University of Sheffield

PhD Thesis

Identification of evidence for key parameters for health economic models used to evaluate the cost-effectiveness of health care technologies

Johanna Lister

Year of submission: 2024

Title page

Identification of evidence for key parameters for health economic models used to evaluate the cost-effectiveness of health care technologies

Johanna Lister

PhD in Health and Related Research, Sheffield Centre for Health and Related Research (part-time), University of Sheffield

Supervisors:

Dr Chris Carroll (from June 2018)

Dr Suzy Paisley

Prof. Paul Tappenden

Prof. Eva Kaltenthaler (until June 2018)

Registration date: October 4, 2017

Thesis date: October 31, 2024

Word count: 74,922 (limit 75,000)

Acknowledgements

Being able to complete this thesis while working, having two young children, an energetic dog and a big garden to care for, reflects the unwavering support I have received from my family, friends and colleagues. My parents and grandparents nurtured my curiosity from an early age, and supported my educational endeavours, even if it eventually led me to move abroad to pursue a life outside Finland. I want to thank my two beautiful daughters, Helena and Iida, for inspiring me to be the best version of myself every day. I'm sorry I missed so many evenings, weekends and vacation days with you. Now let the fun begin! To my rock, James, for your practical and emotional support every day. I know the dinner doesn't cook itself or the kitchen doesn't clean itself. Now it will be your turn to pursue your dreams. To my beautiful Juliana, for your patience and understanding, especially during the most demanding phases of this research. I am also thankful for my friends, who provided both distractions when needed and encouragement when it seemed impossible to continue. I won't be leaving parties at 10 pm anymore because I need to work on my PhD the next morning! I'm also grateful to Alex for getting me addicted to Latin dance and keeping me fit throughout the studies.

The biggest and most heartfelt thank you goes to my three supervisors, Chris, Suzy and Paul. Your mentorship extended beyond the academic realm into personal guidance, providing comfort and motivation during though times. There were many times when I almost gave up and asked myself is this worth it. Your encouraging words and your ability to see how much I had already achieved assured me that I was making an original contribution. Thank you for being excited about my work, and always being available, even at short notice. Chris, you found time in your busy schedule for me *weekly* to make sure I was progressing and not having any problems. Suzy, you continued to supervise me years after you left the university, and even in your own personal time. Paul you were always findings ways to make my workload manageable, even if it meant looking for yet another executable excel model. I could not have asked for a better, more supportive team. Your presence and insights have been pivotal in allowing me to navigate the complexities of my research topic.

Abstract

Objectives. The objective of this PhD study was to compare different information retrieval methods that can be used to identify health economic model inputs. **Methods**. Existing search methods were compared to two alternatives (iterative searching and rapid review), using three health technology assessment (HTA) case studies in ulcerative colitis, thyroid cancer and breast cancer tumour profiling risk stratification. Key criteria for selecting the case studies were the availability of an executable Excel model. Two model inputs were chosen to be tested: health state utilities and baseline risk of clinical events. Usual practice searches were updated, and alternative search methods (iterative searching and rapid review) were conducted, and the differences in model inputs identified by each search approach were analysed. Differences were evaluated in terms of time taken to search, sensitivity, burden (precision and number needed to read) and relevance of identified information. The identified model input values were tested in an executable health economic model, and, when feasible, the model results were compared in order to understand the impact on model outputs. **Results**. Usual practice for identifying health state utility inputs was a systematic review in all except one case study, where a previous health economic model output was used. In all case studies the alternative search methods were mostly less resource intensive and resulted in identical or similar model inputs, with no changes to the conclusions drawn from the health economic model. Usual practice for identifying baseline risk of clinical events varied from no recorded search steps to a systematic review. When the effort for usual practice could not be estimated due to the lack of recorded search steps, the time difference could not be estimated. However, it was clear that applying alternative search methods increased the transparency. **Conclusions**. Alternative search methods were more efficient and more transparent than established search methods, without impacting the health economic model conclusions. Further case studies are required to examine whether this conclusion remains generalisable, and applies to other health economic model inputs.

Abbreviations

AE	Adverse event
AHRQ	Agency for Healthcare Research and Quality
AMCP	Academy of Managed Care Pharmacy
AML	Acute myeloid leukaemia
AMNOG	Arzneimittelmarktneuordnungsgesetz
AOTMiT	Agencja Oceny Technologii Medycznych i Taryfikacji, Poland
APR-DRG	All Patient Refined DRGs
AWMSG	All Wales Therapeutics and Toxicology Centre
BCFI–CBIP	Belgian Centre for. Pharmacotherapeutic Information
BSC	Best supportive care
CADTH	Canadian Agency for Drugs and Technologies in Health
CCTR	Cochrane Controlled Trials Register
CDK4/6i	Kinase 4/6 inhibitors
CDS	Cochrane Database of Systematic Reviews
CEA	Cost-effectiveness analysis
CES	Collège des Économistes de la Santé, France
ČFES	České farmakoekonomické společnosti, Czech Republic
CHEPA	Centre for Health Economics and Policy Analysis, McMaster University
CL	Clinical Guideline
CRD	Centre for Reviews and Dissemination, University of York
DARE	Database of Abstracts and Reviews of Effectiveness
DG	Diagnostic Guidance

DM	Distant metastasis
DRFI	Distant recurrence-free interval
DSU	Decision Support Unit
EAG	External Assessment Group
EDMUS	European Database for Multiple Sclerosis
EED	Economic Evaluation Database
EMA	European Medicines Agency
ENTREQ	Enhancing transparency in reporting the synthesis of qualitative research
EQ-5D	EuroQOL five dimensions questionnaire
ER	Estrogen receptor
ESRC	Economic and Social Research Centre
ET	Endocrine therapy
EUnetHTA	European Network for Health Technology Assessment
FDA	Food and Drug Administration
GEAR	Guide to Economic Analysis and Research
HAS	Haute Autorité de santé, France
HERU	Health Economics Research Unit, University of Aberdeen
HILA	Lääkkeiden hintalautakunta, Finland
HIQA	Health Information and Quality Authority, Ireland
HRQoL	health-related quality of life
HRQoL	Hazard ratio
HTA	Health technology assessment
HTAi	Health Technology Assessment International

IBD	Inflammatory bowel disease
ICER	Institute for Clinical and Economic Review
ICER	Incremental cost-effectiveness ratio
INA-BCHRQoL	Indonesia Breast Cancer Health-Related Quality of Life
INAHTA	The International Network of Agencies for Health Technology Assessment
INESSS	Institut national d'excellence en santé et services sociaux, Quebec Canada
Infarmed	Instituto Nacional da Farmácia e do Medicamento, Portugal
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
ISSG	Information Specialists Subgroup
ITT	Intention-to-treat
KCE	Health Care Knowledge Centre, Belgium
LN	Lymph node
MAIC	Matching-adjusted indirect comparison
MS	Multiple Sclerosis
MTA	Multi-technology assessment
MTC	Medullary thyroid cancer
NCCT	National Collaborating Centre for Methods and Tools
NHS	National health service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
NMA	Network meta-analysis
NNR	Number needed to read

NOMA	Norwegian Medicines Agency, Norway
OGYEI	Országos Gyógyszerészeti és Élelmezés-egészségügyi Intézet, Hungary
ONS	Office for National Statistics
OS	Overall survival
PBAC	Pharmaceutical Benefits Advisory Committee, Australia
PFS	Progression free survival
PHARMAC	Pharmaceutical Management Agency, New Zealand
PICO	patient/population, intervention, comparison and outcomes
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality-adjusted life year
RAMESES	Realist And Meta-narrative Evidence Syntheses: Evolving Standards
RCT	Randomised controlled trial
ScHARRHUD	School of Health and Related Research Health Utility Database
SCI	Science Citation Index
SMC	Scottish Medical Consortium, Scotland
SoC	Standard of Care
STARLITE	Sampling strategy, type of study, approaches, range of years, limits, inclusion and exclusions, terms used, electronic sources
SuRe Info	Summarized Research in Information Retrieval for HTA
TAG	Technology Appraisal Guidance
TARCiS	Terminology, application, and reporting of citation searching
TaSPOR	Taiwan Society for Pharmacoeconomic and Outcomes Research, Taiwan
TSD	Technical Support Document
TSU	Technical Support Unit

TTO	Time to trade-off
TVL	Tandvårds- och läkemedelsförmånsverkets, Sweden
UC	Ulcerative colitis
UK	United Kingdom
WHO	World Health Organization

Contents

Tables

Figures

1 Introduction

The topic of this thesis is the identification of evidence for key input parameters for health economic models used to evaluate the cost-effectiveness of health care technologies within health technology assessments (HTAs). Its purpose is to define and to test empirically search methods that can be used for the identification of evidence for health economic models. In 2012, Paisley found that no studies reported established search methods for the retrieval of the full range of model input parameters.[1] Usual search practice can be described as commonly used or described search practice, such as full systematic literature review searching for estimates of relative treatment effects, primarily from randomised controlled trials (RCTs). These search methods do not address the needs of health economic models that additionally require different types of non-RCT data retrieved from a variety of sources. Further, the scope of the evidence required to inform health economic models often cannot be pre-defined as the modeller's understanding of the decision problem addressed by the model can evolve through the course of the model development process. Due to the lack of guidance, there remains a tendency to rely on the current, systematic literature review search methods, rather than an approach that might be a better fit for modelling evidence requirements. Choosing the right input parameters for health economic models is essential to ensure the accuracy and reliability of the model's results.[2] Model parameter estimates without a clear method or rationale of how data sources were identified and selected, can lead to a lack of transparency. This may have negative consequences in terms of the confidence that decision-makers and other model users can place in the results of those models.

1.1 Aim and Objectives

The aim of the thesis is to explore the impact of different search approaches in the identification of certain evidence for health economic models. The objectives are:

- To identify and develop two alternative search approaches judged to show potential for identification of evidence for health economic models
- To test and compare the two approaches with usual search practice
- To evaluate all three approaches in terms of efficiency (time taken to search), burden (precision and number needed to read), relevance and impact on model outputs, using a case study approach
- To develop a search framework for the reporting of search methods for health economic models

Definitions for usual search practice and measures of search outcomes are provided in the Methods Chapter 4.2.2. More specifically, subchapters 4.2.2.1 and 4.2.2.3 provide definitions for usual search practice and search performance outcomes, respectively.

This thesis will specifically focus on models of health technologies (drugs, genomic test) used to treat non-communicable diseases.

1.2 Motivation for the Study

The motivation for the thesis originates from both the importance of health economic models in HTA decision-making, as well as personal experience working with systematic reviews and health economic models for HTAs.[3]

1.2.1 Identification of Health Economic Model Inputs for HTA

Health economic models play a pivotal role in HTA and influence drug coverage and reimbursement decisions across various countries.[3] The models developed significantly impact on the estimated value of a health technology.[3] To ensure that health economic models are robust, it is crucial to employ rigorous and efficient search methods for identifying the health economic model inputs. Currently, HTA agencies primarily focus on systematic literature searches to inform relative treatment effect inputs. 'Systematic searching' in this context refers to information retrieval that is scoped according to a clearly focussed search questions, and comprehensiveness in terms of the retrieval of evidence. Additionally, 'systematic searching' applies well-known information retrieval guidelines, such as those from Cochrane or The Centre for Reviews and Dissemination (CRD).[4,5] However, specific methods for conducting searches in health economic modelling remain under-addressed in HTA guidelines. Consequently, reliance on traditional literature review approaches persists. As health economic models are associated with multiple evidence needs, it may not be feasible or efficient or even advisable to apply systematic searching to all the evidence required for the models. Another consideration is whether systematic searching is necessary for every model input. However, if no information retrieval steps are recorded, this can lead to potential transparency issues.[6] Although systematic searching is typically necessary to obtain the evidence needed for health economic models, HTA guidelines do not specify particular methods.[7]

Limited guidance or publications exist regarding the process of conducting searches in the specific field of health economic modelling. Golder *et al*. published a study on the feasibility and efficiency of database searching to populate health economic models, and identified forty-two information requirements across multiple different types of model inputs.[8] The study used eighteen search strategies that were aimed at high precision rather than high sensitivity (see definitions for precision and sensitivity in Section 4.2.2.3). The study demonstrated that health economic models are associated with multiple information needs and that these relate to different types of information. A publication by Philips *et al*. reviewed fifteen published health economic modelling good practice guidelines, including guidance on data.[9] In this study, content relating to data identification was summarised. The guidelines were found to make general statements about needing to identify data for the models in a systematic and transparent manner but included little or no procedural guidance beyond identification of relative treatment effect estimates.

Paisley (2016) proposes minimum requirements for identifying evidence for key model parameters, contributing as a search framework to providing initial guidance across all key model parameter inputs.[10] Further, The International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force has recently published guidance on iterative searching on utilities in the context of cost-effective modelling.[11] This publication describes challenges associated with searching for utility values using systematic, sensitive approaches and recommends iterative searching. Some details on methods are given

in the Task Force report, including considerations for search terms, subject headings, free-text searching, where to search, and supplementary search techniques. Considerations for sensitivity versus precision are also given. The report notes that insufficient empirical evidence exists on the most effective method for conducting searches to identify utility inputs. The implications of doing less than comprehensive searching are rarely addressed.

The absence of clear guidance on search methods for health economic modelling may be attributed, in part, to the lack of established alternative approaches. Systematic review methods relating to parts of literature review other than information retrieval through database searching (e.g., appraisal, synthesis, and analysis in different types of reviews) have seen ongoing development, in response to the increasing range in the types of questions addressed using the systematic review approach (e.g., diagnostic test accuracy, prognostic, aetiology, qualitative).[12-15] In contrast, information retrieval methods research, in the context of systematic reviews, has continued to focus on retrieval processes but not on the underlying premises of using a pre-defined search question and aiming for high sensitivity.[16,17] Systematic searching is typically assessed based on precision and sensitivity, but it has been acknowledged that even in optimal conditions, measuring the completeness of identified literature remains challenging.[18] Sensitivity and precision are defined, respectively, as the ratio of relevant citations identified to the total number of relevant citations, and the proportion of retrieved citations that are relevant.[19] That is, sensitivity is the proportion of relevant items that exist and that have been retrieved by the search (the extent to which the search has not missed anything relevant), and precision is the proportion of items retrieved by the search that are relevant (the extent to which the search only retrieves relevant items). Searches aimed at high sensitivity tend to take a long time to conduct and can be associated with large amounts of identified records, meaning resources are required to process large amounts of irrelevant information (i.e. burden of review). Searches aimed at high precision minimise the retrieval of irrelevant information but risk missing relevant information.

Additional information seeking practices have been described in the literature, for example: berry-picking;[20] citation tracking;[21,22] author/expert consultation[22,23] and grey literature searching.[24] However, the empirical evidence assessing the effectiveness of different information retrieval methods is lacking.[25] This is particularly true for the identification of health economic model input parameters, where an evaluation should not only consider the indicators of completeness but also indicators of impact on model conclusions, i.e., the relevance of the data to the decision problem that the model is addressing. From a practical perspective, it is unlikely to be possible to conduct a full systematic review search for every single information need in a health economic model. Rather, it has been argued that search activities for a model should focus on searching for data related to those aspects of the model that have the most impact on the model results, putting the emphasis on identifying sufficient information to understand the implications of uncertainty associated with the model.[26]

The extent of the trade-off between sensitivity and precision in identifying model input parameters using different search methods remains uncertain. The lack of understanding regarding potential trade-offs between review burden (maximising sensitivity) and the relevance of identified sources (maximising precision) poses challenges for researchers considering different, including alternative, search

approaches. Greater understanding of the impact of sensitivity and precision, particularly in terms of the effectiveness and efficiency of all types of search approaches would enhance confidence in selecting appropriate approaches and understanding associated trade-offs.

1.2.2 My Experience as a HTA Evidence Generation Specialist

My motivation for the study comes from my own experience working as a health analyst carrying out literature reviews and developing health economic models for HTA submissions. One key difference that I have observed is the difference in analytical frameworks of systematic reviews and health economic models; the purpose of an economic model is different from that of a systematic review, and these differences need to be taken into account when searching for evidence to inform model development.

Usually the information retrieval strategy to identify relative treatment effect estimates is to perform one large, global systematic literature review on the intervention and comparator relative treatment effects. This approach can cause some practical problems, as searches are required to aim for sensitivity rather than precision, rendering the search process inefficient and very resource intensive when all the possible comparators of interest around the world are added. For example, in Eastern European countries, older therapies can still be in use and they can be associated with vast amounts of evidence. In some countries, such as in France, up-coming therapies also need to be included in the economic model (and hence in the systematic search for clinical evidence).[27] In recent years, it has become more challenging to apply systematic review search methods and to manage the review within the timeframe and budget available. This is due to the high number of records identified by sensitive search strategies. This has an impact on the resources available for other evidence requirements of the health economic model. This is particularly important for those evidence requirements that may significantly impact the model results.

In my work, I would ideally like to be able to distribute the input parameter identification efforts as appropriate for each health economic model (e.g., focus on those model inputs that are most impactful on the model results). More knowledge on the trade-off between sensitivity and efficiency of different search methods (might vary from one model or type of model parameter to another), as well as detailed descriptions of the required minimum sources and accepted search methods could help to reduce the unnecessarily high proportion of the research effort being spent on only a few, labour intensive full systematic reviews. Doug Altman, a late medical statistician, has argued that the society needs "*less research, better research and research done for the right reasons*".[28] In some instances, a full systematic review is required (e.g., clinical efficacy for the HTA dossier), but in the instances when full systematic review may not be the optimal solution, another search method that would need less research effort without compromising the quality of the output or transparency in reporting, would be in line with Altman's principle of reducing waste in research while ensuring high quality and replicability of it. The evidence generation activities that I typically face at work require me to consider several, sometimes conflicting perspectives, such as global versus local perspectives or clinical dossier sections versus the economic model. Health care technology manufacturers often submit reimbursement applications for the new technology in several countries in parallel, and the requirements, processes and comparators vary. Further, many reimbursement dossiers contain two distinctive elements: clinical efficacy and economic value, each with their own data search recommendations and needs. Identifying key input parameters for

health economic models often needs to be fitted into the overall evidence generation activities. Existing systematic review can also be utilised. Careful consideration of all of this (i.e., the needs of the clinical and health economic dossiers around the world and existing systematic searches) is done in the interest of avoiding running several similar (but not identical) searches.

1.3 Overview of the Thesis

In Chapter 1 of this thesis, the context of the study is described, including aims and objectives (Section 1.1), motivation for the PhD study (Section 1.2), and overview of the thesis (Section 1.3). Chapter 2 describes the role of health economic models in HTA, including a chapter overview (Section 2.1), economic model information requirements (Section 2.2), types of evidence required for health economic models (Section 2.3), an overview of existing search methods (Section 2.4), and an overview of types of data sources used in health economic models (Section 2.5). Chapter 3 provides a narrative review of literature, including existing literature on search methods, especially in systematic literature reviews and in evidence-based medicine (3.1) and search methods specific to health economic modelling (Section 3.2). Chapter 3 also includes a review of search guidance given in HTA and modelling guidelines (Section 3.3). Chapter 4 is an overview of the methods used, including the rationale for the methodological approach. (Section 4.2) and ethical approval (Section 4.3). Chapters 5, 6 and 7 each report the methods and results of the three case studies that comprise the PhD study. Chapter 8 provides a discussion of the case study results, and Chapter 9 is a discussion that compares the search methods and puts the PhD study in context of HTA methods development. It also includes the strength and limitations of the PhD study and the contribution of the thesis.

2 Health Economic Models in Health Technology Assessment

2.1 Chapter Overview

In Chapter 2 the context of the study is described, including model information requirements (Section 2.2), types of evidence required for health economic models (Section 2.3), an overview of existing search methods for health economic models (Section 2.4) and an overview of data sources (bibliographic databases, specialist databases and other sources) that are being used to identify inputs for health economics models (Section 2.5).

2.2 Overview of Health Economic Model Information Requirements

Health economic models[1] can provide a useful synthesis of data in the absence of sufficient primary evidence and help decision makers in their task of determining whether a new technology represents an efficient use of scarce health care resources. In doing this, modelling provides a means of assembling a broad range of different types of evidence within a single decision-analytic framework. As a result, the modelling process can generate multiple, complex information needs often incorporating various types of evidence from different information resources such as bibliographic databases, specialist databases and non-database searching e.g., trial registries, grey literature, expert contact, regulatory agencies, national reference cost databases, national health information databases, and national statistics.[26,30-32]

Systematic literature reviews are a key element of evidence-based medicine. They are widely used to assess and synthesise evidence on a particular research question. Systematic reviews enable collection of evidence from a variety of sources in a comprehensive manner.[33] While a valid finding for a systematic literature review question could be that there is insufficient evidence to draw conclusions on the question being addressed,[34] this is not helpful when a decision about health care technology reimbursement needs to be made. It is possible that existing evidence alone cannot fully answer the questions associated with a specific decision problem. Uncertainties associated with reimbursement decisions may be better understood/reduced through further primary research. However, often decisions need to be taken even though the available evidence is limited or uncertain.

Search methods for the identification of RCT evidence on treatment effect exist. While search methods for other types data also exist (e.g. search filters on ISSG resource[35]) these are less established in the context of health economic modelling. To populate health economic models a broad range of data is required; a classification of model input parameters identified fourteen different types of information, such as treatment effect, adverse events (AE), costs, resource use, health state utilities, and baseline risk

[1] Defined as a mathematical framework used to simulate the economic and health outcomes of different healthcare interventions.[29]. Michael F. Drummond MJS, Karl Claxton, Greg L. Stoddart, and George W. Torrance. Methods for Economic Evaluation of Health Care programmes. 4th ed. ed: Oxford University Press; 2015.

of clinical events.[36] Methods exist for the identification of other types of evidence, for example based on the range of filters available on the ISSG website.[37] However, beyond searching for RCTs there is limited empirical evidence on the effectiveness of methods for identifying other types of evidence, particularly non-standard format evidence such as real world data sources.

Systematic reviews aim to address a single focused question and use this question to underpin the design of the review. The underlying principle of the searching is to identify studies that address the same question as the review. Modelling on the other hand can be aimed to address the same decision problem, but it does so by drawing on many different types of evidence. A single research question is inadequate given the diversity of the health economic model information requirements. A systematic review might conclude that there is insufficient evidence to draw a conclusion. However, health economic models will be used for drawing conclusions about cost-effectiveness of one treatment over another, even in presence of imperfect evidence, usually with some assessment of uncertainty. Health economic models can be a tool to help determine what can be concluded based upon the evidence that is available, so that a decision can be made about the reimbursement of a technology and potentially where future research might be valuable. A conflict arises when the research driven systematic literature review tradition is imposed upon the modelling tradition that cannot be clearly defined by one research question. Table 1 summarises the differences in modelling and systematic literature review traditions.

