discontinuation); ii) lack of effect on seizure rate (leading to early discontinuation); iii) partial efficacy with tolerable or no AEs; iv) complete seizure freedom with tolerable or no AEs. Those who experienced outcomes 1 and 2 (early discontinuation) progressed to the next choice monotherapy, or opted to discontinue drug treatment. Those who entered outcome 3 could stay on current drug, try next choice monotherapy, try next choice add-on therapy (*it was assumed that the willingness to try an alternative treatment depended on the number tried at this point: as the number of drugs tried increased, the pt was more likely to try add-on therapy and less likely to try further monotherapy*), or discontinue treatment. Pts who achieved outcome 4 were assumed to withdraw from the drug after a given period (sampled in the model), or remain on current drug if reluctant to withdraw. It was	6 health states: the 4 main treatment outcome states plus successful drug withdrawal (pt has seizure freedom), and unsuccessful drug withdrawal (pt is not seizure free but prefers to remain untreated). Treatment sequences were compared in terms of the average time spent in each of the 4 main treatment outcomes. The time spent in outcome states were sampled from distributions for every pt. An assumption was made that longer durations in states with reasonable efficacy and side-effect profiles represent a positive outcome. The likelihood for a pt reaching a particular treatment outcome (*proportions achieving main model outcome states*) for each drug were based on a SR of RCTs, with the use of a reduction factor and assumptions to account for treatment sequencing effects. Data for other clinical parameters (*durations and proportions moving into secondary model states*), such as proportions discontinuing treatment, time to discontinuation or with drawl, and likelihood of moving on to add-on therapy were based on epidemiological studies and clinical advice. *A set proportion of pts was assumed to discontinue drug treatment and receive successful*

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
					assumed that the proportion discontinuing due to late toxicity or reduction in efficacy over time was negligible.	*surgery after achieving outcomes 1-3 after 1st-line treatment.*
Denis model (*Denis, 2008*) Not applicable (Clinical effectiveness)	5 years (1 month)	Comparison of two fixed sequences of intraocular pressure lowering agents. Patients were treated with 1st line prostaglandin (latanoprost or travoprost) followed by the addition of timolol (beta-blocker). The model was used to estimate probability of starting 3rd-line treatment.	**Glaucoma** Pts with open-angle glaucoma (OAG) or ocular hypertension (OHT). Controlling intra ocular pressure (IOP) is critical to prevent vision damage in pts with OAG and OHT. In most patients, a single agent is no longer sufficient after 2 yrs of treatment to control IOP, and a second agent is often added. The probability of a new visual field defect is known to increase with treatment changes.	Treatment failure and disease progression (visual field defect, VFD)	*The study was poorly reported* The model was described as a discrete event simulation. Glaucoma treatment and clinical outcomes were represented as a chronological sequence of events, with each event occurring at a specified time and denoted a change in the system. However, time was sequenced as a regular cycle of 1 month. Pts were randomly assigned to one of two treatment sequences using a random number generator (RNG). Virtual pts experienced different types of events (treatment failure and disease progression) using a RNG and according to risk functions that were estimated from either RCTs or surveys. Some risk functions estimates were specific to the prescribed treatment. In the model pts could experience events every month. When the model ended, the final patient status was recorded. The model was replicated for 5,000 pts.	2 types of clinical events: i) treatment failure (1st and 2nd line treatment), and ii) disease progression (up to 4 new VFDs) due to poor IOP control. Following 1st-line treatment failure pts received a 2nd-line add-on treatment. Treatment failure was defined as IOP ≥18mmHg at 2 visits. Time to treatment failure was estimated from 2 RCTs (one comparing monotherapies and one comparing combination therapies; prior treatment was not reported). The risk function for the probability of a new VFD was taken from an observational survey. 2 clinical events were excluded from the model due to low probabilities within 5 yrs: 3rd line treatment failure, and >4 VFDs **Implementation:** The DES model was developed with Excel software (Microsoft Corporation).