Table 1. Systematic literature review tradition vs modelling (Kaltenthaler *et al*. 2011[26] & Paisley 2012[1])

Item	Model input identification	'Cochrane-style' Systematic literature review tradition
Research question	Multiple information needs not represented by single question	Clearly focused question
Process	Iterative and dynamic, where information needs can emerge/change during the process.	Pre-defined
Use of PICO (or another relevant, instrument such as SPIDER, PerSPEcTiF or SPICE). [38-41]	Can be suitable for treatment effect parameters but not suitable for all the multiple searches required, and there are complex research needs plus lack of clarity on some information needs (i.e. use of such framework require being able to predefine the information need).	Suitable for Cochrane style searches but increasingly there is dialog about different types of systematic review, including complex topics such as public health topics (see Chapters 2.4 and 3.2).
Emphasis	Identify sufficient information to maximise understanding of the consequences of uncertainty in the model	To minimise risk of bias

2.3 Types of Evidence Required for Health Economic Models

Several types health economic model inputs exist.[32] For the purposes of this research, the focus will be on the following model input parameters: relative treatment effects, harms i.e. adverse events (AEs), health state utilities, costs, resource use and baseline risk of events. Other types of information include adherence, current practice, epidemiology, modelling methods, patient preference, prognosis and results/methods from other models.[32] The focus is on these key model input parameters because they are common to most economic models and because, except for treatment effects,[5] there is a lack of definitive search methods. Additionally, they are featured as the key model input parameters in existing publications.[10,26] For each one of the input parameter types, this section will include potential sources and types of evidence.

2.3.1 Relative Treatment Effects

Health economic models include an underlying disease model, that mathematically describes the disease progression with standard of care (SoC). Standard of care is defined as "the level of care an average, prudent healthcare provider would offer to a patient in a similar situation".[42] In health economic modelling the SoC can be modelled using the data for the control arm of the pivotal clinical study or from an indirect treatment comparison, depending on the context and data availability. The effect of the intervention (versus SoC) is captured through relative treatment effect. Typically relative treatment effects are obtained from meta-analyses or network meta-analyses that include all available relevant RCTs that report the outcome of interest for the intervention(s) of interest.[43] Existing guidelines typically recommend using "Cochrane style" searching to identify relative treatment effect estimates.[44] Cochrane style searching here refers to systematic review searching that has a clearly focused search question and that aims to search extensively as described by well-known information retrieval guidelines, such as those from Cochrane or the CRD.[4,5] Detailed guidance on how to conduct a full systematic review is available, including search methods, and this method is a transparent (albeit resource intensive) way of identifying the pivotal studies from which the treatment effects for health economic models are derived. A hierarchy of evidence suggests that meta-analyses of RCTs or single RCTs are the best sources of evidence.[45] For this reason, the treatment effect searches are often limited to RCTs, where it is perceived that sufficient RCT evidence exists. Sometimes the Cochrane style systematic search on RCTs is extended to include observational studies in an attempt to identify additional evidence on relative treatment effects. In an ideal situation, the systematic literature review that is performed to assess the clinical effectiveness of the new health technology in health technology assessment (HTA), could also be used to generate the treatment effect in the health economic model. HTA is a form of research that generates information about the clinical and cost-effectiveness of health technologies.[46] However, some specific issues related to input parameter identification for health economic models can make it challenging to design and implement one search which is fit for both purposes.[47] The health economic model may require additional input parameters and consequently additional searches, such as comparator treatment effects (to use in indirect treatment comparisons/network meta-analyses), long-term treatment effects, treatment effects of companion treatments/diagnostics or treatment effects of therapies given after the health technology of interest. The information about absolute long-term outcomes data may be limited to the comparator

in the model (i.e., standard of care) as the new health technology is unlikely to have any long-term evidence available at the time of the assessment. Long-term outcomes data from published sources (among other validation exercises) will allow comparing of the model outcomes with findings from studies reported in literature and is therefore an important part of the evidence requirements for health economic models.

2.3.2 Adverse Events

A further aspect of HTA is consideration of AEs. The two main sources for AEs data in the health economic model are the patient-level data from the pivotal Phase 2 and/or 3 trial(s) for the new health technology and RCT publications identified during the treatment effect systematic literature review, but other sources may also be required, especially for considerations of long-term or uncommon AEs, as well as any additional considerations that might be relevant such as impact of AEs in vulnerable groups. The model may include all grades of AEs, or only a selection of AEs (e.g., serious or grade 3 and 4 AEs or those that impact uptake, adherence of persistence). A selection of AEs would typically focus on those AEs that are the most likely to have an impact on costs and quality of life and therefore impact the model results most. The AEs of comparator technologies should also be included. The following inputs would typically be required: frequency of AEs, number of episodes per patient, duration of AE, information/assumption when the AEs typically occur, cost of treating AEs and utility decrements associated with AE. Further AEs may be associated with downstream therapies, such as blood clots or infections from surgeries. Similar inputs would be needed for those, as for the main treatments compared in the model i.e., frequency, number of episodes, cost and utility decrements. Cost and disutility aspects of AE input identification are discussed in more detail in the cost and health state utility sections below (see Sections 2.3.3 and 2.3.5).

As with other model input parameters, the level of detail in which AEs need to be captured in the model depends on the disease area, safety profiles of the health technologies included in the model, the value proposition, unmet need, and sensitivity of the HTA agency to safety issues. The two main sources for AEs data (patient level data and other RCT publications) are usually a good starting place but are often insufficient. The RCT publications may only include summary statistics of the AEs or health technologies can be associated with AEs that are rare and/or do not occur soon enough to be captured by the clinical trial. For example, studies have demonstrated that Esmya (ulipristal acetate) is successful in diminishing bleeding linked to uterine fibroids. However, the European Medicine Agency's (EMA's) Pharmacovigilance Risk Assessment Committee (PRAC) carried out an additional assessment following reports of serious liver injury. The PRAC determined that Esmya could have played a role in the emergence of certain instances of severe liver injury – this would not have been clear from the RCT evidence alone.[48] Even rare AEs can have a significant impact on the health economic model results, if they have devastating consequences for patients and incur high costs. Further, ethical reasons may override cost-effectiveness conventions as the basis for reimbursement decision. For this reason, identifying AE data for health economic models may need to be extended beyond RCTs identified from bibliographic database searches to include observational studies, regulatory data, real world data, electronic health record data, or even social media . Using Cochrane style searches without study type restrictions can result in a wide systematic literature review, especially in a situation where there are numerous comparators to consider. Golder et al. found

that a lower number and a narrower range of side effects are generally reported in published than unpublished studies.[49] HTA agencies in France and Canada are particularly sensitive to safety issues, especially in the generation of the economic evidence.[27,50] For example, health economic models in most countries include a selection of the most impactful AEs (e.g. only severe/serious/etc.), whereas in Canada and France all AEs that occurred in the clinical studies must be included in the health economic model, and their costs and decrement on utility must be considered.

2.3.3 Costs

To estimate costs in health economic models, two elements are required: unit costs (e.g., the monetary value of each specialist appointment and the resource use (e.g., number of specialist visits in a given time period).[10] This Section gives details on cost inputs, and the next Section 2.3.4 details on resource use. These two elements are closely related. Separate searches can be run, but it is possible to put together only one bibliographic database search protocol with the aim of identifying published studies. In my experience, it is possible to identify the relevant publications related costs and resource use, despite these being two distinctly separate items, as long as the search is planned with this dual purpose in mind (from search terms to data extraction fields) and as long as the supplementary activities are specific to cost or resource use (e.g., additional websites searched).

Evidence requirements relating to cost parameters in health economic models are determined by the interventions (drug acquisition costs, administration costs), health states (disease management costs in all relevant stages of the disease pathway), AEs (cost of treating), and other events in the model that have cost implications, such as death that may be associated with end-of-life costs. The types of costs included in the health economic model also depend on the perspective taken for the analysis. Public payer perspective includes direct costs that are all costs that are closely related to healthcare. In France, the economic analysis is conducted from a public payer perspective but it also includes costs to patients, such as cost of travel to health care.[27] In Québec in Canada, the reference case for the economic evaluation is societal, which also includes indirect social and economic costs such as decrease in productivity, early retirement, and income losses.[51]

Paisley (2016) discusses the multiple sources that are used to estimate costs for health economic model input parameters, such as RCTs, observational studies, secondary economic evaluations, drug tariff lists, administrative information, and routinely collected statistics.[10] To estimate drug costs used in the model, national drug tariff (e.g., British National Formulary (BNF) in the UK or Lauer-Taxe in Germany) lists are often used as a source.[52,53] In some countries, not all drugs are included in the national drug tariffs lists, and other sources such as published cost studies or hospital websites may need to be used. For example, in Germany certain high-cost drugs, including some oncology medications, may not be listed directly in the Lauer-Taxe. Instead, these drugs are managed through separate funding mechanisms or special programs.[53,54] Often, national drug tariff lists are not freely accessible (outside the jurisdiction) and knowledge of local sources and language are required to access drug prices, as well as costs of other types of interventions such as medical devices or diagnostics test. For example, in Germany Lauer-Taxe also contains prices for some medical devices used in outpatient settings.[53] The Federal Institute for Drugs and Medical Devices (BfArM) publishes some additional information on the costs of medical devices.[55] Global

departments in pharmaceutical companies responsible for health economic evidence generation work together with their local affiliates, who have detailed knowledge of the local best sources to use, what type of cost to include and how to include it in the model. Published sources identified from bibliographic databases are can be useful for identifying packages of care such as costs of end-of-life care and for identifying and specifying relevant cost items. They are most useful for identifying costs for items not of primary interest in the model, such as an aggregate cost of downstream therapy given after the treatment of primary interest. Published economic evaluations may also be used for checking what has been used in previous analyses, and reference checking may prove to be useful. Prior HTA reports can be checked for previously identified and used inputs.[26] Using the same approach as in previous HTA reports would need to be justified to ensure that the past approach is appropriate in the current context. Any differences may similarly need to be explained also. It is worth noting that unit cost prices are recommended to be from the jurisdiction of interest (see also next Section 2.3.4 Resource use).[56]

2.3.4 Resource Use

Bibliographic searches can provide a useful overview of the available published evidence for resource use.[26] There is an element of conceptualisation for defining the resource use (and costs) required in the model.[26] This can be done by determining standard of care or recommended clinical practice although care needs to be taken as actual practice can vary from recommended practice.[57] Other sources can also be important in determining resource use input parameters for the health economic models, such as national statistics, administrative databases, previous HTA assessment reports or even expert survey for the jurisdiction in question.[26] Depending on the context, both recommended clinical practice (e.g. in clinical guidelines) and actual clinical practice (e.g. observed from real world evidence) may be of relevance. For example, in the US the National Comprehensive Cancer Network (NCCN) guidelines contain evidence-based recommendations that could be used to model resource use in a given cancer.[58] The relevance of using clinical guidelines depends on how up to date the guidelines are and how closely the guidelines are followed by the clinicians. This can differ between jurisdictions as well as between disease areas, and even organisations that publish the guidelines. Other potential sources include literature sources, health care records and expert opinion. When assessing the relevance of different sources, HTA-specific recommendations as well as the evidence-based medicine pyramid can be used when deciding on the most relevant data source.[59] The challenge remains that to appropriately interrogate these data sources, specialist knowledge of the sources and local language is required.

2.3.5 Health State Utilities

If the clinical trial collected health outcomes data uses a standardised instrument such as the EuroQOL five dimensions questionnaire (EQ-5D), most guidelines recommend using these estimates in the health economic model to estimate health state utilities.[44,60-67] If EQ-5D or other standard instrument is not available from the pivotal trial, then one option is to derive estimates from a review of the literature. When mapping is required to derive utility estimates, it may be necessary to perform a bibliographic database search to identify validated mapping algorithms. With respect to the EQ-5D, instrument country-specific tariffs for many countries have been published for either the 3-level or 5-level versions.[68] These tariffs have been developed to account for the loss of individual patient preferences.[69] However, health

outcomes data from trials may not always be available, and even if health outcomes were collected in the trial, they may not fulfil all the utility needs of the model. For example, a cancer trial may collect EQ-5D data at each study visit until disease progression and only include one further visit after disease progression where the EQ-5D questionnaire is administered. This is an issue because the utility estimate is unlikely to be reflective of the utility in people with progressed disease over their remaining survival time. Further complications may arise from comparators other than those included in the trial; for example, consideration needs to be made whether it is appropriate to apply the trial health state utility value (even if health state specific and not treatment-specific) to the additional comparators. Utility decrements for AEs may need to be included in the model, and these would then need to be identified from literature (non-disease specific searches) if it is not feasible to use the trial data to conduct these analyses. The model may include health states that did not occur in the trial, and hence they are not captured by the trial. An example of this may include downstream therapy or other long-term consequence associated with the disease/condition that occurs after the end of the trial duration. Paisley (2016) discusses the challenges in identifying health state utility values from literature. Two further publications provide recommendations on searching health state utilities.[123,124] These publications are discussed in detail in section 3.2.5.1.4 (searching for Utilities).

2.3.6 Baseline Risk of Clinical Events

Baseline risks of clinical events are used to model the natural history of the disease, as well as to inform model structure. In some instances, identification of baseline risk of clinician events will consist of deriving the baseline model from the placebo/standard of care arm(s) of a trial(s) (e.g., in advanced cancer). In other cases, natural history studies will be used, such as for multiple sclerosis or rheumatoid arthritis. Each source is associated with advantages and disadvantages, such as clinical trials populations being highly controlled due to inclusion criteria, registry data being incomplete or unavailable. This is a very difficult input parameter to search for because data requirements vary significantly between different disease areas and can be specific to jurisdictions. Therefore, the potential sources also vary greatly: from published studies identifiable from bibliographic databases to registries and national statistics, clinical/economic professional organisations and charities. Examples of these include Surveillance, Epidemiology and End Results Program (SEER), British Heart Foundation (BHF) and European Cystic Fibrosis Society Patient Registry (ECFSPR).[70-72]

2.4 Overview of Search Methods

This Chapter gives a brief overview of information retrieval methods in general, as well as specifically for health economic models.

2.4.1 Information Retrieval

The focus of information retrieval of systematic searching is typically on sensitivity/recall, but it has also been shown that even under the best of circumstances it is not possible to measure the completeness of literature identified.[18] Recall measures the proportion of relevant documents that are successfully retrieved out of all relevant documents available (i.e. documents retrieved divided by total number of

existing documents).[21] There is a trade-off between recall and precision, which quantifies the fraction of retrieved documents that are relevant (i.e., relevant documents retrieved divided by the total number of documents in the search).[21]

The extent to which relevant evidence can be identified by searching databases depends on factors such as terminology, abstracting, format of research question and type(s) publication bias present in a particular field.[16] Delaney & Tamás (2018) discuss issues that they have encountered in their own work in the low consensus in the field of international development, and argue that the reliance on database searching in the absence of stable terminology, standard abstracting and standard research questions should be more critically considered. Examples of low consensus fields include political science and sociology that distinguish themselves by an openness to diverse perspectives. Physics and health sciences are examples of high consensus field. For health sciences the high consensus is more applicable for information retrieval related to clinical efficacy or safety from RCTs, but other health science topics, such as input parameter identification for health economic models, can be impacted by the same issues as the low consensus fields described by Delaney & Tamás. There are inconsistencies in terminology (e.g. utilities), increasing variations in study design that are associated with less standardised reporting structures and variations in format of evidence sources, and publication formats that are not standardised (e.g. real-world sources, unindexed grey literature).[73]

Additional information seeking methods and practices have also been described: berry-picking,[20] citation tracking,[21,22] author/expert consultation,[22,23] and grey literature searching.[24] some of these can also form part of systematic searching, especially grey literature searching.[4] However, the empirical evidence assessing the value of different information retrieval methods is limited.[23,25,74,75] This is also true for the identification of health economic model input parameters where an evaluation should not only consider the indicators of precision/specificity and sensitivity/recall but also indicators of impact on model conclusions. From a practical perspective, it is not possible to conduct full Cochrane style systematic searching for all model input parameters and other information needs of the model. Rather, it has been argued that search activities for a model should focus on searching for the aspects of the model that have most impact on the model results, putting the emphasis on identifying sufficient information to understand the implications of uncertainty associated with the health economic model.[26]

In 2012, Paisley conducted a literature review on information retrieval methods in HTA as part of her PhD study.[1] Paisley categorised the selected literature and identified two different themes deemed to be relevant to the development of search methods for economic models: 'efficient approaches to searching' and 'searching for specific types of information'. Paisley recognised the time and resource costs of undertaking searches for models as an important issue.[1,8] Of the studies identified, one study reported on a high precision RCT filter.[76] Other publications reported on how search techniques such as search filters, limiting number of databases searched and database facilities can be used to alter the levels of sensitivity and precision.[77] Paisley found that relatively little evidence is published on the implications of doing less than comprehensive searching and that most studies focus on identifying clinical effectiveness studies. However, health economic models require data from different study types (not just RCTs), and from different sources (not just bibliographic databases). The review found no publications on established

search methods for a whole range of evidence needed for the models. Exceptions included some studies that were found to report on approaches to identify information other than RCTs, such as observational studies, information on AEs, and information from routinely collected data sources.[78-80]

2.4.2 Information Retrieval for Health Economic Models

The application of systematic review search methods for the identification of model input parameters is associated with practical and theoretical issues. In practice, it is often not possible to allocate sufficient time and resource to carry out a systematic literature review on identification of all model input parameters. This is because models address real-world decision problems and results will be used for decision-making, and therefore models are associated with vast information requirements. Additionally, data sources are not limited to research-based sources and include non-research-based sources, for example registries, administrative or routine data sources and expert opinion.[26] It may not be possible to assess the completeness of these data sources, and in fact, it may not even be appropriate if resources are spent processing large volumes of irrelevant information at the expense of other analytical activities. It may be more useful to focus on those model inputs that have the most impact on the model outputs. Table 1 (on page 26) summarises some of the key conceptual information retrieval differences between models and systematic review search methods. As part of her PhD thesis, Paisley identified only one publication on search methods for health economic models.[1] This is a study by Golder *et al*. on the feasibility and efficiency of undertaking systematic searching to populate health economic models.[8] The publication is a case study of the prophylactic use of antibiotics in children to prevent urinary tract infections. The study identified 42 information requirements, that can be grouped into five categories (treatment effects, health-related quality of life [HRQoL], resource use and unit costs, baseline event rates and antibiotic resistance). The study used 18 search strategies that were aimed at high precision rather than high sensitivity. The study demonstrated that health economic models are associated with multiple information needs and that these relate to different types of information. Paisley also identified a relevant publication by Philips *et al*. that reviewed 15 published health economic modelling good practice guidelines, including guidance on data.[81] In this study, content relating to data identification was summarised. The guidelines were found to make general statements about needing to identify data for the models in systematic and transparent manner but included none or little procedural guidance beyond identification of treatment effects. Paisley (2016) proposes minimum requirements for identifying evidence for key parameters, contributing to research on prioritisation of efficiency versus sensitivity. So far, these minimum requirements have not yet been tested empirically. Hence, it is not known to what extent there is a trade-off between sensitivity and efficiency of the search. This lack of knowledge of the potential trade-offs makes it challenging for researchers to have the confidence to rely on alternative approaches, especially in the context of reimbursement submissions to HTA agencies.

2.5 Overview of Data Sources

This Chapter gives an overview of various data sources (bibliographic databases, specialist databases and other types of searching), that are being used to identify inputs for health economics models.

2.5.1 Bibliographic Databases

2.5.1.1 General Bibliographic Databases and Filters

Medline and Embase are electronic bibliographic databases that are often used in health science research and also in HTAs. They have been recommended as minimum sources by the Cochrane Collaboration and the National Institute for Health and Care Excellence (NICE).[5,32,44] Additionally, Paisley (2016) also recommends using at least one of these bibliographic databases as a minimum search requirement for health economic models.[10] Search filters can be used to identify specific types of information or study design in these bibliographic databases. Search filters can be defined as search strategies that include a series of pre-elaborated free text terms/text words/phrases and subject headings for a given concept, idea, or study design.[82] This is done by restricting findings by using a combination of keywords. A collection of search filters has been compiled by the InterTASC Information Specialists Subgroup (ISSG).[83] The HEDGES project has also developed a broad range of search filters aimed at maximising sensitivity and precision.[84] Searching bibliographic databases to identify data for health economic models are generally the easiest and least time-consuming way to identify a set of relevant reports/studies, especially those reported in RCTs (treatment effect, AEs). The databases can be searched electronically, using words in the title or abstract. It is also possible to use controlled vocabulary. If carefully reported, searching bibliographic databases is replicable, increasing the transparency of how model inputs have been searched and selected.

2.5.1.2 Specialist Bibliographic Databases

Specialist databases also exist that are relevant for the identification of input parameters for health economic models. For searching RCTs and systematic reviews, the Cochrane Library can be used for identifying Cochrane systematic reviews and controlled clinical trials.[85] The CRD Database of Abstracts and Reviews of Effectiveness (DARE) also has systematic review abstracts, but only dating from 1994 to 2015 as it is no longer updated.[86] Other commercial producers, such as KSR Evidence, now provide access to systematic reviews in healthcare.[87] AEs can be identified from generic bibliographic databases (Medline, Embase). However, some specialist databases such as Derwent Drug File or International Pharmaceutical Abstractsexist.[88-90] Specialist databases for identifying economic evaluations include national health service Economic Evaluation Database (NHS EED) produced by the CRD (not updated since 2015),[86] the HTA database,[91] and the Tufts cost-effectiveness analysis (CEA) Registry.[92] There are no specialist bibliographic databases for identifying resource use, but the above economic evaluation databases may be useful. The Tufts CEA Registry[92] can also be searched for utilities data.[93]

2.5.2 Non-database Searching

In addition to database-based information accessed via generic and specialist databases, health economic models almost always require information from non-database based sources, such as trial registries (relative treatment effects), regulator websites (treatment effects, AEs), national drug cost lists (costs), health/social care service unit costs (resource use), hospital statistics (costs, resource use), and national statistics (baseline risk of clinical events). For cost data, it is especially important to use evidence from the

relevant jurisdiction(s). An example of a non-database source is the NHS Payment Scheme that can provide costs for health economics models. These costs can be accessed through a website (https://www.england.nhs.uk/publication/2023-25-nhs-payment-scheme/). For other model input parameters non-database source searching is also recommended by Paisley but may not be possible.[10] For searching cost input parameters for health economic models, local knowledge/expertise is required due to a wide range of non-database based sources being available. This is because the non-database sources for health economic models can be highly specific to each jurisdiction and local language skills may be required to explore them fully. Clinical guidelines may be consulted for resource use information to get an understanding of clinical practice. It is also crucial to evaluate if there is a discrepancy between the recommended practice and real-world practice in a given area. Information on baseline risk of clinical events can be searched from national statistics/administrative data, as well as disease/epidemiology registers and surveys. Many models require mortality data. If disease-specific mortality is not available/applicable, nationally collected statistics such as life tables are useful additions in models. These can be accessed e.g. through the WHO website.[94]

2.6 Summary

Health economic models are associated with multiple and complex information requirements; the type of information that needs to be identified is varied and can be in a non-standard format. All this complexity has to be managed with multiple search strategies, although a single search could potentially identify evidence for more than one information need. In this chapter the issue of identifying the optimal data for health economic models has been explored. Also, an overview of search methods and potential data sources has been provided. Techniques for retrieving information, designed for high precision, can be employed to streamline searches and effectively handle various informational demands. This chapter also considered what might be the most relevant model parameters in this context. The development of the model will utilise diverse information types for its inputs. Therefore, multiple sources are searched through various information retrieval strategies and techniques. The intended use of the evidence may dictate the information type, and the extent of information retrieval required. However, it is still impractical and unnecessary to uncover all evidence for every informational need in the model.

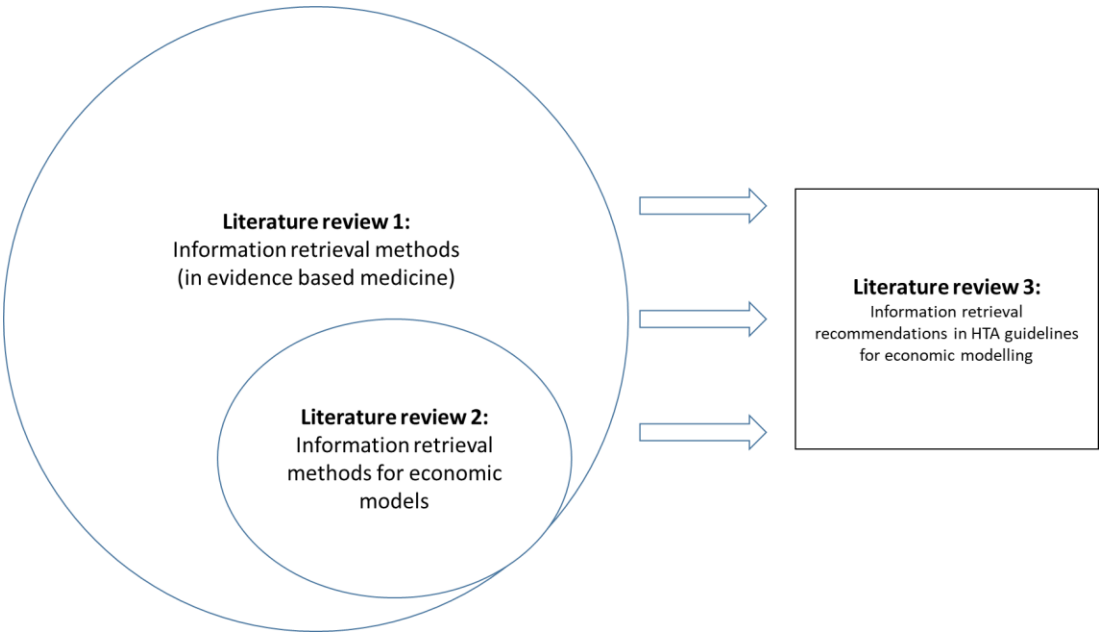
In the next chapter (Chapter 3) the published literature is reviewed in order to understand what information retrieval methods are available in evidence-based medicine generally, with the view of exploring them in the context of identifying health economic model parameters during the course of the PhD study. Further, Chapter 3 reviews health economic modelling and HTA guidance to gain an understanding of what search methods, if any, are recommended for the various key model input parameters.

3 Narrative Review of Existing Literature

3.1 Chapter Overview

This chapter presents three narrative literature reviews that each support the thesis. The first review looks at information retrieval methods within evidence-based medicine. The second review looks at what evidence or recommendations specifically exist related to searching for data for health economic models. The third review looks at if and how these recommendations are reflected in the health economic modelling guidance documents, such as those published by HTA agencies or professional agencies. The first two reviews cover how information retrieval can (at least in theory) be carried out in the context of evidence-based medicine (and within that, models), with Review 2 being a subset of Review 1 (see Figure 1). Review 3 examines how the information retrieval methods identified in Reviews 1 and 2 have been recommended in HTA guidelines for economic modelling. The aim is to form a picture of how well they correspond. The objective is to summarise the key methods, themes and publications, and to place the PhD study in the context of the existing literature in this area.

Figure 1. Relationships between the three literature reviews undertaken



The objectives and the scope of these three literature reviews are presented in Sections 3.2.1 (Reviews 1 and 2) and 3.3.1 (Review 3). The review methods are detailed in Sections 3.2.2 (Reviews 1 and 2) and 3.3.3 (Review 3). The results are presented in Sections 3.2.4 (Review 1), 3.2.5 (Review 2) and 3.3.4 (Review 3). Finally, Section 3.4 includes the summary of these narratives of relevant literature.

3.2 Literature Review on Information Retrieval Methods

3.2.1 Objectives and Scope

These two literature reviews aim to answer the following research questions:

- W*hat search methods are available to identify literature in evidence-based medicine (Review 1)?*
- W*hat search methods are specifically available to identify inputs for health economic modelling (Review 2)?*

The aim of these two reviews is to establish an overview of what official guidance has been produced by recognised review producing organisations and what additional methodological advice has been published in peer reviewed journals by authors with experience conducting literature reviews. The aim is to widely explore literature relating to searching in evidence-based medicine (Review 1) and specifically as they relate to health economic modelling (Review 2).

3.2.2 Search Methods

These two reviews can be described as narrative reviews, rather than as systematic reviews, loosely following methods described by Gasparyan *et al*. 2011 and Whittemore *et al*. 2005.[95,96] To identify available search methods guidance documents, multiple search methods were used:

- Reviewing the PhD thesis by Suzy Paisley.[1]
- Hand searching and browsing journal websites for the last five years (see Table 2).
- Searching websites of organisations producing guidelines for literature reviews (see Table 3).
- Searching Medline (see Table 4).
- Reference checking of relevant identified publications and authors from above steps 1 - 4.

Table 2. List of journals hand-searched

Journal	Website
BMC Medical Research Methodology	https://bmcmedresmethodol.biomedcentral.com/
Health Information and Libraries Journal	https://onlinelibrary.wiley.com/journal/14711842
International Journal of Technology Assessment in Health Care	https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care
Journal of Clinical Epidemiology	https://www.jclinepi.com/
PharmacoEconomics	https://www.springer.com/journal/40273
Research Synthesis Methods	https://onlinelibrary.wiley.com/journal/17592887
Systematic Reviews	https://systematicreviewsjournal.biomedcentral.com/
Value in Health	https://www.valueinhealthjournal.com/

Table 3. Guidance producing websites searched (from Sutton *et al.* 2019[97])

Organisation	Website
Agency for Healthcare Research and Quality (AHRQ)	https://www.ahrq.gov/
Campbell Collaboration	https://campbellcollaboration.org/
CADTH (Canadian Agency for Drugs and Technologies in Health)	https://www.cadth.ca/
Centre for Health Economics Policy Analysis, McMaster University	https://chepa.mcmaster.ca/
Centre for Reviews and Dissemination (CRD), University of York	https://www.york.ac.uk/crd/
Cochrane	https://www.cochrane.org/
Department of Health (UK)	https://www.gov.uk/government/organisations/department-of-health-and-social-care
Economic and Social Research Centre (ESRC)	https://esrc.ukri.org/
EUnethta (European Network for Health Technology Assessment)	https://www.eunethta.eu/
Healthcare Improvement Scotland	http://www.healthcareimprovementscotland.org/
Health Economics Research Unit (HERU), University of Aberdeen	https://www.abdn.ac.uk/heru/
Health Technology Assessment International (HTAi)	https://htai.org/
INAHTA (The International Network of Agencies for Health Technology Assessment)	https://www.inahta.org/
Institute for Clinical and Economic Review (ICER)	https://icer-review.org/
ISPOR (International Society for Pharmacoeconomics and Outcomes Research)	https://www.ispor.org/
Joanna Briggs Institute, The University of Adelaide	https://joannabriggs.org/
National Collaborating Centre for Methods and Tools	https://www.nccmt.ca/
National Institute for Health Research (NIHR)	https://www.nihr.ac.uk/

Table 4. Medline (via ProQuest Dialog) search strategy

#	Search terms	Hits
1	MESH.EXACT("Information Storage and Retrieval")	20446
2	MESH.EXACT("Databases, Bibliographic")	6048
3	MESH.EXACT("Abstracting and Indexing")	4699
4	(information retrieval).kw	957
5	(literature search).kw	196
6	1 or 2 or 3 or 4 or 5	30226
7	MESH.EXACT("Decision Making")	97966
8	MESH.EXACT("Evidence-Based Practice")	10662
9	MESH.EXACT("Evidence-Based Medicine")	73959
10	MESH.EXACT("Technology Assessment, Biomedical")	10249
11	MESH.EXACT("Models, Economic")	10471
12	7 or 8 or 9 or 10 or 11	197925
13	6 and 12	1498
14	limit 13 to yr="2012 -Current"	332

Hand searching of journals and websites was carried out between May and August 2020. The latest document was retrieved on 25th August 2020. The list of journals was determined together with the PhD supervisors and the list of websites to browse was taken from a publication by Sutton *et al.* 2019.[97] A bibliographic database search was carried out in Medline database (via ProQuest Dialog) on the 17th August 2020. The database was searched since 2012 to August 17th 2020. The time limit of 2012 was

applied, as this is the year of Suzy Paisley's PhD thesis and it reported on a similar literature review of search methods.[1] A time limit was applied to the bibliographic database search, but from the website searches the latest documents from each website were retrieved even they were first published before 2012. The reference lists of relevant, identified publications were hand searched and further publications by key authors followed to identify additional publications.

Inclusion and exclusion Criteria

The following types of studies were included: advice/guidelines for searching in evidence-based medicine (review 1) or health economic models (review 2). Additionally, publications covering approaches to searching, development of search strategies, search filters and reporting standards were all of interest. This review excluded guidelines on 1) case studies on searching, 2) comparisons between bibliographic databases or search platforms or filters, and 3) descriptions of information resource development. Studies were also excluded if they did not provide new information for the research questions. Inclusion and exclusion criteria are summarised in Table 5.

Table 5. Inclusion and exclusion criteria

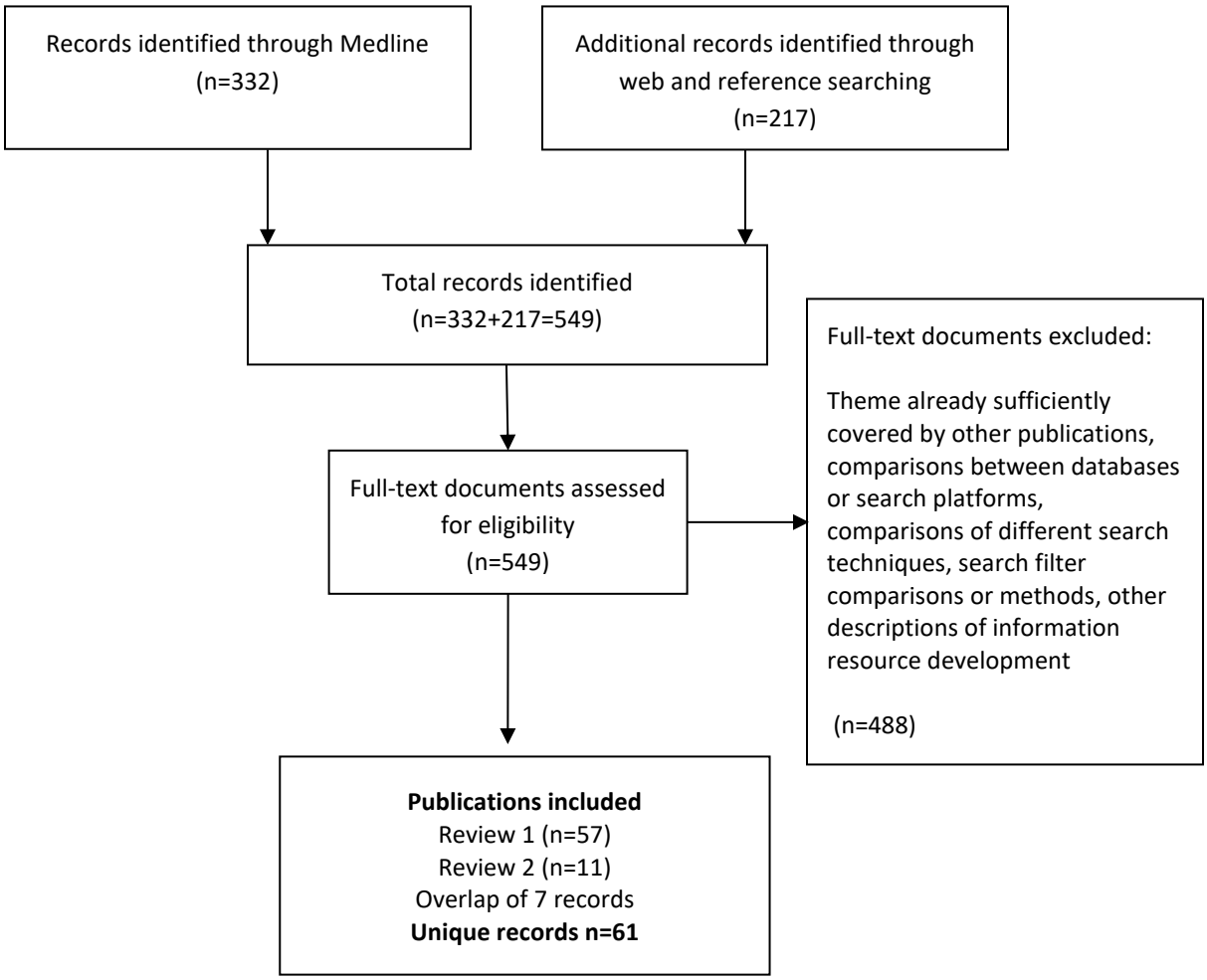
Inclusion criteria
Advice, guidance or current practice on searching in evidence-based medicine (Review 1)
Advice, guidance or current practice on searching for economic models (Review 2)
Publications covering/describing approaches to searching (e.g., citation searching, automated retrieval methods, hand-searching)
Development of search strategies
Development of new search filters
Reporting standards
Exclusion criteria
Case studies on searching
Comparisons between databases or search platforms (e.g., Medline and Embase)
Other information related to search filters, such as (comparative) critical appraisal of existing search filters, surveys of filter performance
Other descriptions of information resource use development

The eligible documents were entered into EndNote Reference Management software and the full text of each document was included in the database. A data extraction grid was developed in Microsoft Excel and was focused on extracting information on three components. Firstly, general bibliographic details, such as year of publication, type of document, authors or agency, etc. Secondly, search methods and approaches to searching. Finally, recommendations on how to search by health economic model input type. A pilot test of three documents was carried out to ensure the extraction sheet was adequate for the task. A descriptive analysis across key model inputs was taken. This was focused on what search methods have been recommended by review type (review 1) and by model input type (review2).

3.2.3 Results

One search was run for both of the reviews (reviews 1 and 2). The search resulted in 332 records potentially includable for either Review 1 or 2. As the records were reviewed, they were separated into those relevant for Review 1 (reported in Section 3.2.4) and those relevant for Review 2 (reported in Section 3.2.5). In total 488 records were excluded as not relevant for either Review 1 or 2, based on the title/abstract. Therefore, 61 publications were considered to be relevant for either Review 1 (n=57) or Review 2 (n=11). Seven records were included in both reviews. This literature review does not discuss all the papers identified, rather the aim is to identify the different search approaches in the literature, and to group similar approaches. Figure 2 displays the PRISMA diagram. The list of 61 included publications is displayed in Appendix 1 (page 268).

Figure 2. PRISMA chart – Literature reviews 1 and 2



3.2.4 Review 1: Results: Information Retrieval Methods in Evidence-based Medicine

3.2.4.1 Guidance Documents

Eleven guidance documents from ten recognised organisations that either produce or commission reviews were identified (Table 6). These documents provide methodological guidance for the conduct of different types of reviews, including methods guidance on searching. The searching guidance is given for the different types of reviews, namely those that are relatively established, such as systematic reviews, rapid reviews, qualitative reviews, and mixed method reviews. The focus in Table 6 is on how the documents relate to literature searching (indicated in the last column of the table). Sutton *et al*. 2019 provide a recent exploration and grouping of reviews type families, that have been utilised in this report.[97] The review families have been defined by Sutton *et al.* and those relevant for this report are shown in Table 7. The rest of the definitions for other review types can be found in Table 3 on page 206 of Sutton *et al.* 2019.

Table 6. Guidance documents from recognized organizations producing reviews

Organisation	Publication	Search guidance by review type
AHRQ	Relevo R, Balshem H. Chapter 5: Finding Evidence for Comparing Medical Interventions. Methods guide for effectiveness and comparative effectiveness reviews Agency for Health Care Research and Quality (AHRQ); 2014.[98]	Comparative effectiveness reviews
Campbell Collaboration	Kugley S, Wade A, Thomas J, *et al*. Searching for studies: a guide to information retrieval for Campbell systematic reviews. Campbell Systematic Reviews 2017; 13(1): 1-73.[99]	Systematic reviews
CRD	Centre for Reviews and Dissemination University of York. Centre for Reviews and Dissemination. Systematic reviews: CDR's guidance for undertaking reviews in health care. York, 2009.[4]	Systematic reviews
Cochrane	Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.0 [Updated July 2019]. 07/2019 2019. https://training.cochrane.org/handbook/current (accessed 25/08/2020 2020).[5] Harris JL, Booth A, Cargo M, *et al*. Cochrane Qualitative and Implementation Methods Group guidance series-paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. Journal of clinical epidemiology 2018; 97: 39-48[100]	Systematic reviews Qualitative reviews
EUnetHTA	EUnetHTA. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness: European network for Health Technology Assessment, 2019.[101]	Systematic reviews
ISPOR	Hoaglin DC, Hawkins N, Jansen JP, *et al*. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2011; 14(4): 429-37. [102]	Network meta-analyses

Joanna Briggs Institute	Aromataris E, Munn Z (Editors). JBI Manual for Evidence Synthesis. JBI, 2020.[103]	Different types of systematic reviews, such as qualitative systematic reviews, Effectiveness systematic reviews, Review of economic evaluations, and Scoping reviews, among others
NCCMT	Dobbins M. Rapid Review Guidebook. Hamilton, ON, Canada: National Collaborating Centre for Methods and Tools, 2017[104]	Rapid reviews
RAMESES	Greenhalgh T, Wong G, al. e. The RAMESES Project. 2013-2020. http://www.ramesesproject.org/Standards_and_Training_materials.php.[105]	Meta-narrative review Realist Synthesis
SuRe Info	SURE Vortal; H. 2019. http://vortal.htai.org/?q=sure-info[106]	Broad, including: Cost and economic evaluation. (SuRE Info). 2020[107]

Abbreviations: AHRQ: Agency for Healthcare Research and Quality ; CRD: Centre for Reviews and Dissemination, University of York; EUnetHTA: European Network for Health Technology Assessment; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; NCCMT: National Collaborating Centre for Methods and Tools; SuRe; Summarized Research in Information Retrieval for HTA; RAMESES: Realist And Meta-narrative Evidence Syntheses: Evolving standards

Table 7. Review type definitions (from Sutton *et al*. 2019[97])

Type of review	Search methods with key publications
Systematic reviews	○ 'Seeks to systematically search for, appraise and synthesis research evidence, often adhering to guidelines on the conduct of a review'
Reviews of reviews	○ '*summary of the [medical] literature that attempts to survey the literature and describe its characteristics*' ○ '*are intended primarily to summarize multiple Cochrane Intervention reviews addressing the effects of two or more potential interventions for a single condition or health problem. In the absence of a relevant Cochrane Intervention review, Cochrane Overviews may additionally include systematic reviews published elsewhere*.'
Rapid reviews	○ *'a type of knowledge synthesis in which components of the systematic review process are simplified or omitted to produce information in a short period of time'*
Qualitative reviews	○ *'Method for integrating or comparing the findings from qualitative studies. It looks for "themes" or "constructs" that lie in or across individual qualitative studies'*
Mixed method reviews	○ *'any combination of methods where one significant component is a literature review(usually systematic). Within a review context it refers to a combination of review approaches for example combining quantitative with qualitative research or outcome with process studies'*
Purpose specific reviews	○ *'Synthesis method used to identify concepts, viewpoints or ideas. Focuses on identifying the defining attributes of the concepts and can be used to develop a synthesis model.'*
Realist Review	○ *'Answers the question "What works for whom under what circumstances?" rather than "What works?". Specifically, it seeks to 'unpack the mechanism' of how complex programmes work (or why they fail) in particular contexts and settings'.*

3.2.4.2 Methodological Advice

In addition to the above guidance documents for specific types of literature reviews, methodological advice, current practice and case study documents were identified for a wide variety of search-related topics. Some publications focus on methodological advice for specific types of review, either complimenting the guidance documents listed in the previous section or in some cases (especially for newer/less established review types), providing recommendations in absence of any guidance. These are summarised in Section 3.2.4.2.1. Additionally, the publications cover approaches to searching (e.g., citation searching, automated retrieval methods, hand-searching), developing search strategies (e.g., sensitivity versus precision or structuring the search), search filters (e.g., features and how to identify/choose one) and reporting standards (Section 3.2.4.2.2). Reporting standards are reported in Section 3.2.4.3. Methodological advice also exists on searching for specific aspects of HTA such as clinical effectiveness, safety, costs and economic evaluations (Section 3.2.5.1). The remainder of this chapter is organised to cover these areas under the relevant subheadings.

3.2.4.2.1 Methodological Advice by Literature Review Type

Table 8 provides identified methodological advice and current practice on search methods, by literature review type.

Table 8. Methodological advice or current practice documents/sources

Review type	Search methods with key publications
Systematic reviews	○ **Established search methods** covered by guidance documents (Table 6): ○ Searching is characterised by comprehensive approach and includes bibliographic database search in more than one database ○ Often focused on particular types of studies or study designs ○ Grey literature searches should be included but guidance is sparse ○ Further search methods include hand searching, web searching, reference list checking citation searching and contact with experts ○ Some **further case studies** exist e.g. searching for network meta-analyses[108] ○ **Living systematic review** (defined as "an approach that aims to continually update a review, incorporating relevant new evidence as it becomes available" by Simmons *et al*.[109]) is a relatively new method, and often considered a sub-type of systematic literature review. Elliott *et al* 2017 propose a new approach for updating systematic literature reviews through "living systematic review" to address the challenges for evidence synthesis while retaining the strengths of traditional systematic reviews.[110] The Cochrane Living Systematic Review Network has published advice for conducting living systematic reviews.[111-114] Millett reports on feasibility and acceptability of living systematic reviews, including challenges that need to be addressed.[115]
Reviews of reviews	○ Focuses on **one study type: systematic reviews** ○ Searching is focused on databases indexing systematic reviews and/or using review filters to search bibliographic databases ○ Grey literature searches are also recommended e.g. searching PROSPERO for prospective reviews ○ Further search methods include e.g. reference list checking ○ An **overview of methods** given in presentation by Wright & Walwyn 2016[116]
Rapid reviews	○ Rapid reviews abbreviate/deviate from conventional systematic reviews to deliver them within the specified scope/timeline

	○ Some forms of rapid reviews abbreviate the search process and others abbreviate other elements of the process ○ Search **methods depend on agreed methodology**. Methods and implications of rapid reviews are given by Ganann *et al.*[117]

Qualitative reviews	○ Qualitative review is a **broad term** for a group of methods (such as thematic synthesis, meta-ethnography, critical interpretive synthesis, framework synthesis, and meta-narrative) used to undertake systematic reviews of qualitative evidence
	○ Cochrane Qualitative and Implementation Methods Group (Harris *et al*. 2017) has published a paper on methods for question formulation, **searching and protocol development for qualitative evidence synthesis**.[100] Published a year earlier, Booth *et al*. 2016 provides an overview to **help navigate the methodological choices** associated with qualitative reviews, including appropriate reporting standards (ENTREQ and RAMESES).[118]
	○ Sutton *et al*. 2019[97] provides useful considerations when choosing appropriate search methods for qualitative reviews:
	  ○ Is the review aggregative or interpretative?
	  ○ Is theory expected to play an important role?
	  ○ Are differences in context likely to be important?