Heeg (schizophrenia) model *(Heeg, 2008)* CUA	5 years	Comparing atypical vs conventional antipsychotics (as a class) as 1st-line treatment, within a fixed sequence of 4 treatments. For each comparison, both treatment arms included the same 2nd-, 3rd-, and 4th-line treatments. The model was described as a nonproduct-specific DES, which represents the use of pharmacological agents in day-to-day clinical practice in the UK.	**Schizophrenia** Pts with chronic schizophrenia requiring antipsychotic treatment. Pts enter the model while suffering an episode for which the care of a psychiatrist is sought. It is assumed the patient is presenting early on in the course of the illness, but it is not the first episode of psychosis (as distinct from first episode of schizophrenia), because the diagnosis of chronic schizophrenia cannot be made based on a single psychosis. Therefore, patients may not be treatment naïve.	Positive and Negative Symptom (PANSS) Score; number and duration of psychotic relapses	Simulation started with selecting a number of fixed (time independent) attributes from a set of pre-specified probability distributions: pt profile (which determines whether a pt recovers fully (38%) or partially (62%) between relapses in terms of symptom score); severity of illness; social and environmental factors; and whether the pt will suffer from side-effects when put on a specific medication. Once the time-independent attributes were assigned to a pt, the model simulated disease progression based on a number of interdependent time-dependent variables. The two major time-dependent variables were the pt's health state at a certain moment (in relapse, or between relapses) and the result of reassessment of medication and treatment location during psychiatric visits. The treatment characterises (treatment and location) were determined during psychiatric visits, based on the interdependencies between the time-dependent and time-independent variables.	The variables included in the model were either fixed (pt characteristics/attributes) or time dependent. An example of a patient history from time of entering model during relapse at visit to psychiatrist was presented in Figure 2. This included additional time-independent and time-dependent variables not listed in the text. Other time-independent variables included: gender, age, pt type. Other time-dependent variables included: episode (relapse); psychiatric visits; compliance; line of treatment; PANSS; disability to take care of oneself; risk to self/others; treatment setting; quality of life. When the patient enters the model or switches treatment, an antipsychotic was selected based on UK market share data. All patients switched to clozapine after the third treatment. **Implementation:** The DES model was programmed using Extend software (Imagine That Inc., San Jose, CA).

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
					The decision to switch to another treatment depended on whether or not the pt was in relapse while on current medication or the occurrence of a side effect.	
Non-terminating population based simulation						
Cardiff T2DM model (McEwan, 2010) CUA/CEA (life years gained)	Lifetime (run over 100 years; data collected over last 10 years)	The study included the comparison of 4 treatment strategies, representing treatment escalation, all starting with monotherapy (metformin MF) followed by double then triple combination therapy (3 treatment lines). *The model was used to examine the effects of strategies on HRQoL improvements associated with different hypoglycaemia profiles (and side effects), rather than the efficacy variables, such as change in blood glucose levels (HbA1c). But the latter were used as an indicator for treatment switching.*	**Type 2 diabetes mellitus** Patients with type 2 diabetes mellitus (T2DM) starting treatment with antihyperglycaemic drugs. The baseline cohort characterises were drawn from the UK Prospective Diabetes Study (UKPDS) 68 outcome study	*HRQoL*	The model was implemented as a non-terminating simulation. The model was initialized with a prevalent population profile (*prevalence based cohort*) and utilised the annual T2DM incident rate to allow new cases to enter the model each year. Pts exited the model through diabetes-specific or all-cause mortality. The model required specification of population profiles, in terms of baseline demographics (*age, gender, duration of disease etc.*) and modifiable risk factors (*total cholesterol, HbA1c, body weight, blood pressure etc.*) [*The model was capable of modelling changes in modifiable risk factors*] Pts started 1st-line treatment as they entered the simulation. Following the application of a treatment effect modification to each pts' baseline HbA1c, the model used dynamic equations to project HbA1c over time. Pre-specified HbA1c threshold values were used to invoke an escalation in therapy to 2nd- or 3rd-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 to 0.51, 0.57, and 0.62 for 1st-, 2nd-, and 3rd-line treatment, respectively. This ensured that the model predicted a constant proportion of subjects on 1st-, 2nd-, and 3rd-line treatment to that seen in the UK when applying the specific thresholds to 2nd- and 3rd-line (7.6 and 7.9% respectively).	Pre-specified insulin (HbA1c) threshold values used to invoke treatment switching to second and third line, which was varied in the scenario analysis. These represented thresholds used in primary care, and taken from an observational study. *Disease progression was simulated using data from a published prospective study.* Time-dependent evaluation of risk factor profiles (*predicted complications*) were implemented using equations reported in the UKPDS 68 outcome study. Therapy profiles associated with 1st-line treatment (metformin) were taken from a Cochrane SR, and selected RCTs for 2nd-line and 3rd-line treatments. **Implementation:** 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 states prior to collecting summary statistics. The model was run over 100 yrs and data collected over last 10-yr period.