Mixed method reviews	○ Mixed method reviews either aim to identify studies that include both qualitative and quantitative types of data or identify studies which either include **quantitative or qualitative data** and integrate these in the mixed method review. Search approaches include use of filters, generic terms to retrieve specific types of studies, or running a broad search without filters.[97] ○ Methods for rapid realist review are published by Saul *et al* 2017[119] ○ The RAMESES project gives some methods for searching in the context of realist reviews[120] and Pawson has published methods with colleagues.[121] ○ A recent publication by Booth, Briscoe and Wright provides a six-component framework for realist review[122] ○ Noyes *et al*. 2018 aims to clarify how to integrate quantitative and qualitative evidence.[123] The publication concludes that mixed methods can be **particularly helpful in understanding how complexity impacts** health technologies in particular settings. ○ Mixed method reviews often **follow up index papers** to find related papers. This process has been formalised in the CLUSTER procedure.[124]
Purpose specific reviews	○ Sutton *et al*. define "purpose specific reviews" as a **heterogeneous group of review types and methods**.[97] Multi-question systematic reviews for HTA would represent an example of this. Many HTA agencies produce their own guidance for searching that are often based on the guidance documents listed in Table 6 (Section 3.2.4.1). ○ **Guidelines exist for clinical effectiveness review and reviews of economic studies**, e.g. from EUnetHTA or from Summarised Research in Information Retrieval or HTA (SuRE info HTAi IRG) (see Table 6). ○ Sutton *et al*. 2019 emphasize that the most effective search process is determined by aligning the review's goals with the types of studies and samples needed. **No methodological guidance exists**, and adapting an established method can be considered, such as realist synthesis, meta-ethnographies or qualitative evidence syntheses.[97]
Other	○ Smythe and Spence re-visit an alternative literature review method which recognises that there is **no final understanding of the relevant literature**, but rather it is a hermeneutic process and requires constant re-interpretation that leads to deeper and **more comprehensive understanding of relevant literature**.[125] The hermeneutic literature review gets away from the **assumption that there is only "*one way to do a literature review*"**. ○ Greenhalgh & Shaw 2017 provide **a case study** in application of telehealth in the management of heart failure **to make sense of complex literature** by using hermeneutic systematic review.[126]

3.2.4.2.2 Approaches to Searching

Bibliographic Database Searching in Systematic Literature Reviews

Bibliographic database searches are recommended for Cochrane style systematic literature reviews, especially in the context of evaluating if health technologies work as well as in the assessment of the potential magnitude of that benefit. Study designs used to assess treatment effect include RCTs and observational studies (see Section 3.2.5.1.1). Searching bibliographic databases for systematic literature reviews are described in guidance documents published by e.g., the Campbell Collaboration, CRD, Cochrane, EUnetHTA and Joanne Briggs Institution.[4,5,99,101,103] Searching observational studies (e.g., for baseline risk of clinical events) poses some specific challenges due to poor and inconsistent indexing in databases.[127] A study published 2019 by Li *et al*. evaluated the sensitivity and precision of search strategy that incorporated a filter to identify observational studies in Medline and Embase.[128] They identified eighteen methodological filters across two eligible studies. The first study's filters, targeting observational studies, had higher sensitivity and precision and were externally validated. The second study's filters, focused on comparative non-randomised studies and had lower sensitivity and precision. The filters were not externally validated. Due to limited and heterogeneous evidence and methodological limitations, the authors of the publication called for further research and improved indexing. Haynes et al. 2005 have published "optimal Medline" (i.e. precision-maximising) search methods, where trade-offs between sensitivity, specificity, precision and accuracy were looked at.[129] Their work validated optimised retrieval for researchers who want to retrieve little non-relevant material.

Citation Searching

Citation searching uses a technique where the citations surrounding a particular publication are used to identify further similar publications. Many of the documents identified in Sections 3.2.4.1 and 3.2.4.2 recommend citation searching as a supplementary search method. Both direct and indirect citation methods exist. Direct citation can take the form of forward and backward searching.[4,130] A study by Briscoe *et al.* (2020) found that carrying out backward citation searching was more consistent than carrying out forward citation searching.[131] The authors conclude that this is likely to be due to paucity of practical advice on how to conduct forward citation searching. Searching for the direct citations is a productive search method only if the studies often cite all related work, therefore creating a single network of citations.[132] However, this is often not the case.[133] Janssen and Gwinn (2015)[134] developed a search method to identify related publications using the principles of co-citation. A study by Briscoe *et al.* (2020) found that carrying out backward citation searching was more consistent than carrying out forward citation searching.[131] The authors conclude that this is likely to be due to paucity of practical advice on how to conduct forward citation searching. Searching for the direct citations is a productive search method only if the studies often cite all related work, therefore creating a single network of citations.[132] However, this is often not the case.[133] Janssen and Gwinn (2015)[134] developed a search method to identify related publications using the principles of co-citation.[135] This method uses, not only direct forward and backward citations, but also indirect citations. For example, publications may not cite each other directly (no direct citation) but both could be citations in a new review publication (indirect citation). Between 50% and 88% of included studies were retrieved by using this co-citation method. Since the conduct of

this review, a new guidance on terminology, application, and reporting of citation searching (TARCis) statement)has been published.[136]

Automated Retrieval Methods

Information retrieval for systematic reviews has recently been associated with development of tools and automation technologies. O'Mara-Eves *et al*. (2015) systematically reviewed the literature on text mining for study selection/identification.[137] They found that cutbacks in workload are likely achievable.[137] The overall workload reduction was predicted to be between 30% and 70%, although sometimes this was associated with the loss of 5% of relevant studies (i.e. 95% recall). Therefore, the authors conclude that this method may be considered safe for living reviews, but the use of text mining as a second reviewer should be done with caution and that using text mining to automatically eliminate studies needs more investigation. Some of the study selection tools include Abstrackr, Colandr, DistillerAI, EPPI-Reviewer, Rayyan, and RobotAnalyst.[138-143] A more recent publication by Stansfield *et al.* (2017) found that text mining tools can be used in five different ways for searching: "*improving the precision of searches; identifying search terms to improve search sensitivity; aiding the translation of search strategies across databases; searching and screening within an integrated system; and developing objectively derived search strategies*".[144] A case study is provided by Shemilt *et al*. (2014).[145] Shemilt and colleagues described methods for applying and evaluating text mining technologies to reduce the screening workload in extremely large scoping reviews of public health evidence. This case study demonstrated how text mining could prioritize records for manual screening, significantly reducing the manual screening workload. No guidance on the topic exists but the Campbell Collaboration recommends considering employing the use of text mining for going through search results.[99] The authors conclude that there is a narrow and heterogeneous evidence base, and they recommend advancing research in this area. Additional studies highlight the benefits of text mining tools in developing search strategies for systematic reviews. Stansfield et al. (2017) identified key uses such as enhancing search precision and aiding cross-database strategy translation.[146] McGowan (2021) discussed improving objectivity and reproducibility by revealing high-frequency terms and correlated words.[147] Since the conduct of this review, automation of literature reviews has continued be focus of on-going research and development. A recent breakthrough in regenerative artificial intelligence has opened new possibilities for faster and more efficient literature review process.[148]

Iterative Searching

Sutton *et al*. (2019) collated information on iterative search methods.[97] The authors state that iterative searching is increasing in importance for qualitative and purpose-specific literature reviews. They also state that currently there is no common definition or methodology. Iterative searching is characterised by reoccurring cycles of information retrieval and evaluation, with the possibility of returning to previous steps as judged appropriate by the reviewer. The iterative process can be repeated as many times as needed to reach saturation when no further relevant evidence is being found. None of the existing publications provide practical steps for describing how to carry out iterative searching. Finally, Sutton and colleagues finish by asking a question whether the berry picking technique, defined by Bates in 1989[20], is synonymous to iterative searching.

Web Searching

The identified publications provided limited guidance on web searching, with the CRD Handbook making a separation between an internet search that uses a search engine and internet search that browses relevant websites.[4,99] The Campbell Handbook gives advice on exploring a search engine with some guidance how to undertake the searches.[99] The functionality and non-structured nature of search engines and websites can present challenges.[149] Briscoe, Nunns and Shaw (2020) provide a recent study to characterise and appraise current practice of web searching.[131] The study found that web searches were more simply structured than bibliographic searches and that advanced methods were not broadly used. The most popular search engines were Google Scholar and Google Search, but there is potential to employ a wider array of search approaches in a broader selection of search engines.

Search Filters

Search filters are collections of search terms created to retrieve selections of records from a specific bibliographic database.[150] Search filters can be designed to identify records for a specific study type or topic. The choice of search filter depends on several factors. Some guidance documents recommend specific filters, for example, the Cochrane Handbook recommends highly sensitive search filter for identifying RCTs.[130] Filters can also be used for retrieving economic and cost studies.[107] This literature review identified recent publications on search filters, including study by Waffenschmidt *et al*. 2020 on development and validation of study filters for identifying controlled non-randomised studies in PubMed and Medline,[151] study by Taljaard *et al*. 2020 on search filter development to identify pragmatic trials in Medline,[152] and a study on optimal development of search strategies in PubMed for the purpose of finding treatment predictors.[153] An online resource by ISSG provides a useful collections of search filters.[35] The ISSG resource was updated in October 2020 and it was cross-checked against the findings from this literature reviews. The finding was that all search filter related publications identified in this literature review can also be found in the ISSG resource.[83]

3.2.4.3 Reporting Standards

Transparent and clear reporting is an important aspect of literature review. Documenting the search is important for establishing transparency and reproducibility, and also to facilitate future updates of the search.[154] The following identified publications contained reporting standards: The Campbell Collaboration Guide,[99] the Cochrane Handbook,[130] CRD,[4] ENTREQ,[155] the EUnetHTA guideline for information retrieval,[101] MECIR from Cochrane,[156] the PRISMA checklist,[157] and RAMESES.[158] Of these, the ENTREQ and RAMESES are specifically developed for reporting of qualitative systematic reviews. Booth (2006) has proposed STARLITE as a method for documenting and reporting search methodologies for systematic searches, but this is yet to be widely adopted.[159] PRISMA-S is a complement to the PRISMA statement, offering a checklist that could be used to verify that each of the components of the search is complete and reproducible.[160]

3.2.5 Review 2: Results: IR Methods for Health Economic Models

Paisley (2012) found in her PhD thesis that modelling literature provides little information on formal search methods. Technical Support Document (TSD) 13 for NICE provides details on how to search for and review evidence for models submitted to the NICE appraisal process.[26] In her PhD thesis Paisley writes that she was aware of the study by Golder *et al.* (2005) prior to starting her PhD project, and after searching she was not able to identify further publications.[8] Golder *et al.* (2005) identified a number of issues related to search methods for models. Some of these issues included the range of information and types of evidence required for the models, data quality issues and internal/external validity, and considerations on efficient approaches for searching. Glanville & Paisley (2010) reported that health economic model searches can be less exhaustive and more targeted than review searches since their purpose is to identify adequate evidence to populate the model, rather than retrieve all studies.[47]

This update search identified sparse guidance and methodological advice on how to search for evidence for all the main model input types or for a specific one, such as utilities.[73,158,161,162] The details of searching for a specific model input type are summarised in Section 3.2.5.1 below. Paisley (2016) has suggested minimum searching levels for each model parameter type but notes that these have not yet been tested empirically.[10] In her PhD thesis, Paisley (2012) concludes that the development of search methods for models is recognised as an area of further research and while few additional publications have now been identified, this area remains an area where further research is required.

3.2.5.1 Search methods by data type

This Section summarises the search methods by data type, as relevant for health economic modelling, although these methods are not (necessarily) specific for health economic models.

3.2.5.1.1 Searching for Treatment Effect

Cochrane and CRD provide detailed guidance on how to identify treatment effectiveness information.[4,5] Search recommendations include use of multiple bibliographic databases and using a RCT or systematic review filter, where appropriate. The guidance documents recommend supplementary search methods such as searching trial registries, hand searching, citations searching, grey literature searching and consultation with experts. [17,21]

3.2.5.1.2 Searching for Adverse Events

Cochrane and CRD also provide guidance on adverse event searching.[17,21] The guidance documents cite non-randomised observational studies and non-research based information as relevant additional sources of evidence. A publication by Golder, Peryer and Loke (2019) provides details on challenges associated with searching for AEs, and provides methodological advice to help identify AEs data.[163] The authors list the key challenges as:

- AEs are not pre-specified and therefore it is not easy to pre-specify the search terms;
- Huge range of AEs to consider;
- AE classification/reporting is inconsistent, leading to poor indexing;

- Terminology is inconsistent, leading to multiple synonyms being required;
- AEs for different types of interventions require different approaches;
- Study type limitation to (short term) RCTs may not be appropriate but searching for other types of studies can be problematic due to inconsistent use of terminology and poor indexing;
- No single comprehensive source exists, and unpublished data may be of importance.

3.2.5.1.3 Searching for Costs and Economic Evaluations

Searching for costs and economic evaluations is challenging for two principal reasons: both guidance and reliable/comprehensive economic databases are lacking. Thielen *et al*. (2016) provide methodological advice on how to search for systematic reviews of economic evaluations.[164] Some of the recommendations that the authors make are: recommendations to search bibliographic databases, using "*a wide range of search terms including thesauri and proximity operators as well as truncation options*". Validated filters are available from the ISSG website.[83] It is recommended that as few restrictions are used as possible in the search strategy.[164] The SuRE vortal for costs and economic evaluations gives further details on searching for economic studies, including which sources to search and how to design search strategies.[106]

3.2.5.1.4 Searching for Utilities

Evidence for utilities can be derived from RCTs but might not be available and/or suitable. Studies reporting utilities can be identified from bibliographic databases and/or specialist databases. There are some utility-specific subject headings within MeSH and EMTREE as well as the general subject headings such as 'Quality of life' that are likely to identify the relevant studies, albeit with poor precision.[161] The guidance suggests iterative searching to be combined with usual practice systematic search techniques to investigate a broader spectrum of evidence. [114]

Two pivotal publications fall outside the time frame of this review. Advice on how to retrieve evidence on utilities from literature has been produced by the University of Sheffield for the NICE Decision Support Unit (DSU) in 2010.[72] More recently by an ISPOR Task Force was published on identification, review and use of health state utilities in cost-effectiveness models.[73,162]

3.3 Review 3: Literature Review on Information Retrieval Methods for Health Economic Models Reported in HTA Guidelines

3.3.1 Objectives and Scope

The third literature review focuses on reviewing how modelling guidelines, such as those published by HTA agencies or professional organisations, reflect the evidence available for health economic model input identification. The aim of this search is to determine to what extent information retrieval techniques are recommended in the guidelines and what these recommendations are based on.

A manuscript describing this literature review has been submitted for publication to the Health Information and Libraries journal and is currently being revised following peer review. The title of the manuscript is: "*Identification of evidence for health economic model inputs: a scoping review of health technology assessment agency and health economic modelling guidelines*." The original literature review was conducted in May and June 2020, similarly to Reviews 1 & 2. For the purposes of the manuscript, Review 3 was updated in December 2023.

3.3.2 Introduction

Health economic models for health technology assessment (HTA) require multiple data inputs. Selecting certain data for the health economic model may have important implications for the cost-effectiveness results of a given health technology. Therefore, it is important to understand how different data came to be incorporated in the model. Lack of clarity may result in concerns about transparency, i.e. which criteria were used to select one specific source over another source.[81,165] While there are well-defined approaches for searching for evidence on relative treatment effects, the same level of clarity does not extend to other types of health economic model inputs.[4,166-168] There is a general lack of accepted or widely-used approaches for searching, that are specific for health economic modelling.

Zechmeister-Koss and Schnell-Inderst (2014) conducted a systematic literature review of HTA manuals and health economic guidelines in order to gain an understanding of the appropriate sources for populating health economic models in the context of HTA.[7] The authors reviewed requirements for evidence identification, and noted that several publications required model input identification to be based on a transparent search, but that this requirement was mostly limited to treatment efficacy or effectiveness. Further, the authors found that some guidelines specifically pointed out the lack of available search methods for inputs other than relative treatment effect estimate, and also that searching outside bibliographic databases is needed for some types of model inputs. There are several other reviews published that review health economic modelling guidelines.[169-172] Most reviews of health economic modelling guidelines focus on aspects of health economic modelling, e.g., analytical technique, perspective, comparator, types of costs/outcomes and what type of uncertainty analysis is needed, and have not specifically reviewed recommendations for health economic model input identification.[169-172]
169-172

Owing to the lack of guidance for searching for health economic models, researchers may have a tendency to rely on the traditional search approaches rather than an approach that might be better-suited to evidence requirements for models. Two recent developments in literature review methods have been observed: 1) more different types of reviews and 2) a move to greater systematicity.[97] More literature review types have emerged since literature reviews are no longer only used to consolidate existing evidence but also they serve as instruments in evidence-based decision-making.[118] Despite this, much of the information retrieval advice remains generic and search techniques for specific review types remain a challenge.[97]

Researchers working on global health economic models, which are developed with the purpose of being adapted for several different HTA agency submissions, need to consider guidance from multiple HTA agencies. One important aspect to consider is how data for the model are identified. This includes considerations on what type of searching should be done, and how exactly the searching should be conducted. The search methods may be described directly in the HTA/health economic modelling guidelines, or the documents may refer to an external search method publication. Health care technology manufacturers often submit reimbursement applications for a new technology in several countries in parallel, and the requirements, processes and comparators vary. Further, many reimbursement dossiers contain two distinctive elements: clinical efficacy and economic value, each with their own data search recommendations and needs. Identifying key inputs for generating health economic models often needs to be fitted into the overall evidence generation activities. This is done in the interest of avoiding running several similar (but not identical) searches. To manage these several, sometimes conflicting perspectives in the most efficient manner, it is of interest to have a clear picture of which search methods have been recommended in the HTA and modelling guidance documents. Therefore, this review is of specific interest to organizations submitting evidence to HTA and reimbursement agencies, where there is a set of requirements that have to be met.

The objective of this scoping review is to understand the extent to which search methods published by specialist organisations, such as Cochrane, are referred to and what other recommendations for model input identification might be present in the modelling guidelines. Additionally, the aim is to demonstrate the general absence of standard approaches for searching that are specific for health economic model inputs.

3.3.3 Methods

A search was performed to identify either country-specific or general (not country-specific) health economic guidelines. The review included guidelines from HTA agencies that set out requirements for the conduct of modelling for HTA, as well as other good practice guidelines for modelling closely associated with specific HTA agencies, such as the NICE Decision Support Unit Technical Support Documents.

Scoping review was chosen as a methodology, as scoping reviews serve as an ideal tool for assessing the extent and coverage of existing literature on a particular topic.[173] They can be particularly valuable when dealing with emerging evidence, as they help to identify gaps or specific questions that can be addressed more precisely by further research.[174] For this literature review it was recognised up-front that country-

specific HTA guidelines are not commonly indexed in bibliographic databases. Therefore, a web-search based approach was adopted. Initially two websites were searched, ISPOR and Guide to Economic Analysis and Research (GEAR).[175,176] The websites of HTA agencies were also searched. This was supplemented by a web search for relevant methodological documents, as well as a hand search in the guidelines reference lists.

The search was undertaken in December 2023. For documents retrieved via the ISPOR website, the HTA websites were still directly visited for verification purposes to check whether the document on the ISPOR website was up-to-date, and to check whether any further information or relevant documents were available directly from the website. If multiple versions were available, the most recent one was included. No quality or reporting guidelines were included in this review. No restrictions were applied to country or publication date. Only English language guidelines were included in this review. Moreover, recommendations on evidence identification or potential sources needed to be included for at least one of the key model inputs. Key model inputs are defined here, as well as elsewhere, as 1. relative treatment effect estimates (i.e., efficacy/effectiveness), 2. adverse events, 3. health state utilities, 4. costs, 5. resource use and 6. baseline risks of events.[32] Models have broad evidence requirements, but the focus is on the six key model inputs because they are common to most economic models and because, except for relative treatment effect estimates, there is a lack of definitive search method recommendations.[6,167] They are also featured as the key model inputs in existing publications.[10,26]

The eligible documents were entered into EndNote Reference Management software and the full text of each document was included in the database. A data extraction grid was developed in Microsoft Excel, and was focused on extracting information on three components. Firstly, general bibliographic details, such as year of publication, type of document, authors or agency, etc. were extracted. Secondly, search recommendations by model input type were also extracted. Finally, citations of external search methods documents were recorded. A pilot test of three documents was carried out to ensure the extraction sheet was adequate for the task.

A descriptive analysis across key model inputs was taken. This was focused on whether there were required/recommended information retrieval methods, and whether any external search methods guidance documents were referenced. The results are reported according to the model input type, and include a list of external specialist search method guidelines.

3.3.4 Results

Overall, 59 publications were identified for inclusion in the pool of publications to be screened (Figure 3). Of these, 49 sources were identified from the HTA websites, and a further ten were identified from web searching. One source was not available online.[177] Sixteen exclusions were made because they were not in English language. [62,178-190] Therefore, of the initial 59 references, 42 met the inclusion criteria for this review. Publications from 32 countries were included, as well as ten guidelines that were not specific to a country. Some countries, such as the Baltic countries (Estonia, Latvia, Lithuania) have published a common set of recommendations.[191] For some other countries more than one publication was identified; England,

Wales and New Zealand.[26,65,73,192-194] Table 9 lists all the publications included in this review. Table 10 displays the types of key model inputs associated with search recommendations in each publication. Most of the publications were associated with recommendations for identification of relative treatment effect estimates (81%), with 43% of them only including relative treatment effect estimate recommendations and nothing else. Search methods for cost/resource use and health state utilities were reported in 33% and 31% of publications, respectively. Only few a publications included details on identification of baseline risk of clinical events (12%) or AEs (10%).

Figure 3. PRISMA flow diagram

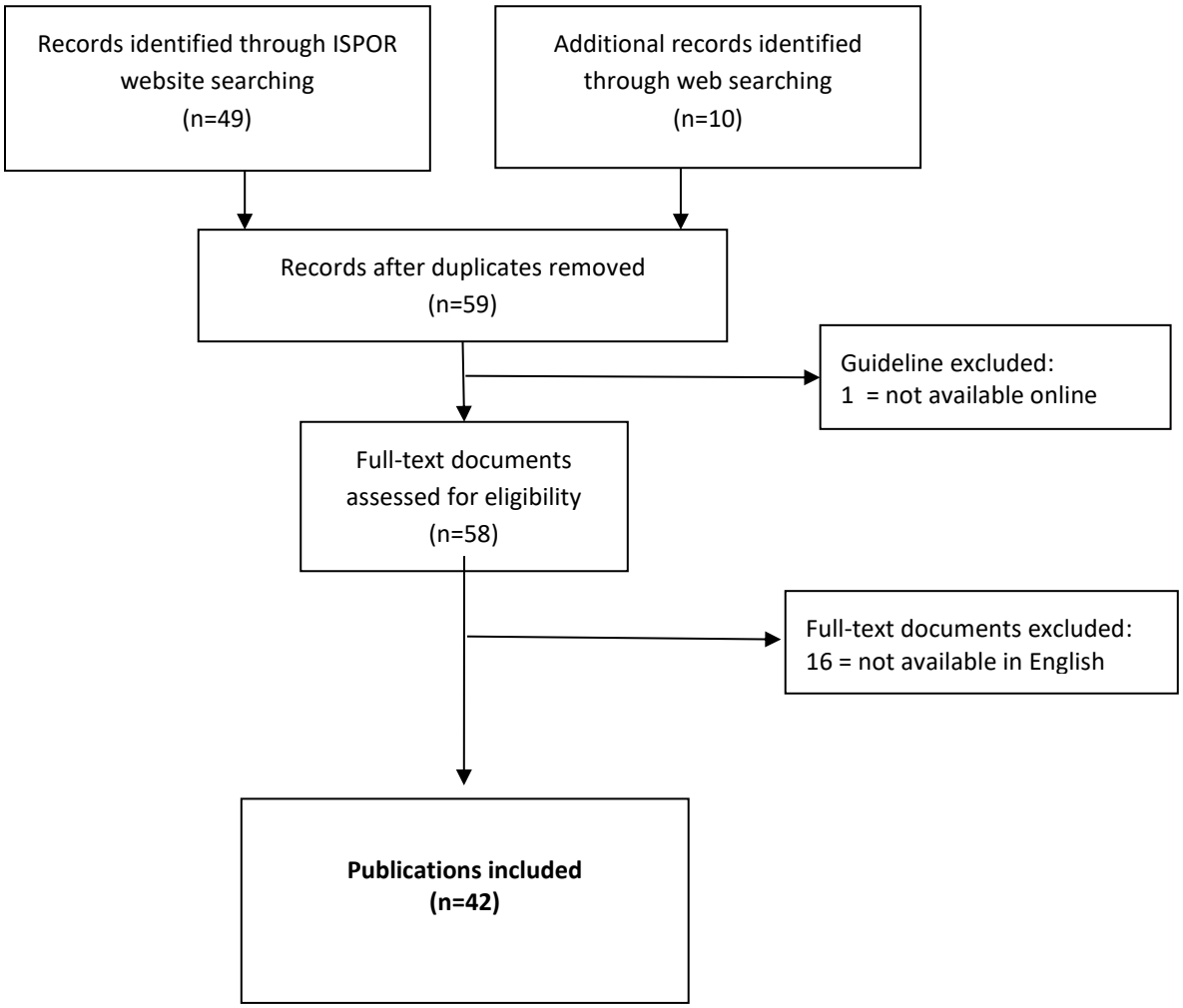


Table 9. Included publications alphabetically by country

Author, year/agency or country	Country	Reference
Pharmaceutical Benefits Advisory Committee	Australia	Pharmaceutical Benefits Advisory Committee, Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0). 2016, Australian Government, Department of Health.[64]
Austria	Austria	Walter E, Zehetmayr S. Guidelines on health economic evaluation, consensus paper. Vienna: Institute for Pharmacoeconomic Research; 2006.[195]
Baltic Health Authorities	Estonia Latvia Lithuania	Behmane D, Lambot K, Irs A, Steikunas N. Baltic guideline for economic evaluation of pharmaceuticals (Pharmacoeconomic Analysis). 2002.[191]
Health Care Knowledge Centre	Belgium	Cleemput I, *et al*. Belgian guidelines for economic evaluations and budget impact analyses. Health Technology Assessment (HTA). Brussels: Belgian Health KCE. 2012. KCE Report 183C.[196]
The Canadian Agency for Drugs and Technologies in Health	Canada	CADTH, Guidelines for the Economic Evaluation of Health Technologies: Canada. 2017.[50]
Agency for Quality and Accreditation in Health Care	Croatia	Agency for Quality and Accreditation in Health Care, Department for Development, Research and Health Technology Assessment. The Croatian Guideline for Health Technology Assessment Process and Reporting, 1st ed. Zagreb: 2011.[197]
Denmark	Denmark	Alban A, Gyldmark M, Pedersen AV, Søgaard J. The Danish approach to standards for economic evaluation methodologies. PharmacoEconomics. 1997;12(6):627–36.[198]
Egypt Ministry of Health	Egypt	Pharmacoeconomic Unit, Central Administration for Pharmaceutical Affairs. Guidelines for reporting pharmacoeconomic evaluations in Egypt. Version 01. Ministry of Health and Population. 2013.[199]
National Institute for Clinical Excellence	England & Wales	National Institute for Clinical Excellence (NICE). NICE Health Technology Evaluations: The Manual. Process and methods. 2022.[193]
NICE DSU TSD, Kaltenthaler 2011	England & Wales	Kaltenthaler E, Tappenden P, and Paisley S, NICE DSU Technical Support Document 13: identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. 2011: Sheffield.[26]
NICE TSU, Papaioannou 2010	England & Wales	Papaioannou D, Brazier J, and Paisley S, NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. 2010: Sheffield.[73]
NICE TSU, Dias 2011	England & Wales	Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU Technical Support Document 5: Evidence synthesis in the baseline natural history model. 2011.[192]
EUnetHTA	Europe	EUnetHTA. Methods for Health Economic Evaluations - A guideline based on current practices in Europe. European network for Health Technology Assessment. 2015.[200]
Lääkkeiden hintalautakunta (HILA)	Finland	HILA. Preparing a health economic evaluation to be attached to the application for reimbursement status and wholesale price for a medicinal product 2019.[201]
Haute Autorité de santé	France	Haute Autorité de santé (HAS). A methodological guide. Choices in Methods for Economic Evaluation. 2012.[27]

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen	Germany	Institute for Quality and Efficiency in Health Care ((IQWiG)). General methods for the assessment of the relation of benefits to costs. Version 5.0. 2017.[202]
Ministry of Health, Indonesia	Indonesia	Indonesian Health Technology Assessment Committee (InaHTAC), Ministry of Health Republic of Indonesia. Health Technology Assessment guideline. Jakarta. 2017.[203]
Health Information and Quality Authority	Ireland	Health Information and Quality Authority (HIQA): Guidelines for the Economic Evaluation of Health Technologies in Ireland.[204]
Agenzia Italiana del Farmaco	Italy	Agenzia Italiana del Farmaco (AIFA). Guidance to applicants for the submission of pharmacoeconomic analysis within the Pricing and Reimbursement Dossier. May 2020.[205]
Central social insurance medical council	Japan	Center for Outcomes Research and Economic Evaluation for Health, National Institute of Public Health (C2H) (Japan): Guideline for preparing cost-effectiveness evaluation to the Central Social Insurance Medical Council (Version 2.0). 2019.[206]
Ministry of Health Malaysia	Malaysia	Ministry of Health Malaysia—Pharmaceutical Services Division. Pharmacoeconomic Guideline for Malaysia. Second edition.[207]
Zorginstituut Nederland	Netherlands	Zorginstituut Nederland. Guideline for economic evaluations in healthcare. 2016.[208]
Pharmaceutical Management Agency of New Zealand	New Zealand	Pharmaceutical Management Agency of New Zealand (PHARMAC): Prescription for pharmacoeconomic analysis. Version 2.2. 2015.[65]
PHARMAC	New Zealand	PHARMAC, Guidelines for funding applications to PHARMAC. 2017, Pharmaceutical Management Agency: New Zealand.[194]
The Norwegian Medicines Agency	Norway	Norwegian Medicines Agency (NOMA). Guidelines for the submission of documentation for single technology assessment of pharmaceuticals. 2018.[209]
Agencja Oceny Technologii Medycznych i Taryfikacji	Poland	Polish Agency for Health Technology Assessment. Health Technology Assessment Guidelines. Version 3. (Warsaw, 2016).[210]
National Authority of Medicines and Health Products (INFARMED)	Portugal	Silva AE, Pinto CG, Sampaio C, Pereira JA, Drummond M, Trindade R. Guidelines for economic drug evaluation studies, 1998.[211]
Scottish Medicines Consortium	Scotland	Scottish Medicines Consortium (SMC) Guidance to manufacturers for completion of New Product Assessment Form (NPAF). 2022.[66]
Agency for Care Effectiveness	Singapore	Agency for Care Effectiveness (Singapore): Drug evaluation methods and process guide. 2018.[212]
South Africa	South Africa	Matsos, M., Guidelines for Pharmacoeconomic Submissions Republic of South Africa. 2012.[213]
Spain	Spain	López-Bastida, J., Oliva, J., Antoñanzas, F. *et al*. Spanish recommendations on economic evaluation of health technologies. Eur J Health Econ 11, 513–520. 2010.[214]
Taiwan Society for Pharmacoeconomic and Outcomes Research	Taiwan	Taiwan Society for Pharmacoeconomic and Outcomes Research (TaSPOR), Guidelines of Methodological Standards for Pharmacoeconomic Evaluations. 2006.[215]
The Medical Association of Thailand	Thailand	Health Intervention and Technology Assessment Program (Thailand): Ministry of Public Health guidelines for health technology assessment in Thailand (2nd ed). 2014. https://www.hitap.net/documents/168738.[67]

Academy of Managed Care Pharmacy (AMCP)	United States	AMCP, The AMCP format for formulary submissions. Version 4.1. 2016.[216]
Drummond *et al*. 2015	International	Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.[217]
EUnetHTA	International	EUnetHTA. Methods for Health Economic Evaluations - A guideline based on current practices in Europe European network for Health Technology Assessment, 2015.[200]
Gold *et al*. 1996	International	Gold MR. Cost-effectiveness in health and medicine: Oxford university press; 1996.[218]
The International Decision Support Initiative	International	Wilkinson T, Sculpher MJ, Claxton K, Revill P, Briggs A, Cairns JA, *et al*. The international decision support initiative reference case for economic evaluation: an aid to thought. Value Health. 2016;19(8):921–8.[219]
ISPOR	International	Weinstein MC, O'Brien B, Hornberger J, *et al*. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2003; 6(1): 9-17.[165]
ISPOR	International	Brazier J, Ara R, Azzabi I. Identification, review and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. . Value in Health 2019; 22(3): 267-75.[162]
Kobelt 2013	International	Kobelt G. Health economics: an introduction to economic evaluation. London: Office of Health Economics, 2013.[220]
Philips *et al*. 2006	International	Philips Z, Bojke L, Sculpher M, Claxton K, Golder S, Philips Z. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. PharmacoEconomics 2006; 24(4): 355-71.[9]
World Health Organisation	International	Finc J. Making choices in health: who guide to cost effectiveness analysis. Geneva: World Health Organization, 2003.[221]

Table 10. Types of key inputs that are associated with search recommendations

HTA agency/publication	Systematic literature review required for:				
	Treatment effect (n=35)	Adverse events (n=4)	Utilities and utility decrements (n=13)	Cost and resource use (n=14)	Baseline risk of clinical events (n=5)
Australia [64]	✓		✓		
Austria [195]	✓				
Baltic Health Authorities [191]				✓	
Belgium [196]	✓			✓	
Canada: CADTH [50]	✓		✓		✓
Croatia [197]	✓		✓	✓	
Denmark [198]	✓			✓	
Egypt [199]	✓			✓	

England & Wales: NICE [193]	✓		✓		
England & Wales: NICE TSU [26]	✓	✓	✓	✓	✓
England & Wales: NICE TSU [192]					✓
England & Wales: NICE TSU [73]			✓		
Finland [201]	✓				
France [27]	✓			✓	
Germany [202]	✓			✓	✓
Indonesia [203]	✓			✓	
Ireland [204]	✓	✓			
Italy [205]				✓	
Japan [206]	✓				
Malaysia [207]	✓				
Netherlands [208]	✓				
New Zealand: PHARMAC PE guide [65]			✓	✓	
New Zealand: PHARMAC Submission [194]	✓				
Norway [209]	✓		✓		
Poland [210]	✓	✓	✓		
Portugal [211]				✓	
Scotland [66]	✓				
Singapore [212]	✓				
South Africa [222]	✓				
Spain [214]	✓				
Taiwan [215]	✓				
Thailand [67]	✓		✓		
United States: AMCP [216]	✓				
Drummond *et al.* 2015 [217]	✓				
EUnetHTA [200]	✓				
Gold *et al*. 1996 [218]	✓				
iDSI [219]	✓				
ISPOR [165]	✓		✓		
ISPOR [162]			✓		
Kobelt 2013 [220]	✓			✓	
Philips *et al*. 2006 [9]	✓	✓	✓	✓	✓
WHO [221]	✓				

Relative Treatment Effect Estimate

The identified HTA and modelling guidelines varied greatly in their recommendations for searching for evidence to inform relative treatment effect estimates, from none to explicit recommendations on search strategy, specific databases, non-database searching, filters, types of studies/data to be searched, the external guidelines cited and whether identification should also be carried out for treatment extrapolation. Some guidelines did not make specific recommendations in relation to identifying evidence for relative treatment effect estimates, such as the Portuguese and Finnish guidelines.[201,211] Furthermore, a significant proportion simply state that comprehensive searching should be carried out, but without providing many details or references to external guidelines.[9,62,81,191,223-228]

In the Japanese guidelines, no specific databases are recommended but it is a requirement to report those which have been used.[206] Guidelines from Australia, Belgium, Canada, Croatia, England and Wales, France, Ireland, Japan, Malaysia, Netherlands, New Zealand, Norway, Poland, Scotland, South Africa, Taiwan, Thailand, and the United States include a clear recommendation that systematic literature review is required.[9,27,50,64-67,193,196,197,204,206-210,215,216,222] The databases that are recommended include: Medline, Embase, The Cochrane Controlled Trials Register (CCTR), The Cochrane Database of Systematic Reviews (CDR), The CRD DARE and Trip. Further databases suggested by the New Zealand PHARMAC health economic guidelines include sources shown in Table 11. Non-database searching in the following sources is also recommended: clinicaltirals.gov, International Clinical Trials Registry Platform, Australian Clinical Trials Registry, Internal registries, and grey literature searching. Grey literature has been defined as a vast range of different information created outside of standard publishing routes, and which is frequently not well represented in indexed bibliographic databases.[229] The PHARMAC guidelines further note that it may be useful to check the reviews of clinical evidence undertaken by national HTA organisations whose reviews and decisions can be publicly accessed, such as those shown in Table 12.[64]

Table 11. Recommended additional sources in PHARMAC guidelines

Name of the source	Website
Cochrane	http://www.cochrane.org/
UK Medicines Information	http://www.ukmi.nhs.uk/default.asp
Evidence-Based Medicine	http://ebm.bmj.com/
BMJ Clinical Evidence	http://www.clinicalevidence.com/ceweb/conditions/index.jsp/
Prescrire International	http://www.prescrire.org/

Table 12. Recommended national agencies to check (PHARMAC)

Agency	Website
National Institute for Health and Care Excellence (UK)	http://www.nice.org.uk/
NIHR Health Technology Assessment Programme (UK)	http://www.hta.ac.uk/
Canadian Agency for Drugs and Technologies in Health	http://www.cadth.ca/
Scottish Medicines Consortium	http://www.scottishmedicines.org.uk/
Australian Pharmaceutical Benefits Scheme	http://pbs.gov.au/
Belgian Health Care Knowledge Centre	http://kce.fgov.be/
Swedish Agency for HTA and Assessment of Social Services	http://www.sbu.se/en/
All Wales Medicines Strategy Group	http://www.wales.nhs.uk/
CEA Registry	https://research.tufts-nemc.org/cear4/

Some of the guidelines name specific bibliographic databases that should be searched. The submission guideline from Poland recommends that as a minimum, Medline, Embase and the Cochrane Library are searched.[210] The guidelines also state that further medical databases may need to be searched in certain cases, as recommended by the EUnetHTA guidelines.[166] Further search tactics recommended are reference searching, searching at least two clinical trial registries (clinicaltrials.gov and clinicaltrialsregister.eu), consultation with experts, non-systematic searches of specialist journals, contacting authors for unpublished data, internet search engines and consultation with manufacturers using regulatory dossiers. A requirement for a written search protocol is present in 15 of the included guidelines.[27,50,64,65,193,194,196,197,204,206,208-210,216,222] Two guidelines make specific reference to the use of search filters.[26,64]

Eleven of the identified documents included references to particular external search method guidelines. The most commonly referenced external guidance documents were the Cochrane Handbook for systematic reviews and University of York CRD's guidance for undertaking reviews in health care.[4,167] The list of external guidance documents and the HTA guidelines where they have been cited in is shown in Table 13.

Table 13. External searching guidance documents and their citing HTA guidelines

External Guidance	HTA/modelling guideline where mentioned
Agency for Health Care Research and Quality (AHRQ). Methods guide for effectiveness and comparative effectiveness reviews, 2014 [98]	AMCP US [216]
Centre for Reviews and Dissemination University of York. Centre for Reviews and Dissemination. Systematic reviews: CDR's guidance for undertaking reviews in health care. York, 2009 [4]	Belgium [230]
	NICE England [193]
	Kaltenthaler *et al*. 2011 [26]
	Ireland [204]
EUnetHTA. HTA Core Model® EUnetHTA Domain 4 - Clinical effectiveness (EUnetHTA Joint Action 2, Work Package 8. HTA Core Model, version 3.0 (PDF) 2016: Methodology/Where to find information? p. 146 [166]	Poland [210]
Higgins J, Green S. Highly sensitive search strategies for identifying reports of randomized controlled trials in MEDLINE. Cochrane Handbook for Systematic Reviews of Interventions 4.2. 5 [updated May 2005]; Appendix 5b. The Cochrane Library 2005;(3) [167]	PBAC Australia [64]
	Croatia [197]
	Kaltenthaler *et al*. 2011[26]
	Taiwan [215]
	Thailand [67]
Hoaglin DC, Hawkins N, Jansen JP, ***et al***. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2011; 14(4): 429-37 [102]	AMCP US [216]
Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness R. In: Eden J, Levit L, Berg A, Morton S, eds. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC): National Academies Press (US) Copyright 2011 by the National Academy of Sciences. All rights reserved.; 2011 [231]	AMCP US [216]
McDonagh MS, Jonas DE, Gartlehner G, ***et al***. Methods for the drug effectiveness review project. *BMC Med Res Methodol* 2012; 12: 140. [232]	AMCP US [216]

Adverse Events

Some guidelines, such as those for the Baltic countries and the Netherlands, do not mention AEs as a model input.[191,208] However, most guidelines identify AEs as an important model input, but do not specifically give details on which AEs should be included in the model or how information related to the AEs should be identified. Some of the most detailed recommendations relating to the safety/AE profile of the treatment to be assessed were laid out in the Polish guidelines and follow the EUnetHTA model.[166,210] The general recommendation is that the identification process should follow a similar literature review methodology as that used to identify evidence on relative treatment effect estimates. The NICE Methods Manual for England does not include specific information on identifying AE data.[193] Additional checks were made on the NICE website to identify any further documents that may contain information on identifying AEs. The NICE company evidence submission Document B template for submitting companies does not contain any specific instructions for identifying AE data.[233] However, the Document B template is associated with a User Guide that included wording around a requirement that details of the methodology used for identifying AEs should be included.[234] The CRD's guidance for undertaking review in health care is provided as an example how identification of AEs should take place and be reported.[4] Some of the

guidelines do not make specific recommendations for bibliographic database searching, but recommend reviewing online sources. The New Zealand PHARMAC guidelines recommend reviewing the following sources for safety information Medsafe: http://www.medsafe.govt.nz/, U.S. Food and Drug Administration (FDA): http://www.fda.com/ and European Medicines Agency (EMA): http://www.ema.europa.eu/.[65] These sources will give case study reports that can helpful to grow knowledge body e.g. by describing newly recognised/suspected AEs. On the other hand, case studies do not allow generalisability, causal interference or estimation of epidemiological quantities (e.g., proportion of patients who experience a certain AE).[235]

Health State Utilities and Utility Decrements

The most extensive HTA document to provide details on identification of utilities comes from England. The NICE Methods Manual states that the evidence to inform utilities must be systematically identified and selected, although there is little guidance on how exactly to do this.[193] The NICE DSU Technical Support Document Number 9 provides details of how to search for utility literature.[73] The document recommends that the scope of evidence identification is kept broad initially, and that the literature review will be refined according to the evidence uncovered. Therefore, an iterative process is recommended, and a variety of sources and methods should be used to identify relevant studies e.g., electronic database searching, reference list checking and expert opinion. This document states that the framework used for identifying relative treatment effect estimate literature is not useful for scoping health state utility reviews. This is mainly because searching for utilities is iterative in nature, not focused on specific study type (such as randomised controlled clinical trials [RCTs]) and because searching usually happens for health states and not for intervention and comparators. The NICE Methods manual is focused on how changes to health-related quality of life should be measured (method and instrument), as well as sources, in case (EQ-5D) data is not available from the clinical study. The manual simply states that if EQ-5D is not available from the study, values should be sourced from literature using a systematic search, without any further details or references to search method manual. The NICE DSU Technical Support Document includes more details on how to search, and it also considers the challenges associated with identification of health state utilities, which is in line with this PhD study.

A recent guideline from ISPOR also recommends searching utilities in an iterative manner.[162] Overall, the health economic guidelines referenced the Cochrane Handbook and NICE DSU document 9 for possible methods, and some other guidelines make a recommendation to simply follow good scientific practices for health state utility identification.[4,50,64,73,167]

Cost and Resource Use

The guidance documents included recommendations for cost and resource use input identification for the purpose of including them in the model. Additionally, several documents contained recommendations for conducting a systematic literature review of existing economic evaluations.[27,193,230] The latter is not specifically done to identify costs and resource use data for the model. Rather, it is used to identify if there are any existing economic studies that could be used instead of undertaking a new analysis.[236,237] Although not its primary purpose, the literature review can also be a helpful starting point to obtain an overview of available cost and resource use literature. Some of the guidelines make recommendations for

bibliographic database searching. The Belgian guidelines state that databases to be searched for this literature review should include Medline, Embase, and NHS Economic Evaluation Database (NHS EED). However it should be stated that NHS EED is increasingly out-of-date as no bibliographic records have been added since the 31st March 2015. [27,193,230] The Canadian guideline states that "*In the reference case, researchers should systematically identify, measure, value, and report all relevant resources based on the perspective of the publicly funded health care payer.*"[50] Many guidelines include separate sections for cost and resource use, but these sections or this distinction does not necessarily extend to identification of data for these two distinct types of input data. Some of the most detailed recommendations come from Belgium.[230] In the Belgian guidelines there is a section on costs and measurement of resource use. The resource use section in the guidelines mentions the types of data that are acceptable, which include clinical trials, prospective observational studies, databases and patient charts. Potential databases in Belgium are listed in Appendix 6 of the guidelines and the federal level databases include Cellule Technique pour la gestion des données RCM-RFM, Bases de données Agence Intermutualiste (AIM), Databanken Intermutualistisch Agentschap (IMA–AIM), INAMI–RIZIV, BCFI–CBIP, SPF Santé publique FOD Volksgezondheid, and ISP–WIV.[230] Further community level databases are listed as well (see Appendix 6 in Cleemput et al. 2012).[230] It also discusses the process of validating resource use data from other countries to the Belgian context, including sourcing expert opinion. Many Belgian data sources are mentioned, such as All Patient Refined Diagnosis Related Group (APR-DRG) for mean length of stay in hospital in Belgium (for resource use) and the Belgisch Centrum voor Farmacotherapeutische Informatie (BCFI) website for unit prices for drug cost.[238,239] Other guidelines, including those from Croatia, Ireland, and Malaysia also state that cost and resource use information in the model needs to be systematically identified and detail some country-specific cost (and rarely resource use) sources, but no further details on recommended information retrieval methods are given.[197,204,207] Some further guidelines list specific sources, such as Statistics Netherlands (Centraal Bureau voor de Statistiek, CBS), National Institute for Rijksinstituut voor Volksgezondheid en Milieu (Public Health and the Environment) and L'Institut national de la statistique et des études économiques (INSEE) in France.[208,209,223,224,240] Other documents focus on the types of costs that should be included in the model over specific sources or methods of identification.[191,208,210]

Baseline Risk of Clinical Events

The included guidelines focus on the types of sources of information rather than information retrieval methods. Dias *et al*. 2011 provide the most relevant guidelines that focus on evidence synthesis for the baseline natural history model (the NICE DSU Technical Support Document 5). The publication states that a common approach for identifying sources for this input has been to use the same trials that have provided information on relative treatment effect estimates but restricting to the standard of care arm(s) of the trials. The authors conclude that this may create potential issues with drawing general conclusions from the model for the target population, and further sources such as registry data can also be considered. Previous health economic models are also mentioned as one possible source, but it is worth noting that using existing models might restrict the new model to the limitations of existing models.[192] Most of the documents focused on the synthesis of data. Local data and expert opinion are also mentioned as potential sources. In the New Zealand PHARMAC guidelines, emphasis is placed on ensuring that any

modelling of natural history of disease is sufficiently adjusted for local age-ethnic-specific population data.[65] Specific sources, such as a government website http://www.stats.govt.nz/ are recommended. The submission guidelines from New Zealand provide more details on specific sources for New Zealand, such as detailed burden of disease estimates for New Zealand (including e.g., Māori versus non Māori), but they are not explicitly stated to apply to economic analyses.[194]

3.3.5 Discussion

Review 3 was undertaken with the objective of reviewing HTA and generic health economic modelling guidelines in order to summarize and compare the recommendations regarding input identification. An additional objective was to identify which externally developed; specialist information retrieval techniques were referenced in the health economic model guidelines. This scoping review provides the first comprehensive account of search methods for health economic model parameters in HTA and generic health economic modelling guidelines.