Terminating population based simulation						
Cardiff T2DM model (Erhadt, 2012) CUA	40 yrs (1 yr)	The comparison of 2 drugs used as 2nd-line treatment within a fixed sequence starting with the metformin (MF) monotherapy followed by the addition of a new drug (saxaglibitin vs sulfonylurea) as combination therapy, followed by the same 3rd line treatment (metformin plus insulin). Scenario analyses were also conducted where pts entering the model received	Pt population of interest was pts with T2DM receiving second line treatment after failure of first-line treatment. In the treatment pathway for the base case, all pts received fist-line monotherapy metformin. The baseline characterises and risk profiles for simulated pts were based on the pt population in a recent RCT of saxagliptin+MF vs sulfonylurea+MF in pts with inadequate response to MF.	HRQL plus changes in glycated haemoglobin (HbA1c) and body weight; occurrence of hypoglycaemia.	The model was described as a discrete event simulation, and as a 'fixed-time-increment stochastic simulation' based on UKPDS 68 outcomes equations. This time the Cardiff T2DM model was implemented as a terminal simulation. As well as the standard model outputs developed by the core equations in the Cardiff model, the current model was adapted to accommodate the following treatment effects: changes in HbA1c and body weight, and occurrence of hypoglaecaemia. AEs other than hypoglaecaemia were not incorporated as they were considered similar between comparator treatments. The model was run using annual cycles in which treatment-dependent risk factor profiles, including HbA1c and body weight, were modelled dynamically.	A HbA1c level of 7% was used as a threshold for moving from 1st to 2nd-line treatment, and 7.5% for moving to 3rd-line treatment. Data on clinical effectiveness of metformin taken from SR; 2nd-line treatments form head-to-head trial; and 3rd-line treatment from SR. Disease progression based on data from the UKPDS study 68. **Implementation:** A cohort of 1000 individuals was run 10 000 times.

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
		combination therapy as their 1st treatment.				

Abbreviations: AE adverse event or effect; CEA cost-effectiveness analysis; CUA cost-utility analysis; DES discrete event simulation; T2DM type 2 diabetes mellitus; pt, patient; mth month; RCT randomised controlled trial; SR systematic review.

Table E2.3: Discrete event simulation models used by rheumatology studies

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
PBM [early version of BRAM] (*Jobanputra, 2002*) CUA	Lifetime (Time spent on DMARD)	DMARD sequence with and without ETA (at different points)	Pts first stating DMARD treatment, with decision population developed within the model (data on baseline characteristics taken from published large observational study)	QoL (continuous)	Each pt followed a fixed sequence of treatments. The model cycles were based on the time spent on a particular DMARD. At the end of each cycle the maximum time on a DMARD (based on sampled time on DMARD) compared with the patient's remaining lifetime was calculated. A logic node was used to determine whether a pt would transfer to death or next DMARD. e.g. *if the pt was given a remaining lifespan of 6 yrs, and max times of 2yrs on SSZ (1st treatment) and 5yrs on MTX (2nd treatment), then the pt actually spent 2 yrs on SSZ and 4 yrs of MTX. The sample times for the other DMARDs were then not used for this pt*. In the case of pts moving from MTX or CyA, it was determined whether the reason for quitting was toxicity. Transfer from CyA was to palliative treatment if either of these was toxicity. *Pts switched to the next DMARD when the current DMARD was ineffective or produced toxicity.*	*Each pt followed a pathway containing a fixed treatment of 9-1 cDMARDs followed by palliative care.* **Implementation:** The model was constructed using Tree Age, and run for 10,000 virtual pts in each strategy.