Overall, 42 HTA or generic health economic modelling guidelines were identified that included recommendations for identifying model inputs. The level of detail in which information retrieval for model inputs was addressed varied greatly between the guidelines, and also by model input type. The issue of information retrieval was not handled systematically across guidelines, and the sections discussing information retrieval were not always easily found within the documents. Where information retrieval methods were specified, there was little discussion on alternative information retrieval techniques for searching bibliographic databases or other search methods.

The most common model input type that was associated with search method recommendations concerned relative treatment effect estimates. Search method recommendations were largely based on publications that provide methods for comprehensive searching (e.g., Cochrane or CRD guidance).[4,167] Methods recommended by HTA guidelines for searching for AEs also varied from no explicit recommendations to recommendations to use comprehensive search methods that are focussed on sensitivity over precision.[210] For the identification of health state utility values, iterative searching is emerging as a search technique but empirical research is still lacking.[73,168,241] For costs, resource use and baseline risks of clinical events, no external information retrieval method documents were referred to in the guidelines. Instead, where guidance existed, it was restricted to the naming of local sources such as websites or databases. These sources can be complex to navigate requiring local language skills, or a paid subscription is required to permit access. Searching for baseline risks of clinical events was not associated with clear search methods in the guidelines, possibly because baseline risk of clinical events represents a heterogeneous model input, with variability coming from the differences in disease areas and model structure.

There is a paucity of literature that summarises available search methods for health economic modelling. There are very few published reviews of health economic modelling methods that also include evidence identification as one of the components summarised, and nothing after 2014.[242] However, it is interesting to note that the high level findings from what is published are aligned with the findings of this much more up-to-date and comprehensive scoping review, i.e., that there is a lack of health economic modelling

specific search methods, especially outside identification of relative treatment effect estimate.[242] Further, this scoping review has examined the search recommendations by model input type and reports on these findings in more detail.

The results from this study should be interpreted in light of the following limitations. Firstly, only English language guidelines were included. Secondly, this review is not intended to be a systematic literature review associated with high sensitivity. The intention was not to capture every piece of literature that exists on this topic, but to ensure that key documents were included. It is unlikely that any other currently available evidence would not alter the conclusions that can be made from the study. Thirdly, no quality appraisals were carried out as no generalised tool was available to assess the quality of the recommendations in this field. Finally, the search was limited to the identification of the evidence used to inform key model inputs, and did not consider other aspects of model development such as understanding of the decision problem, or conceptualisation of the model structure.

The comprehensive search methods, such as those described by AHRQ, Campbell Collaboration, Cochrane, and CRD do not address all search needs for health economic models that additionally require different types and format of evidence from peer-reviewed scientific evidence (e.g., RCTs) to routinely collected / administrative data.[4,98,99,130] Further, health economic modelling study questions often cannot be pre-defined and therefore health economic modelling evidence requirements cannot be fully defined until the underlying conceptual model has been determined. This study provides useful insights for researchers working in health economic modelling or in information retrieval, especially in the context of HTA submission development.

More formal and reproducible approaches to identification and assessment of quality of model inputs are required to reduce the 'black box' nature of decision models, and lead to less scepticism regarding model outputs.[6] Systematically identifying evidence for inclusion in a model does not necessarily mean following a Cochrane style review for every health economic model input, as this is neither feasible nor required. Different types of search methods exist, including search methods to improve efficiency (rapid reviews and text mining), complexity of research questions (realist reviews) and sensitivity/sufficiency (qualitative reviews and rapid reviews). Realist reviews are known for their iterative nature, which is a key aspect of their methodology.[243] This approach involves continuously refining the review process based on emerging insights and data. The iterative process typically includes defining the review scope, developing initial program theories, conducting evidence searches, selecting and appraising data, and synthesizing findings.[243] As new information is gathered, the initial theories and search strategies are revisited and adjusted to better understand the contexts, mechanisms, and outcomes of the interventions being studied.

There is a lack of empirical evidence to show which search method is not only the most efficient, but also a transparent method for identifying model inputs. The lack of clear guidance in this area may result in more burden in terms of time and resources for searching and reviewing literature for model inputs. Future empirical research investigating the suitability of information retrieval techniques such as rapid searching, iterative searching or search automation would have major benefits in the context of model input identification.

3.3.6 Conclusions

This study shows that only limited types of search methods for health economic model inputs have been recommended in the HTA and general health economic modelling guidelines and they are mostly associated with the identification of evidence on relative treatment effect estimates. There is a lack of widely-used standard approaches for searching for health economic model inputs in HTA. Further research into suitable search methods can help to increase the efficiency and transparency of model input identification for health economic models, and therefore increase the transparency of the conclusions drawn from those health economic models for decision making.

3.4 Summary of the Information Retrieval Literature (Reviews 1-3)

This PhD study aims to select and develop alternative search approaches for identification of evidence for health economic models. Existing literature was reviewed in order to form a picture of information retrieval methods in evidence based medicine (Review 1). It was of interest to review information retrieval literature more widely in evidence based medicine, rather than directly focussing on information retrieval for health economic models, as it was expected that literature relating to health economics models would be scarce and/or limited in methodological scope. Further, it was assumed that information retrieval in evidence based medicine may include search methods that show potential for health economic model input identification, even if they have not yet been applied in this context. A sub-search to identify information retrieval methods specific for health economic models was also undertaken (Review 2). Finally, published HTA and health economic modelling guidelines were reviewed to better understand which information retrieval methods are cited in the published guidelines (Review 3). This both helped to cross-check if any new search methods would be identified from these guidelines, and to understand the current practice in searching for health economic model inputs, although it is recognised that real world practice may differ from HTA/modelling guideline recommendations.

The review found several established methods published for searching in systematic reviews (e.g. CRD, Cochrane Handbook).[4,5] Some guidance also exists for searching in other contexts, such as for a specific review type (e.g. rapid review) or for a specific type of data (e.g. clinical efficacy, safety, health state utilities). However, there is little established practical guidance on how to conduct searching for reviews other than Cochrane style systematic reviews. Some guidance exist on how to search for health economic models, for example NICE TSD 13.[26] The importance of the iterative nature of searching is recognised in this document. More recent recommendations, such as ISPOR Task Force Report for identifying health state utilities, also recommend adopting an iterative search approach, although the recommendations are not based on empirical evidence.[162] Intuitively, an iterative process for health economic model input identification makes sense, given that health economic model development is also iterative, and the modelling and/or searching approach evolve as knowledge builds up. Given that iterative searching is both mentioned in the literature and matches the practicalities of health economic modelling, iterative searching will be one of the two search methods to test in this PhD study.

In Review 2 several of the HTA/health economic model guidelines required searching for model inputs to be "*systematic and transparent*", often without specifying exactly how the searching should be conducted or reported. The most common exception to this was clinical efficacy, where Cochrane style systematic searching was often recommended. In a health economic model clinical efficacy informs the relative treatment effect modelling, often through an indirect treatment comparison or network meta-analysis. Rapid reviews have the potential to be systematic and transparent, if appropriately reported. Rapid reviews use several methods to simplify or omit some of the processes used in systematic reviews, including reducing databases, allocating one reviewer for each review stage, omitting or minimizing the use of grey literature, and doing so narrowing the scope of the review.[244] They share methodological similarities with full systematic reviews but are faster to conduct. Rapid reviews therefore have the potential to address a key challenge associated with health economic input parameter identification (i.e.

using full systematic reviews is too resource intensive). Rapid review was therefore chosen as the second method to be tested in this PhD study.

A further objective of the PhD study is to adapt or develop a framework for the reporting of search methods for health economic models. Existing reporting standards were identified and summarised as part of Review 1. One or more of these will be considered in the development of a reporting tool for the searches in this PhD study.

The key gap observed in the literature was that none of the recommendations for health economic model search methods were based on empirical research. Further, there are no established, practical step-by-step guidelines on how to conduct searching using iterative searching or rapid review methods, especially in the context of health economic modelling. This PhD study seeks to make a contribution to filling both gaps. This is to be done by providing an empirical comparison of health economic model input identification methods, including testing the search outputs in an executable health economic model to understand the marginal impact of additional information. This PhD study also seeks to contribute to the development and description of iterative searching and rapid review methods, as well as reporting standards. The next Chapter defines usual practice searches as well as giving details of the two chosen alternative search methods (iterative searching and rapid review). The step-by-step practical methods were partially based on the available (but limited) literature as well as specifically developed for this PhD study.

4 Methods

4.1 Chapter Overview

This chapter provides details of how the methods were chosen for the primary research, given the research objectives. The methodological approach and implementation are described (Section 4.2). This also includes description and explanation of model inputs included in the project (Section 4.2.1) and search methods (Section 4.2.2). The search methods are described in Section 4.2.2.1 (usual search practice) and in Section 4.2.2.2 (alternative search methods). This PhD study adopts case studies as the methodological approach (Section 4.2.2.3). The performance of usual practice and alternative search methods were evaluated by using outcomes described in Section 4.2.2.3. Approaches to the reporting of the searches are described in Section 4.2.3.

4.2 Choice of Methodological Approach

The key question that this research is aiming to answer is "*Does the use of alternative search approaches for the identification of evidence for the key model inputs lead to a more efficient approach to searching, when compared against current searching practice?*" A key motivator for the PhD study was the lack of empirical research on different information retrieval methods, especially in the context of health economic model input identification. Doing so can lead to either resource intensive, Cochrane style search methods being used for some of the inputs, while the majority of other inputs are not associated with any record of how they come to be incorporated in the model, leading to lack of transparency. Conducting a survey to establish usual practice would have been one potential method to establish usual practice, but it was determined to be out of scope for this PhD study. The empirical focus of this PhD study was the testing of the search results in the health economic models. Usual practice was determined through the case studies instead.

This PhD study focuses on two model inputs: health state utilities and baseline risk of clinical events. These types of inputs were judged 1) not to be associated with established and tested search methods in the context of economic models 2) to often be impactful model inputs, and 3) to be representative of the challenges that are also associated with other model inputs. The details on selecting these input types is given in Section 4.2.1.

For this study, a desire existed to gain a better understanding of available alternative search methods. An early literature review on available search methods was conducted during this PhD study, resulting in an improved comprehension of the applicable search methods. Among the various information retrieval methods identified, iterative searching and rapid review were selected for testing in this research project. These two methods were deemed most suitable for the type of searching often associated with health economics models, which involves multiple and evolving information needs. Usual practice, to which these alternative search methods are compared, is explained in Section 4.2.2.1. The rationale behind choosing the two alternative methods is explained in Section 4.2.2.2.

Case studies can be useful to explain, describe and explore events or phenomena in the everyday contexts in which they occur.[245] A case study approach was chosen for this PhD study because it allows practical testing of the alternative search methods. No empirical research on the impact of alternative search methods in economic models was identified in literature. It is therefore unlikely that other research methods, such as interviews or focus groups, would give insights based on experience into using these alternative search methods. Green and Thorogood defined a case study as "*in-depth study undertaken of one particular 'case', which could be a site, individual or policy*".[246] In this PhD study a 'case' is a health economic model that is associated with health state utility and baseline risk of clinical event inputs, and one of the usual search practices. The case study approach is explained in detail in Section 4.2.2.3

Finally, to assess the performance of different search methods, metrics to measure their differences were needed. These are used to measure multiple aspects, such as the time it took to carry out the search, and the number and relevance of findings. These outcomes, by which the performance of search methods is measured, are described in Section 4.2.2.3.

4.2.1 Model Inputs Included in the Project

A selection of which key model input parameters to test the alternative search methods on needed to be made. This section explains which inputs have been selected and why. Table 14 below summarises which model inputs have been included in this PhD study.

Table 14. Summary of included and excluded parameters

Model parameter	**Included or excluded**
Relative treatment effects	Excluded
Adverse events	Excluded
Costs	Excluded
Resource use	Excluded
Health state utility values and utility decrements	Included
Baseline risk of clinical events	Included

Identification of relative treatment effects are excluded from this research project. There may be room to improve the efficiency of existing search approaches, but given the many explicit and implicit recommendations to use Cochrane style searching in the HTA guidelines, relative treatment effects are excluded from this project. There is a general consensus in the HTA guidelines (Review 3) that a comprehensive approach is needed for this parameter. Further, in HTA treatment effect is not needed solely for health economic models, but also for clinical assessment of the health technology.

AEs are also excluded from the project, on similar grounds to treatment effect (i.e. HTA guidelines recommend an established search approach). Searching for AEs in the context of health economic modelling for HTA is challenging for many reasons. Some of the HTA guidelines include very detailed

search requirements for the assessment of safety of the new product *outside* of the modelling context. It is unclear if these same search requirements/recommendations apply for the identification of AEs for health economic modelling, and if so, whether they are specific for the new intervention only or also for the comparator(s).

Health state utility values are included in the study, because they are often an impactful model input parameter and are challenging to search for. Further, an ISPOR Task Force has recently published guidance on iterative searching on utilities. This can be tested empirically and compared with other alternative search methods.

The issues associated with searching for cost, resource use and baseline risks of clinical events are to an extent similar. This is not only because all of these inputs are associated with searching observational data and other data collected for routine purposes, but also because the data are likely to come from variety of sources, such as bibliographic databases, specialist databases and from non-database searching. The ultimate aim is to develop a framework that allows for efficient searching of these multiple sources and that the searches can be reported transparently. Some guidance for searching for economic evaluations and cost studies does already exist,[27,193,208,209,223,224,230,240] but very few guidance documents are available for identification of baseline risk of clinical events. Therefore, baseline risk of clinical events is included. Costs and resource use are excluded.

4.2.2 Search Methods

4.2.2.1 Usual Practice (Control Arm)

The usual practice search approach served as a control arm against which the alternative search methods could be compared. There are three possible usual practice search methods that were included in this project, as described below.

Full Systematic Literature Search

Full systematic ('Cochrane style') literature review refers to systematic searching that has a clearly focused search question and that aims to search extensively as described by well-known information retrieval guidelines, such as those from Cochrane or CRD.[4,5] HTA agencies have published guidance on how to conduct systematic searching, focussing on searching for treatment effect data in Cochrane style searches.[247] For other types of evidence, including utilities and baseline risk of clinical events, input identification usually needs to be systematic but no specific methods are described. Due to this lack of guidance, the tendency has been to rely on the traditional literature review approaches even in situations when an alternative approach may be a better fit to modelling requirements.[10,31] Therefore, Cochrane style, full systematic literature review is included as one of the possible usual practices.

Minimum Search Recommendations

Paisley (2016) proposes minimum requirements for identifying evidence for key model parameters, including utilities and baseline risk of clinical events.[10] For utilities it is recommended that one bibliographic database (Medline) is searched, together with specialist databases, if accessible. The

specialist databases may include Tufts CEA Registry, School of Health and Related Research Health Utility Database (ScHARRHUD) (no longer updated) or instrument websites such as EQ-5D. For baseline risk of clinical events, one bibliographic database (Medline), specialist database (if accessible) and non-research and non-standard-format sources are recommended, as appropriate. Specialist databases could include jurisdiction specific incidence and prevalence databases. Non-database searching may include national and international websites, such as Office of National Statistics (ONS), US CDC or World Health Organisation (WHO). For the purposes of this research project, if a search was identified that was similar to Paisley's minimum search requirements, it was also considered 'usual practice' and therefore included.

No Specified Search Method/Rationale

Not all health economic models include details of how model inputs were identified. The searches for model parameters other than those searched for in full systematic searches tend to be done on an *ad hoc* basis, without a pre-specified strategy and/or proper recording of the steps taken to explain how particular inputs came to be incorporated in the model. Baseline risk of clinical events is a particularly challenging model input parameter type to search for because it involves combining standard and non-standard format sources, and is characterised by a lack of consistent terminology with which to describe the evidence. For this reason, not having a clear search method will also be considered usual practice for this study.

4.2.2.2 Alternative Search Methods (Experimental arms)

There is little methodological research comparing different search methods, and it is not known to what extent there is a trade-off between sensitivity and efficiency for identifying model input parameters. This lack of knowledge of the potential trade-offs makes it challenging for researchers to have the confidence to rely on alternative approaches, especially in the context of reimbursement submissions to HTA agencies. The alternative search methods that will be compared to the 'usual practice' control arm are iterative searching and rapid review, as described in the sections below.

Iterative Searching

Iterative searching has been recognised to be increasing in importance for qualitative and purpose-specific literature reviews.[97] However, currently there is no common definition or methodology with detailed practical steps describing how to carry out iterative searching.[97] Iterative searching can be deemed to be a potentially useful method for model input identification because of the nature of model development i.e., it is rarely possible to pre-define the search questions fully upfront. A recent publication by an ISPOR Task Force on searching for health state utilities also recommends an iterative search and review approach for utility identification.[162] This project provides an opportunity to test an iterative approach empirically and to compare it with rapid review methods as well as the usual search practice. For the purposes of this study, iterative searching is not simply the refinement of a search strategy. Rather, it is the running of additional searches (iterations) to identify different information from one or more previous searches. In this more specific sense, this iterative searching, as defined, is akin to/in the manner of the technique or approach as it is described and employed in realist synthesis.[121]

The iterative searching will be done in the following stages: (1) develop information requirement(s), (2) build the search and (3) perform the search. These tasks count as one iteration of the search, and this sequence can be performed as many times as needed. Each stage is informed by what is learned in the previous stage and/or the current stage. The searching is repeated until saturation is reached i.e., no new relevant citations/topics are being identified. The iterative search process is shown on Figure 4.

Developing Information Requirement

The first iteration can be either a focused or wide search, with the sensitivity or research question changing with each subsequent iteration in response to findings. In this way, the iterative process allows the search approach to be explorative and adaptable. Iterative processes can also include searches of reviews and overview articles.

Publications by key authors can also be scanned to identify key papers and to find relevant key words/phrases. Backward and forward citation tracking can be used to increase the body of knowledge.[248] In backward citation tracking references cited in the publication are examined to increase knowledge of the topic and to identify experts. [249] Further publications by authors are examined to review previous work for relevance. In forward citation tracking, new publications that reference the original article are identified.[249,250] This allows for expansion of the knowledge of the topic through follow-up publications. This process closely matches the berry-picking information seeking model.[20] These different information retrieval techniques can be used in any combination in iterative searching, as well as in other search methods. This also includes Cochrane style searching, but unlike with iterative searching, the information requirement in Cochrane style searching cannot be modified. In contrast in iterative searching, the information requirement can be modified or fine-tuned at any point, as necessary, based on the increased understanding of the topic of interest. This understanding can improve both through searching or through modelling activities. This step ensures that the focus is on what is perceived to be the most important aspects of the topic. Based on the revised information requirement, the most relevant articles will be selected for inclusion in the health economic model.

The process of screening publications and modifying the search strategy ("search and select") in between search iterations will be documented. The key feature of iterative searching is that in iterative searching, 'patches' (i.e. some or all citations from one iteration of a search) of information are evaluated and further searching is then initiated if judged relevant. Therefore, unlike in Cochrane style searching, the search is not necessarily only run once using one pre-defined search protocol.

Build Search Strategy

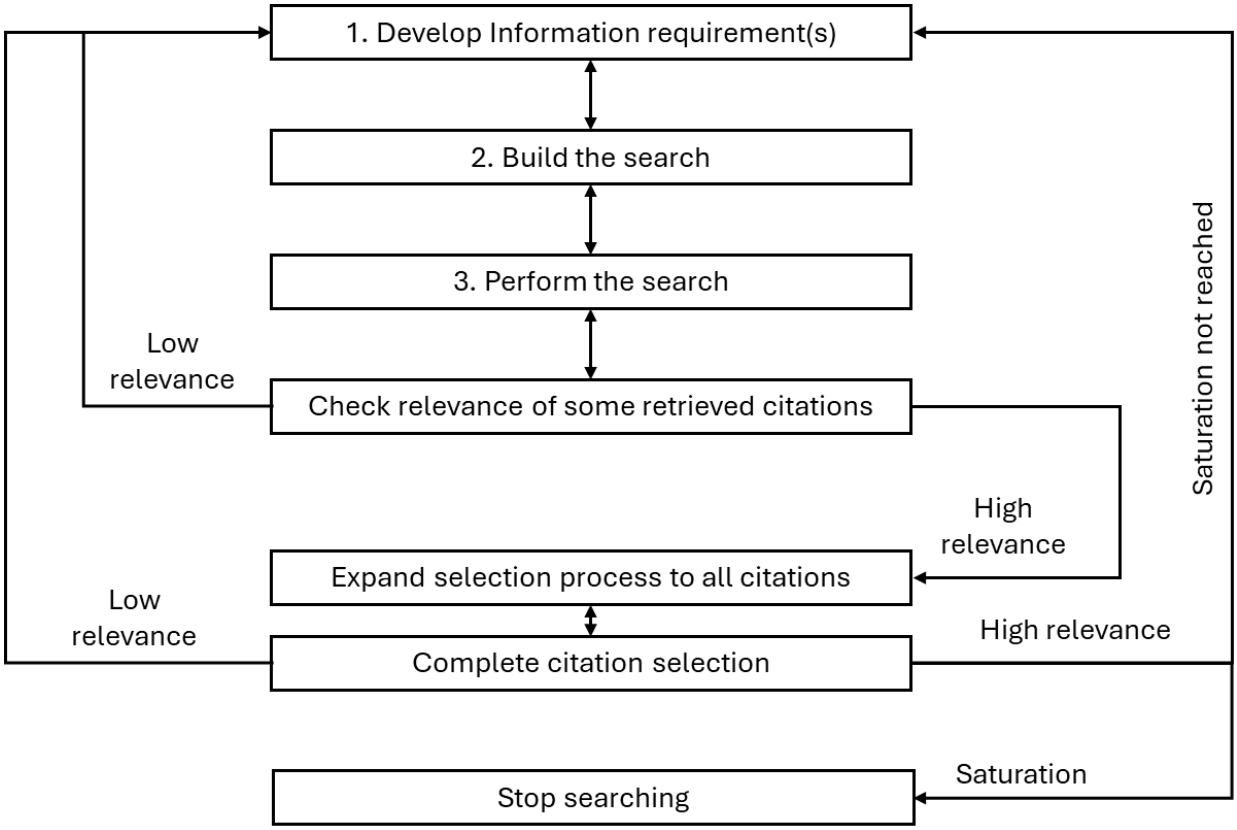
An initial search strategy is constructed, and if it includes a bibliographic search, it is then executed on the relevant databases. The researcher can start to review the titles/abstracts, and if needed re-do the search without finishing reviewing all sources. This would count as the first iteration. It is not necessary to have only one search query, but it can be series of search queries in case the search concepts are scattered.

Performing the Search

When the search is performed, some or all of the citations may be screened. The patch of information may be highly relevant and it makes sense to continue to screen the rest of the citations. Similarly, it may

be apparent that little of no relevant data was found. If only some citations are screened (i.e. not the whole patch), in this PhD study the rest of the citations will be screened as a validation to ensure that no further relevant citations were found in the discarded part of the patch. The final search results may contain findings from any or all of the search iterations, or 'patches'.

Figure 4. Iterative search method concept



Decision to Stop Searching

The search process for iterative searching should be stopped when no further relevant information is identified. A key concept is therefore making judgements about the relevance of already retrieved information and the potential gains from further searching. Two concepts can be helpful here: relevance and saturation. This Section summarises and discusses how saturation of relevant evidence might be defined in the context of searching for health economic models.

Saturation

In qualitative research, the concept of saturation is applied as a benchmark to determine when to cease data collection, and can be helpful in determining when to stop searching for model inputs during an

iterative search. Saturation can be defined in different ways. Saunders *et al*. 2018 presents four models of data saturation (Table 15).[251]

Table 15. Models of saturation (source: Saunders *et al*. 2018[251])

Model	Description
Theoretical saturation	Relates to the development of theoretical categories; related to grounded theory methodology
Inductive thematic saturation	Relates to the emergence of new codes or themes
A priori thematic saturation	Relates to the degree to which identified codes are exemplified in the data
Data saturation	Relates to the degree to which new data repeat what was expressed in previous data

The first three of the above four saturation models are based on grounded theory.[252] Theoretical saturation is based on the development or refinement of (existing) theoretical categories of data.[253,254] The researchers sample until no further data is found for the different categories. The second model, inductive thematic saturation, defines saturation as a point in analysis when no new themes occur in the data and there are mounting instances of the same themes.[255] Unlike theoretical saturation, theoretical inductive saturation relates to the emergence of new themes rather than the degree of refinement of those themes already identified. The third model, *a priori* thematic saturation, defines saturation in reverse compared to the previous models: *Given the theory, do we have enough data to illustrate it?*[256] The final model is related to data saturation instead of theoretical saturation. Data saturation aims at understanding how much data is needed until nothing new is apparent or once information redundancy is reached.[257]

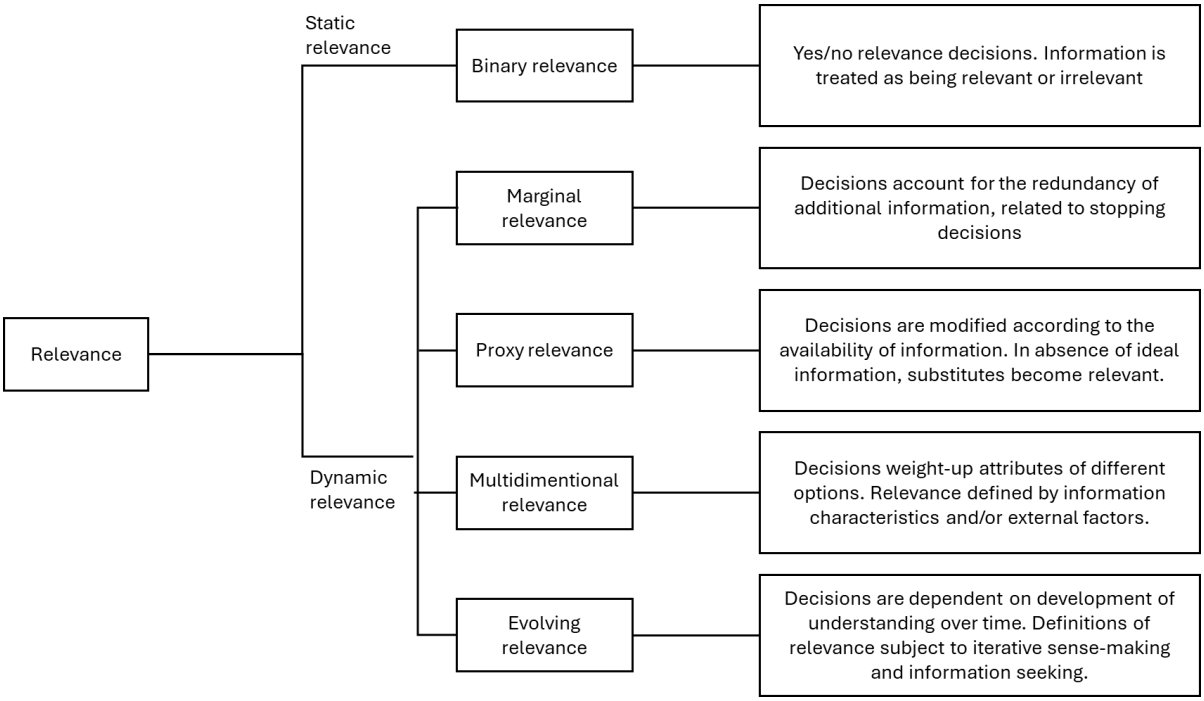
Relevance

Relevance has been identified as a key concept for explaining information seeking behaviour of the model development process by Paisley.[1] Model development is a complex task that requires an understanding of the issues relevant for the decision problem and the translation of this understanding into a mathematical model. Therefore, Paisley does not interpret relevance only as a binary, predetermined concept, but also as a dynamic multidimensional concept. Paisley describes how modellers and reviewers in her case studies used different strategies to maximise the allocation of available resources in understanding potentially relevant information. This is to prevent spending excess time in processing information that is unlikely to be relevant. Parallels to information seeking behaviour models were described by Paisley (berry-picking model and information foraging theory).[20,258] A key concept of the berry-picking model is evolving query.[12] For information foraging theory, important concepts include completion of complex task, where relevance is defined as anything that helps to complete that specific task.[13] Also, information foraging theory recognises that processing information is associated with resource and opportunity cost. Both of these information theories recognise that naturalistic behaviours like browsing, 'bit-at-a-time' and enrichment strategies are being used to maximise the relevant information. Paisley concludes that the classic information retrieval model i.e. Cochrane style comprehensive searching, does not have a conceptual framework that can accommodate the complex and ambiguous nature of information retrieval for many of the health economic model information needs.

Cochrane style systematic reviews require a binary *a priori* understanding of the information need which means the information need can't be changed and evidence can only be judged as relevant or not relevant, and may therefore result in a conclusion that no evidence exists to permit estimation of (cost-) effectiveness. Decision making usually cannot be delayed, and therefore the aim of searching should be to identify the best possible evidence for use in the health economic model at the time at which a decision must be made. That being said, there is not a complete absence of clearly defined information needs at the outset of the modelling process, and some of these such as clinical efficacy may require comprehensive information retrieval. But for many model parameters a predefined definition of relevance would restrict the extent to which a search can identify the full scope of potential information that can be used in the model, therefore potentially resulting in bias. Relevance should be defined in such way that it can accommodate the complex and emerging nature of relevance to prevent loss of relevant information for populating the health economic model, and also so that it provides a way of defining saturation (marginal relevance in Figure 5).

Paisley refers to information science literature to provide definitions of relevance. These are summarised in Figure 5 (source: Figure 8.1 in Paisley 2012[1]). She writes that most health economic modelling information queries can be characterized as dynamic relevance definitions, rather than static although static information needs also exist. For example, a cost is identified from a drug tariff list and this tariff list is accepted (e.g. by the decision-maker) as an authoritative source. Therefore, other sources with the drug cost become not relevant. There is a stopping rule because of the authority of the drug list, and therefore it is not necessary to identify all occurrences of the same drug cost.

Figure 5. Specifications of relevance (source: Paisley 2012[1])



Practical Implications of Saturation and Relevance for Stopping Searching

The concept of relevance establishes that making judgements about the relevance of information is not always binary for health economic models, and search methods need to support the dynamic nature of information retrieval for models. Saturation is also part of the concept of relevant. If further information adds nothing to the task, then it is not relevant (marginal relevance). Therefore, relevance is not intrinsic to the information itself but it is embodied in the relationship between the information/task for which is used, and the user of information. When searching for health economic model inputs, a judgement needs to be made about stopping searching i.e., when *saturation* of *relevant* information is achieved.

The different concepts of saturation and relevance will be used in this research project to characterise the rationale for stopping searching. Constant comparison is a technique that will be used in the search process, and this will also provide support to arrive at conclusions about relevance and saturation, and therefore when to stop searching. Both saturation and relevance as dynamic concepts are open to interpretation, and therefore the decision to stop searching will be different for different searches. In this study, an important part of saturation and relevance assessment was done by running the health economic model with the identified inputs, and making observations about the additional information that they bring to the assessment of the decision problem that the particular model aims to address. It is important to consider the context in which saturation is being used.

In this thesis, the aim was to transparently report how decisions about relevance and saturation have been made for each of the iterative searches. A careful evaluation was made how relevance and saturation are conceptualized and operationalized for each search. This was done through the concepts identified by Paisley 2012 (Figure 5) and Saunders 2018 (Table 15).[10,259] This should lead to a consistent approach for

stopping searching; not in the sense that the terms are always used in the same way, but rather in relation to consistency between the theoretical and analytical framework adopted. This will allow stopping searching in a way that best meets the aims and objectives of each search.

Rapid Review Methods

A rapid review aims to accelerate systematic review processes by limiting various stages to provide more timely information for decision making.[260] The methods of conducting rapid reviews vary widely, but are characterised by shorter timelines. The Search, Appraisal, Synthesis and Analysis (SALSA) framework for rapid reviews outlines some basic characteristics: limits on the search through time constraints; the quality appraisal (limited or not performed); narrative and tabular synthesis only; and analysis focussing on overall quality/direction of effect literature.[261] Rapid review is not necessarily less systematic and can still follow the principles of systematic reviews.

Rapid review methods may be useful in the modelling context, because one aim is to understand whether an alternative search method can reduce the burden to review excessive numbers of citations. A previous comparison has shown that the main conclusions that can be drawn from a systematic literature review and rapid review do not differ significantly.[262]

It is important that rapid review is conducted in systematic fashion, to avoid it simply being a narrative review or poorly conducted systematic review[263] and to avoid bias.[117] A thorough evaluation of suitable rapid review methods was done for each case study. For example, fewer sources may be searched, or other limits imposed (years searched, languages included and sources included). The selection of the exact rapid review methods was made for each case study, utilising those described by Ganann *et al*. 2010.[117] In iterative searching, screening of identified materials takes place between the iterations. In contrast, in rapid reviews, only the final results are screened thoroughly.

Similar to iterative searching, transparency of the methods applied is of critical importance. Poor reporting of rapid review methods and limitations may lead to the impression that rapid review is secondary to full systematic review, whereas this may not be the case if the rapid review is transparently reported. The search reporting framework development within this project aims to minimise these concerns (Section 4.2.3).

4.2.2.3 Selection of Case Studies

To test the alternative approaches, case studies were needed to compare them with usual practice and to assess their relative efficiency. It is important to have a sufficient number of case studies to draw conclusions from, and therefore three case studies were included. Yin (2003) states that a case study can either contain a single case study or several.[264] Having more than one case study allows an understanding of similarities and differences between the case studies to develop, and according to Baxter & Jack the evidence generated from multiple case studies can be strong and reliable.[265] Any observations made in this PhD study from multiple case studies would therefore be more intensely grounded in a range of empirical evidence, making it more convincing theory than if only one case study was used.[266] On the other hand, multiple case studies can be time-consuming to process. Gerring (2004) argued that the more case studies are included, the less observation time has been spent on each individual case.[267] It was

judged that three case studies would be an appropriate number of case studies for this PhD study, given the time constraint of the project and still allowing for high quality observations of each case.

The emphasis was on the identification of case studies using the following criteria:

- A Modifiable Microsoft Excel model must be available that includes health state utility input data or baseline risk of clinical event input data, or both.
- One of the 'usual practice' methods (as described in Section 4.2.2.1) must be available.
- For both model parameters, at least three case studies needed to be selected.
- Priority is given to any case study that includes both health state utilities and baseline risk of clinical events.

An open-access model or a model that is otherwise available (which does not contain confidential information) is needed to test the impact of alternative search methods, without having to re-build the whole model. Building a new model would take time away from testing the alternative search strategies, and therefore reduce the resource available to address the main aim of this research project. A feasible source was the NICE Multiple Technology Appraisals that SCHARR team members have access to, and this was deemed to provide the best approach to gain access to both modifiable Excel models and associated records of the associated searching, as these are routinely reported both as NICE Assessment Reports and as HTA monographs.

The previous searching should have been done using usual, current practice, although at times this may be difficult to determine for health state utilities and baseline risk of clinical events. The possible definitions of usual search practice were described in Section 4.2.2.1.

This study is at risk of bias as a single researcher (myself) conducted an updated version of the usual practice search as well as the two alternative search approaches. Once a reviewer's knowledge of a topic grows, this can be assumed to make the searching more efficient. Therefore, if the usual practice search is always constructed/updated first, it is possible that it is the slowest of the search methods due to that being the first search method applied. However, in this case it may not have a major impact as the usual practice search will be updated from an existing search i.e., not started from scratch. Still, to minimise this potential bias, and as no additional reviewers will be available, the searching started with a different search technique for each case study. The approaches for each case study were conducted in the following order:

- Case study 1: Usual practice, iterative searching, rapid review
- Case study 2: Iterative searching, rapid review, usual practice
- Case study 3: Rapid review, usual practice, iterative searching

4.2.2.4 Outcomes by Which Search Performance was Measured

A comparative analysis of usual search practice and alternative search methods is needed to compare the search methods, and for this a measure of efficiency needs to be determined. For bibliographic database searches, efficiency is usually measured in terms of sensitivity and precision.[5] For this study further measures incorporating the time and burden were needed. Limited publications were available that assess the performance of searching with outcomes other than those associated with Cochrane style comprehensive searching (sensitivity, specificity and precision). Payner *et al*. is a prospective comparison of search strategies developed with and without text-mining tools.[19] This study measured sensitivity, number needed to read, number of hours spent searching and screening, platforms searched and total number of deduplicated citations. For this PhD study, the most relevant outcome measures were adopted. For other types of searching (non-research, non-standard), the comparison was done on lists of input values (if available), time and/or sources obtained.

The efficiency measures for bibliographic databases, adapted from Payner *et al*. 2021[19], are:

- **Time** to measure the efficiency of how long it takes to develop and conduct a search strategy (hours)
    - Measured in 5-minute increments on a tracking sheet to indicate the number of hours spent. Total number of hours spent was calculated for each search option.
- **Sensitivity** to assess the relevant papers identified by the experimental methods
    - Sensitivity is defined as the number of relevant citations identified out of the total number of relevant citations in existence. 'Relevant citations in existence' is defined as the citations identified by the usual practice search. In this PhD study sensitivity is not explicitly calculated but rather qualitatively assessed. This is because it is possible that one of the experimental search methods finds more relevant citations than usual practice, which would result in sensitivity over 100%.
- **Burden** to determine the screening burden
    - **Number needed to read** (NNR) is defined as the number of irrelevant citations a reviewer has to screen for each relevant citation found. This performance measure can be interpreted as the number of articles that researchers has to screen before finding one relevant paper, reflecting the efficiency of the search.[268]
    - **Precision** is calculated by dividing the total number of articles found in bibliographic databases with the number of true positives.
- **Relevance** to determine the relevance of the findings for modelling
    - **Evolving relevance**: Were all known relevant citations needed for modelling?
    - **Marginal relevance**:
        - Not all known relevant citations identified: If all known relevant citations were not identified, did it make any difference for modelling?
        - Relevant, new citations (that were not identified by usual practice search) identified: Does that make any difference for modelling? *See definition for **sensitivity** above*.

The usual practice bibliographic database searches in the identified case studies were updated in order to have comparable, up-to-date results against which to test the alternative approaches. That way, all searches (usual practice and alternatives) have been run on the same, or very similar, dates. The usual practice bibliographic database searches were devised by the authors of the HTA assessment, and therefore it is assumed that in most cases information specialist and/or systematic reviewers were included in the search design. If bibliographic database results were available by search line, these were included in this report. All alternative searches were devised by me (with an oversight from the supervisors) and results by line are reported, as relevant.

An estimation of the time to develop and run the usual practice searches was done retrospectively. The usual practice search consisted of the published search, plus an update search to ensure comparability of the findings with the alternative search methods. The retrospective estimation includes the amount of time taken for the development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening against the inclusion criteria. In the first case study (Chapter 5) the time spent updating the usual practice search was recorded in 5-minute increments. The time for conducting the original search reported in the published report still needed to be estimated, as this was not available from the publication. The estimation was based on using the times from the update search per record screened. A worked-up example is given in case study 1, in Section 5.3.1. For running and updating the original usual practice search, the search protocol from the original search was used to ensure consistency between the original and update searches. The selection of publications from the previous search was also used, and only the records from the update search were reviewed.

The search burden was reported on a table format, that captured the number of citations by review stage and time by task (search protocol development, running of the searches, duplicate removal, title/abstract screening, full paper review). Table 16 shows usual practice search burden reporting as an example. As explained above, usual practice included two elements: the original published search and my update search. The table reflects this. For iterative searches the table format was edited to include as many iterations as was needed for each case study. Rapid review searches only required one column i.e. the table structure was simplified.

Table 16. Search tracker example

Item	Usual Practice Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Original	Update	Total	Original	Update	Total		
Search protocol								
Running searches								
Identified citations								
Duplicate removal								
T**itle/abstract**								
Citations excluded								
Full papers								
Sources in the report								
Sources in model								
TOTAL								

Abbreviations: NNR: number needed to read

The comparison of the length of time is measured proportionally (percentage change) rather than in absolute terms (number of hours). This is because search strategies are usually set up by information specialists, whereas here my inexperience resulted in comparatively longer absolute search times. Measuring the proportional times instead of absolute times, allowed comparing the time taken to conduct searching using the different search methods with each other but no conclusions about the number of hours taken to develop the search strategy can be made.

For the non-bibliographic database searching, it is not possible to estimate the time, sensitivity or burden difference between usual practice and the alternative approaches. For this reason, comparison was done by listing the input values (if any) and sources identified. It was not always be possible to access the identified sources, such as natural history databases. Searching for baseline risk of clinical events was particularly challenging. It can involve searching bibliographic and specialist databases, as well as other non-research and non-standard format sources. The intended development of a search framework (see Section 4.2.3 below) is aimed at capturing the searching performed in transparent manner. One of the key difficulties with the baseline risk of clinical event searching was that the sources were scattered and therefore the traditional search approaches were challenging to implement. Some data are routinely collected for administrative purposes by health insurance companies (e.g., sickness funds in Germany) or electronic medical records (e.g., the French administrative health care database (SNDS)).[269] This type of administrative data can be associated with limitations, such as incompleteness of the data. Many of the data sources do not cover different types of care settings e.g., only secondary care data are included and often disease severity measures are not available, thereby precluding an analysis of patient outcomes being performed. In addition to administrative data sources, disease-specific data sources also exist. They can be databases, registries, or studies using observational methods to evaluate a specific population of

people with a disease. Some of the potential sources may require payment for accessing the data (no funding was available to access these sources in this project), and therefore it is difficult to know the content of the database and therefore difficult to judge the relevance. Examples of disease-specific data sources for Multiple Sclerosis (MS) include European Database for Multiple Sclerosis (EDMUS), the Swedish National MS registry, the Danish National MS Registry and the global MS Registry (MSBase).[270]

When new input values were obtained from bibliographic or other type of searching, they were tested in the Excel models. This was more often possible for utilities than for baseline risk of clinical events for the reasons discussed above. The comparison of original and new values was done for the base case and also for sensitivity analyses to examine whether the amount of uncertainty was impacted by input values identified during the searches.

4.2.3 Search Reporting

Part of this study involves the review and development of reporting methods of the alternative searches, so that transparency of searching is maintained. In Review 1 existing reporting standards were identified and summarised (Section 3.2.4.3). Most of the search reporting standards have been developed for Cochrane style systematic reviews centred around identification of quantified clinical efficacy of a given health technology. Booth (2006) has developed STARLITE as proposed standards for reporting searches, that acknowledges the demands of both quantitative and qualitative literature searches.[159] Searching for health economic model inputs, like for qualitative reviews, includes a broader range of purposes, making Booth's STARLITE reporting standards a good starting point for search reporting framework for health economic models.

Many of the elements such as sampling strategy, type of studies, approaches (e.g., other than searching bibliographic databases), range of years, limits, inclusions and exclusions, terms used and electronic sources included in STARLITE are similar to the elements that need to be reported for the searches for health economic model inputs. Based on STARLITE, a draft reporting tool was produced to capture the important elements of searching for model inputs. A key addition to Booth's STARLITE is the separation of non-database searching into a separate section. In Booth's STARLITE this is covered under "Approaches", where approaches other than electronic subject searches or methods (e.g. hand-searching or citation snowballing) can be added. For the purposes of this project, it was decided that a separate detailed section to capture non-database searching would be beneficial in order to capture details related for e.g. HTA website searching. The structure of the form prompts the researcher to enter details on the source names and date they were accessed. Booth's STARLITE separates out functional and conceptual limitations, whereas for the purposes of this project they were combined into one field. An additional section for this project was added: iterative search specifications. This allows the research to capture the details on iterative search, which was one of the alternative search methods tested in this PhD study. The rapid review (the other alternative) specifications were already considered to be captured by Booth's STARLITE.

This search framework can serve as a vehicle by which the search approach is operationalised and managed. This should help with the planning and recording of the searches. It also provides an easy step to reporting the searches. Central to the search framework are model input-specific considerations, and generic considerations, such as how much searching is enough, what should inform decisions to stop searching, and what to do in absence of optimal (e.g., country/jurisdiction specific) evidence. Table 17 shows the initial version of the draft search framework. This draft was tested in the case studies.

Table 17. Draft search framework (some elements adapted from STARLITE[159] reporting framework)

Bibliographic database search elements	Considerations
Sampling	Should the sampling be comprehensive, selective or purposive? Why was specific type of sampling chosen?
Type of studies	What type of studies will be included?
Sources	Which databases and platforms will be sampled?
Limits	What limits can be applied and if so how can they be justified?
Terms used	What search strategy will be used for the main databases?
Conceptual limitations	Can further conceptual limitations be applied such as by geographical location, setting, specific focus of study etc.? How can these additional limits be justified?
Iterative search specifications	How will the initial search terms be informed? If no suitable evidence is identified, is further database searching is likely to be fruitful? If so, which parts of the inclusion/exclusion criteria should be relaxed? Has progress in the model development process changed the information needs? When to stop searching?
Non-database search elements	**Considerations**
Approaches	What approaches other than bibliographic database searching will be implemented?
Source names	What will be the exact sources to be accessed and what is the link (e.g., web address), if applicable?
Search dates	When will the source be searched?

4.2.4 Summary

The next three chapters (Chapters 5 - 7) report three case studies where iterative searching and rapid review methods (Section 4.2.2.2) were tested against usual practice (Section 4.2.2.1). The case studies were selected together with the PhD supervisors, and mostly based on the availability of an executable

excel model. This was a challenge, as most health economic models contain confidential information. PT was able to acquire three executable health economic models (ulcerative colitis, thyroid cancer and breast cancer tumour profiling risk stratification tool), and since they also met the rest of the case study inclusion criteria (Section 4.2.2.3), those three case studies were chosen for this PhD study. The search methods were carried out for the two chosen health economic model inputs: health state utilities and baseline risk of clinical events (Section 4.2.1). The sensitivity and burden performance outcomes described in this Chapter (Section 4.2.2.4) were used to enable comparison between the search methods. The reporting framework reported in this Chapter was also tested by using the case studies.