BRAM (Barton, 2004; Clark, 2004; Chen, 2006; Malottki, 2011) CUA	Lifetime (Time spent on DMARD)	Sequences of DMARDs with and without various bDMARDs (at different points)	Pts first stating DMARD treatment, with decision population developed within the model (data on baseline characteristics taken from published large observational study) *(Model used by Malottki included pts who had inadequate response to their first TNF; matching the decision point*)	*%age change in* HAQ HAQ (continuous)	Pts assigned to different pre-defined sequences based on computer-generated random numbers (numbered 1-16 with the more expensive and more effective strategies put first). The model structure consisted of events (that take no time) and activities (that take a variable amount of time). The main loop ('start new treatment' – 'on treatment' – 'quit DMARD' – 'select next treatment') was followed for each DMARD successively, until no DMARDs remained and the pt 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 4 events (competing risks): death, HAQ increase, need joint replacement, or quitting DMARD.	6 events: start new treatment; quit DMARD; select next treatment; HAQ increase; joint replacement; death 1 activity: 'on treatment' **Implementation:** Two versions of the initial BRAM version were constructed, one in TreeAgree DATA Pro and the other in Borland Delphi. The model was run separately for each strategy being compared, and a set of consecutively run strategies were implemented in the same model.
Lindgren model (*Lindgren, 2009)* CUA/CEA	Lifetime (Time to next event)	bDMARD (3 lines of TNFs evaluated as a class) sequence with and without RTX (added to the start)	Pts with inadequate response to their first TNF. A model of the disease process under previous therapy was developed using registry data - Southern Swedish Arthritis Treatment Group (SSATG) register. (IPD available for up to 3 lines of TNFs). Simulations were performed for a population matching the pivotal RCT for RTX.	*DAS28 (high and low)* *HAQ*	The pt register (SSTAG) provided longitudinal epidemiological data, which allowed a model of the disease process as it evolved under previous therapy to be developed (IPD available on the use of up to 3 lines of TNFs). Pts entering the model were either starting their 2nd TNF or RTX, and stayed on these treatments until discontinuation, according to SSATG data for TNF and RCT for RTX. Pts on a TNF would then re-initiate treatment with their 3rd TNF according to the timings in SSTAG, whilst pts on RTX would start immediately on their 2nd TNF. The simulation could end	3 events: 'start treatment', stop treatment', and 'die'. 3 states: on treatment, off treatment, or dead. The treatment sate was further divided into high or low disease activity (DAS28 score 3.2 used as cut off point). In between treatments, all pts were assumed to have a high disease activity. While on or between treatments, pts had a certain HAQ (functional capacity) and DAS28 (disease activity), which in turn drove the costs and utilities.

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
			Deterministic base case analysis based on a single pt (not stated how selected)		before all pts re-initiated treatment (representing the data in SSATG). When pts failed again they switched to another TNF (assumed to be the same as 3rd-line). A pt could die at any time during simulation. A change is 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 pts relating to gender, age, disease duration, functional, drove the time to the next event.	**Implementation:** Software used to develop the model was not stated. *All time-to-events data for the TNF-inhibitors were based on the registry data. A cox proportional hazard model was used to identify covariates with possible impact on times to event, which included gender, age, disease duration, current HAQ, current disease activity and treatment line. Significant covariates were included and parametric survival models estimated using Weibull models except in the case of time to active disease. Here an exponential model was used instead. As not all patients had a period of low disease activity, the probability of reaching low disease activity was first estimated using logistic regression, and then the time to active disease using survival modelling. Time to death was estimated from age- and gender-specific mortality rates for general population, multiplied by a relative risk of 2.4 considering the disease severity of patients included in registry.*
Tran-Duy model (AkS) (*Tran-Duy, 2011*) CUA	Lifetime (Time to next event)	The model was used for the comparison of two strategies. One strategy included a sequence of five NSAIDs, which were chosen in a random order for each patient from a selection of 10 possible drugs. The second sequence included the same five NSAIDs plus two sequential TNF-inhibitor, which were also selected in a random order. *The next treatment was selected after the primary or secondary failure of the first, based on clinical guideline recommendations, and if BASDAI was ≥4.*	Pts with ankylosing spondylitis (Aks) with axial involvement.	*BASDAI, BASFI*	*General simulation process:* At the start of each simulation a virtual pt is created by drawing values from various appropriate probability distributions of pt attributes such as age (A1), symptom duration (A2), and BASADI (A3). Then the event 'visit a rheumatologist' (E1) occurs and a decision on eligibility for treatment is made. If a new treatments is decided, the event 'select a new treatment' occurs, which is followed by the event 'start a new treatment'. Next the pt enters the state 'on BASDAI decrease' (S1) or 'on BASDAI stability' (S2) depending on whether the treatment is effective. If no treatment is given to the pt at event E1, the state S2 is assigned to the pt. Then the procedure of the BASDAI-related state of the pt is implemented. In this procedure, time to all possible events associated with the BASDAI-related sate is sampled and compared. The event to which the time is shortest is the next event to occur (E_{next}). The simulation time is then advanced to the time at which E_{next} occurred (T_{event}) and the pt's attributes like A1 and A2 are updated at T_{event}. If E_{next} is 'visit a rheumatologist', a new loop starts. If the E_{next} is one of the BASDAI events (which includes 'end of BASDAI decrease', 'loss of response', and 'BASDAI increase by 1 unit') a relevant BASDAI-related state is assigned to the pt	The model included: 7 pt attributes: age, gender, BASDAI, contraindicated to TNF-inhibitor, symptom duration, work disability, having a paid job. 4 treatment related states: no treatment, an NSAID, a TNF-inhibitor, palliative care 3 BASDAI related states: on BASDAI decrease, on BASDAI stability, on BASDAI increase 3 BASDAI-related events: end of BASDAI decrease, loss of response to current drug, and BASDAI increase by 1 unit 5 BASDAI-neutral events: severe toxicity on current drug, visit to a rheumatologist, select a new treatment, start a new treatment, death. *Procedures: visit to a rheumatologist, select a new treatment, start a new treatment, on BASDAI decrease, on BASDAI stability, on BASDAI increase.* **Implementation:** The model was written in Delphi language (CodeGear Delphi. 2009, Embarcadero Technologies). SAS 9.2 (SAS Institute 2008, Cary) and R were used for data handling, and statistical and output analysis.

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
					and the procedure for the BSDAI-related state is implemented, which also starts a new loop. If the E_{next} is 'death', the simulation is terminated.	The size of the population was determined by repeatedly running the simulation with increasing initial population size until means and standard deviations of costs and QALYs became stable (occurred when initial population consisted of 13000 pts).
Tran-Duy model (RA) (*Tran-Duy, 2014*) CUA	Lifetime (Time to next event)	cDMARD sequence with and without a sequence of 4 bDMARD. *Strategy 1* (baseline) included 8 available conventional DMARDs: MTX followed randomly by SSZ or LEF, followed by 5 cDMARDs in random order. *Strategy 2* included the same 8 cDMARDs plus 4 biologics: MTX followed randomly by SSZ or LEF, then 2 TNF-inhibitors, and 2 non-TNF biologics, which were followed by 5 cDMARDs in random order. 2 TNF-inhibitors randomly chosen from 5 treatments; and 2 non-TNFs from 3 treatments.	Pts newly diagnosed with RA, with decision population developed within the model (pts' baseline characterises taken from the Nijmegen Inception Cohort; IPD available) *[The decision population could be developed within the model.]*	DAS28 HAQ	The sequential non-TNF biologics were randomly chosen from 3 drugs, but for pts who were RF +ve, RTX was available as the first drug, and for pts who were RF –ve ABA or TOC were available at the first position. The DAS28-realted states were used to determine the events that may occur given a trend of change in DAS28. Treatment-related states were used to determine changes in DAS28, times to DAS8-related events and a new treatment when the current drug fails. 3 events (severe toxicity, select new treatment, start new treatment) only occurred when a visit to a rheumatologist occurred. The remaining events were competing events. For competing events, the pt 'jumped' to the event to which the sampled time was shortest. When the event occurred, an associated procedure was invoked for implementation, where the pt characteristics were updated and times to next events computed. 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 the Dutch patient registry (DREAM biologic registry).	The model included: 7 pt attributes: age, gender, DAS28, HAQ, rheumatic factor (RF) +ve, disease duration, work disable. 8 treatment-related states: on 1st cDMARD (MTX), on 2nd cDMARD, on 1st TNF-inhibitor, on 2nd TNF-inhibitor, on 1st non-TNF-inhibitor, on 2nd TNF-inhibitor, on 'palliative' treatment 4 DAS28 related states: on DAS28 decrease, on DAS28 maintenance, on DAS28 increase, on DAS28 stability, 3 possible DAS28-related events: end of DAS28 decrease, loss of response to current drug (which caused an increase in DAS28), and DAS28 reaching 1.2 unit higher (this was used to help calculate the rate of DAS28 increase based on the assumption that DAS28 returned to the baseline level at 12 weeks after loss of response to current treatment, and that a 1.2-unit change in DAS28 was significant) 5 possible DAS28-neutral events: severe toxicity on current drug, visit to a rheumatologist, select a new treatment, start a new treatment, death. 8 Procedures: start MTX; visit a rheumatologist; select a new treatment; start a new treatments; and a series of 4 procedures for DAS28-related states (sample time to events that can occur given disease status; find event to which time is shortest; calculate time at which event occurs; advance time to event and update pt attributes). **Implementation:** The model was written in Delphi language (Embarcadero Delphi XE15.0, Embarcadero Technologies). SAS 9.2 (SAS Institute 2008, Cary) and R were used for data handling, and statistical and output analysis. An initial cohort of 10,000 pts was used. The simulation was run until death of the patients. The size if the initial population was determined by

Model name (*related publications*) **Economic evaluation**	**Time horizon (Cycle length)**	**Treatment sequences**	**Model patient population**	**Clinical effect measures**	**How sequences modelled**	**Health states and events**
						repeatedly running the simulation with increasing initial population size until means and standard deviations of costs and QALYs became stable. This occurred when initial population consisted of 10,000 pts, which was the number of simulations used for both first and second-order uncertainty analysis. 95% CIs of incremental cost-effectiveness ratio (ICER) was computed using non-parametric bootstrapping method with 100,000 times of sampling.

Abbreviations: AE adverse event or effect; AkS ankylosing spondylitis; BPM Birmingham Preliminary Model; BRAM Birmingham Rheumatoid Arthritis Model; BASDAI Bath Ankylosing Spondylitis Disease Activity Index; BASFI Bath Ankylosing Spondylitis Functional Index; CEA cost-effectiveness analysis; CI confidence interval; CUA cost-utility analysis; CyC cyclosporine; DAS28 Disease Activity Score 28 joints; DMARD/bDMARD/cDMARD disease-modifying anti-rheumatic drug/biological DMARD/conventional DMARD; HAQ Health Assessment Questionnaire; IPD individual patient data; LEF Leflunomide; MTX Methotrexate; mth month; NSAIDs Non-steroidal anti-inflammatory drugs; pt patient; QALY quality adjusted life year; RCT randomised controlled trial; RA rheumatoid arthritis; RTX Rituximab; SSZ Sulfasalazine; TNF Tumour necrosis factor inhibitor; TOC tocilizumab