4.3 Ethical Approval

This study required no ethnical approval.

5 Case Study 1: Ulcerative Colitis

5.1 Chapter Overview

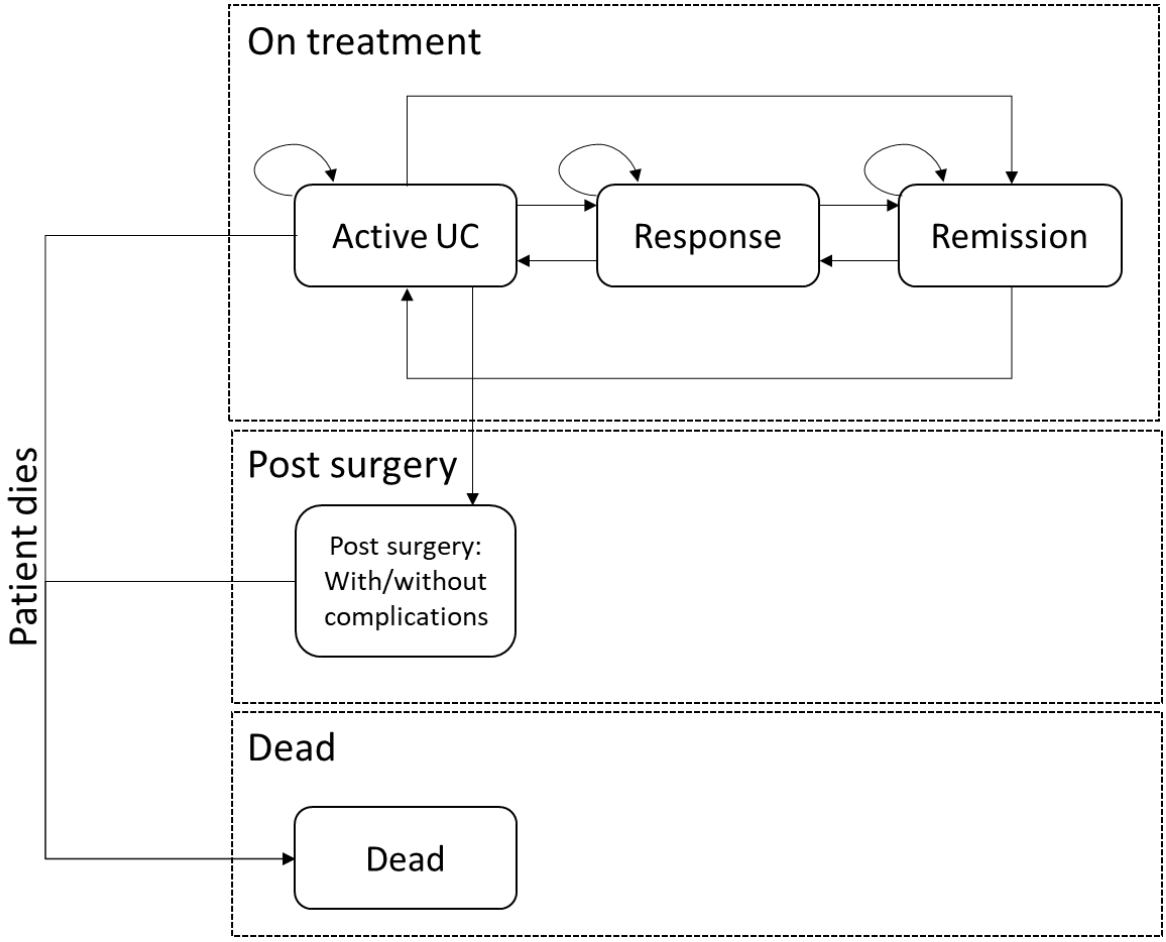
This chapter reports the findings of the ulcerative colitis case study. The Chapter provides an introduction to the case study, and then reports the searching for utility (Section 5.3) and baseline risk of clinical events (Section 5.4) separately. For both model input types, usual practice search, iterative search and rapid review methods are explained. This includes details of how the search was carried out, what the results were, how studies were selected and used in the model. This is followed by an assessment of impact of the sources on the model results, and then an assessment of using the performance measures outlines in the methods chapter (Section 4.2.2.4).

The utility part of this case study has been published in a peer reviewed journal: Lister J, Paisley S, Carroll C, Tappenden P. Empirical Testing of Alternative Search Methods to Retrieve Utility Values for Health Economic Modelling. Pharmacoeconomics. 2024 Aug 6. doi: 10.1007/s40273-024-01414-7.[271]

5.2 Case Study Introduction

In 2014 NICE carried out a Multiple Technology Appraisal (MTA) of infliximab (IFX), adalimumab (ADA) and golimumab (GOL) for the treatment of moderately-to-severely active ulcerative colitis (UC) after the failure of conventional therapy.[272] Ulcerative colitis (UC) is a disease that results in inflammation and ulcers of the colon and rectum, and can have a considerable impact on patients' health-related quality of life (HRQoL).[273] As part of this appraisal, an independent Assessment Group developed a health economic model to assess the cost-effectiveness of second-line IFX, ADA and GOL, conventional non-biological therapies and immediate colectomy. The modelled population was patients with moderate-to-severe UC who have failed at least one prior conventional therapy. The health economic model used a Markov structure (see Figure 6 for simplified model structure). The key model health states were defined according to whether the patient is alive or dead, their current level of disease control (remission, response and active UC) and their prior history of colectomy. A modifiable version of the original Excel model was made available by one of my supervisors (PT) to test the identified utility and baseline risk of clinical event data identified from the case study searches. Further details of the Assessment Group model can be found in Archer *et al*. 2016.[272]

Figure 6. Simplified UC health economic model diagram (source: Archer *et al*. 2016[272])



5.3 Utility Values

5.3.1 Description of the Usual Practice (Control)

Utility Value Identification

In the original search reported in the MTA, literature searches were undertaken to identify utility values measured using the EQ-5D instrument (including both 3 and 5 response levels) in literature relating to UC. In the health economic model, health utility is assumed to be dependent on the level of disease control achieved with drug therapy (active UC, response, remission), whether or not surgery has been performed and whether or not the patient experiences post-surgical complications. The usual practice search strategy) and combined free-text and MeSH or thesaurus terms relating to UC with terms for specific utility measures or more general utility terms. The search was classified as full systematic search (see Section 4.2.2.1 for definitions of usual practice). The MEDLINE search strategy is presented in Table 18. The search strategy was translated across all databases searched (Embase, Cochrane Library, CINAHL, BIOSIS). No date or language restrictions were applied. Literature searches were originally conducted for the MTA during January and February 2014. As part of this PhD study, I updated the searches in January 2023. Full details of the usual search methods can be found in Archer *et al*. 2016.[272]

Table 18. Medline (via Ovid) search strategy for utility values - Usual practice

Search Strategy
1. Colitis, Ulcerative/
2. ulcerative colitis.tw.
3. colitis ulcerosa.tw.
4. uc.tw.
5. colitis ulcerative.tw.
6. Colitis/
7. colitis.tw.
8. colitides.tw.
9. Inflammatory Bowel Diseases/
10. inflammatory bowel disease$.tw.
11. ibd.tw.
12. (col* and ulcer*).tw.
13. colitis gravis.tw.
14. proctocolitis.tw.
15. or/1-14
16. (euroqol or euro qol or eq5d or eq 5d).tw.
17. 15 and 16

Sources were included as part of the pool of studies from which the model base case and sensitivity analysis utility values would be chosen, if they reported EQ-5D utility estimates for multiple UC health states or utility values for post-surgery health states. Figure 7 displays the process of selecting the studies

for the original and the update searches. References were collected in a bibliographic management database (EndNote), and duplicates were removed. In total, there were 562 citations identified, of which 89 were duplicates. At abstract/title level, 473 titles were screened, and 119 full papers were retrieved for more detailed assessment. During this process, 91 studies were excluded as they were not relevant. Therefore, 28 citations were selected as potentially includable in the model, because they reported EQ-5D estimates for one or more health states relevant to the model.[274-299] In this first selection step, the candidate options are selected. In the second step, all the candidate options are weighted up according to the different attributes in order to make the final decisions about what is included in the model. This two-step process is important, as even if some sources are not included in the model (only mentioned in the report as potential candidate options), they are still relevant because they inform the final selections and improve transparency and rigour of modelling process. Inclusion/exclusion is a multi-dimensional process where relevance is assessed by weighting up the different attributes of candidate options.

Figure 7. Study selection results UC utility search – Usual practice (original + update)

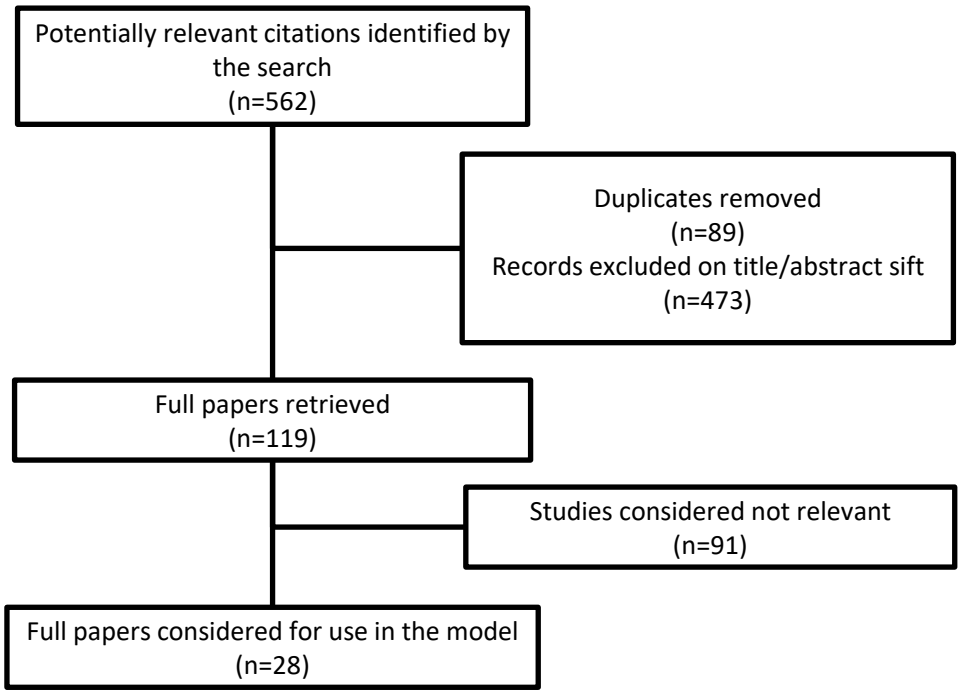


Table 19 shows the 28 studies identified from the usual practice search. Woehl *et al*. and Swinburn appeared to be the most useful among the identified publications, as they were UK-based studies with fairly large numbers of participants and have the widest coverage of health states for the model.[295,298] Woehl *et al*. was used in the original base case analysis as the post-surgery estimate for surgery was more consistent with published literature than that of Swinburn *et al*.[295,298] The publication by Swinburn *et al*.

was used in a sensitivity analysis. Two further sensitivity analyses in the original MTA included replacing utility for response/remission with valuations from ACT1 trial and PURSUIT maintenance trial.[274,275]

During the update search, a further 16 studies were identified out of which two were selected for inclusion in the model as sensitivity analyses.[293,296] The base case did not change. Firstly, Sardesai *et al*. was selected as a sensitivity analysis.[293] The publication reports a cost-effectiveness analysis of tofacitinib versus infliximab, adalimumab, golimumab, vedolizumab and ustekinumab for the treatment of moderate to severe UC in Germany. The utility inputs were derived from *post hoc* analyses of EQ-5D data in the OCTAVE 1, OCTAVE 2 and OCTAVE Sustain studies.[293] Secondly, the TRUENORTH trial was selected.[296] The NICE Committee papers report cost-effectiveness analysis of ozanimod versus current clinical management. The estimates were derived from EQ-5D-5L data collected during the TRUENORTH trial. The other 13 publications were not selected for inclusion in the model because the coverage provided for the health states was poorer than in the included publications. Hernandez *et al*. reported more health state utility values, but at closer examination these were derived from other published sources, primarily Woehl *et al*.,[288] and therefore the publication was excluded from the model.

Table 19. Overview of utility sources – Usual practice (original + update)

Study	Report (n=28)	Included in model (n=6)
ACT1 & ACT2 2014[33,34274,275]	✓	✓
Alrubaiy 2015[276]	✓	
Armuzzi 2020[277]	✓	
Assche 2016[279]	✓	
Beilman 2016[278]	✓	
Biedermann 2022[280]	✓	
Burisch 2022[281]	✓	
CADTH 2019[282]	✓	
Casellas 2003[283]	✓	
Dulai 2021[284]	✓	
Gherardi 2018[285]	✓	
Gibson 2014[286]	✓	
Hagelund 2020[287]	✓	
Hernandez 2020[288]	✓	
Kawalec 2018[289]	✓	
Kuruvilla 2012[290]	✓	
McLeod 1991[300]	✓	
PURSUIT 2014[274,275]	✓	✓
Richards 2001[301]	✓	

Sardesai 2021[293]	✓	✓
Scott 2020[294]	✓	
[c]Swinburn 2012[295]	✓	✓
TRUENORTH 2022[296]	✓	✓
Tsai 2008[302]	✓	
Vaizey 2013[303]	✓	
Van d. Valk 2015[297]	✓	
Woehl 2008[298]	✓	✓
Yue Min Ho 2019[304]	✓	

Footnote: a The utility weights derived by the TTO method are reported, b Licensed arms only, c Approximate estimate based on graph reported in the publication.

Utility Values Used in the Model

This section provides the utility values that were chosen to be used in the model. The health state utility values from Woehl *et al*. were used as the base case.[298] Of the five scenario analyses three were already included in the original NICE MTA.[23,29,293] Two further scenario analyses were added from the update search.[52,55] The utility values run through the health economic model are summarised in Table 20. Not all sources reported post-surgery valuations.[23,29,293,296] The post-surgery valuations in Woehl *et al*. were more consistent with the other published post-surgery valuations, and were used in all scenarios.[305-307]

Table 20. Health state utility values used in the model - Usual practice

	No response	Response	Remission	Post-surgery
Base case : Woehl *et al*.[298]	0.41	0.76	0.87	0.70
Scenario 1: Swinburn *et al*.[295]	0.55	0.80	0.91	0.70
Scenario 2: ACT1[23]	NR	0.82	0.88	0.70
Scenario 3: PURSUIT[29]	NR	0.80	0.89	0.70
Scenario 4: Sardasei *et al*.[293]	0.78	0.87	0.93	0.70
Scenario 5: TRUENORTH[296]	0.68	0.84	0.90	0.70

Abbreviations: NR: not reported. **Footnote**: [a]Post-surgery values from Woehl *et al*. were used for all scenarios.

Health Economic Model Results

Each intervention is compared to the next most effective alternative. This is done by calculating the incremental cost-effectiveness ratio (ICER). Dominating treatment options are both more effective and less costly than the alternatives they are compared to. When a treatment strategy is being dominated, it is both less effective and more costly. Extended dominance rules out any intervention that has an incremental cost-effectiveness ratio that is greater than that of a more effective intervention, meaning that the treatment does not lie on the cost-effectiveness frontier.[308]

The original base-case analysis of the model indicates that colectomy is expected to dominate all other treatments because it generates more quality-adjusted life years (QALYs) at a lower cost. All medical (drug) options were expected to produce considerably fewer QALYs at a greater cost than colectomy (Table 21). The model is sensitive to the utility values of remission, response, active UC and post-surgery. In the sensitivity analysis where utility values from Swinburn *et al*.[295] are used (scenario analysis 1), the results are reversed such that colectomy becomes the least effective option, instead of the most effective. This is also true for using sources from the update searches (scenario analyses 4 and 5); colectomy is the least effective option when using utility values from both Sardesai *et al*. and TRUENORTH studies.

Table 21. Health economic model results – Usual practice

	Incremental cost per QALY gained				
Analysis	**IFX**	**ADA**	**GOL**	**Conventional mgmt.**	**Colectomy**
Base case Woehl *et al*.[298]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 1 Swinburn *et al*.[295]	£178,982	£79,714	Dominated	Ext dom	-
Scenario 2 ACT1[23]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 3 PURSUIT[29]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 4 Sardasei *et al*.[293]	£396,008	£171,229	Dominated	Ext dom	-
Scenario 5 TRUENORTH[296]	£282,898	£128,997	Dominated	Ext dom	-

Abbreviations: QALY: quality adjusted life year, IFX: infliximab, ADA: adalimumab, GOL: golimumab. Ext. dom.: Extended dominance.

Search Efficiency

The usual practice took 1,440 minutes (approx. 24 hours). This included estimations of the amount of time taken for the development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening against the inclusion criteria. Table 22 displays the estimated burden that has gone into searching, using the usual search practice search method, for both the published search (covering time from database inception to Jan/Feb 2014) and an update of that search (covering the time period from Jan 2014 to Jan 2023).

The time spent doing the update search was recorded and rounded to the nearest 5-minute increments. The time for the published, original search needed to be estimated, as this was not available from the publication. The estimation was based on using the times from the update search per record screened. For example, in the update search screening 296 records at title/abstract level took on average 0.8 minutes each, resulting in total of 237 minutes. This was rounded to 240 minutes. The 0.8 minutes was also applied to the 177 titles/abstracts from the published original search, therefore it was estimated to have taken 142 minutes (rounded to 140 minutes) to sift through the titles/abstract in the original search up to January 2014. Therefore, the total time for title/abstract selection in the usual practice was 140 minutes + 240 minutes = 380 minutes. Developing the search protocols for the update search took 115 minutes, and running the searches took in total 70 minutes. It is not possible to estimate the time for the original search protocol development or running of the searches. No additional time was assumed for search protocol development and running of those original searches. It is likely to be an underestimate of the burden, and this represents a limitation of the study.

Table 22. Search tracker - Usual practice

Item	Original Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Original	Update	Total	Original	Update	Total		
Search protocol development						110		
Running searches						70		
Identified citations	195	367	562					
Duplicate removal	18	71	89	10	35	45		
Title/abstract level	177	296	473	140	240	380		
Citations excluded	124	230	354					
Full papers to retrieve + review	53	66	119	370	460	835		
Sources cited in the report	12	16	28				5%	20
Sources used in model	4	2	6				1%	94
TOTAL	**4**	**2**	**6**			**1440**		

Abbreviations: NNR: number needed to read

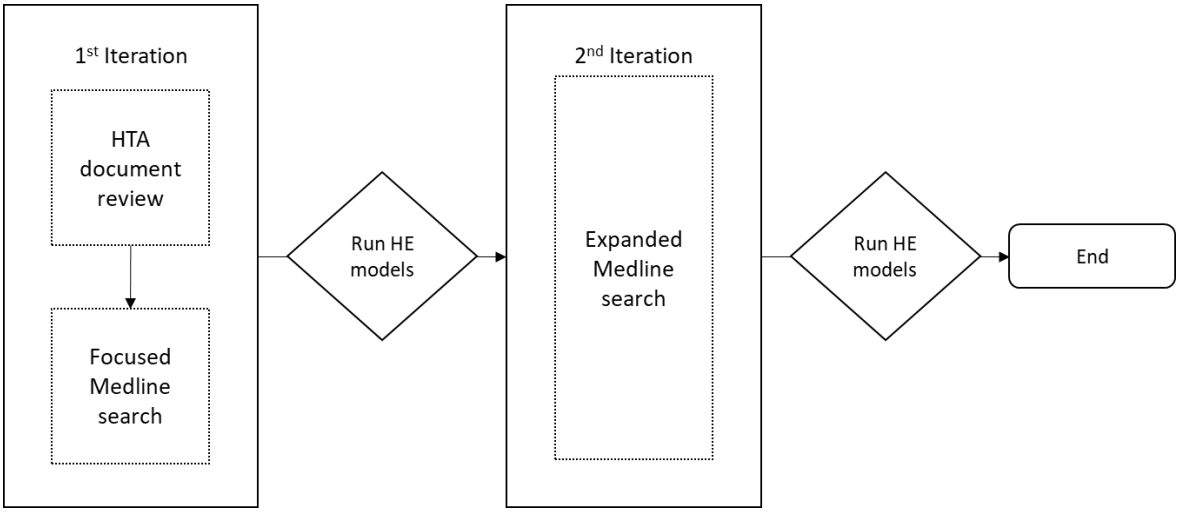
In total, 28 sources were included for consideration in the model from the usual practice search, including both the original and update searches. These sources are listed in Table 19. Six of them were used in the model, either in the base case analysis or in a sensitivity analysis. The search output had a precision of 5% in terms of items cited in the report and 1% for items used in the model (base case and sensitivity analyses). The number needed to read (NNR) was therefore 20 and 94 for sources cited in the report and sources used in the model, respectively.

5.3.2 Alternative Search Method 1

Iterative Searching

This iterative search aimed to maximise the rate of gain for relevant sources, by using searching in stages rather than in one go. The results of one iteration inform decisions about the next iteration (if applicable).[309,310] The first iteration included a highly precise bibliographic search of the most recent evidence (keywords searched in the titles only), as well as a review of HTA documents that include literature reviews of utility values for UC health economics models. The most recent evidence in the title only search was limited to the period since the last published HTA report. The first iteration aimed to identify those utility sources that were already identified through literature reviews by others as relevant for use in health economic models, as well as any new information that may have become available since the most recent HTA review (See Section "Iteration 1" for details). The second search iteration aimed to understand whether further bibliographic searching would provide additional, relevant evidence for the health economic model (see Section "Iteration 2" for details). Figure 8 shows the concept of the iterative search for this case study 1. The health economic model was run between each search iteration to assess the marginal impact of identifying additional evidence on the health economic model results, and to assess whether continuing searching is likely to identify further relevant evidence.

Figure 8. Summary of iterative search



Iteration 1

In the first search iteration, the past HTA submissions that included health economic models were identified and browsed through for any systematic literature reviews of utility values. England was chosen as the reference country for all the case studies, and therefore the NICE Technology Appraisal Guidance (TAG) documents on UC treatments were reviewed (see Table 23). The documents were reviewed to collect information on which utility studies had been identified in past literature reviews, and subsequently selected for inclusion in health economic models. The TAG documents were reviewed for

past health state utility literature reviews, as well as any *post hoc* analyses of clinical study data to derive utilities for health economic models. Such *post hoc* analyses are often not published outside the HTA documents and would not be identifiable through bibliographic database searches. The reference lists were checked for further relevant sources.

To ensure that no recent sources were missed since the most recent TAG document, a bibliographic database search was run in Medline (via ProQuest Dialog). No other databases were searched. The most recent NICE submission was upadacitinib for treating moderately-to-severely active UC (TA856), and the literature search for utilities was performed on the 6th January 2022.[311] Therefore, an update search from January 2022 to present was carried out to retrieve any records that may not have been captured by the literature reviews in the TAG documents. Both the non-bibliographic and bibliographic searching was done in January and February 2023. References were collected in a bibliographic management database.

Table 23. Iterative search: 1st iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit from January 2022 – present
Terms used	(ti(ulcerative colitis) OR ti(inflammatory bowel)) and (ti(health utility))
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	NICE Technology Appraisal Guidance in UC were reviewed: https://www.nice.org.uk/guidance/conditions-and-diseases/digestive-tract-conditions/inflammatory-bowel-disease/products?GuidanceProgramme=TA
Source names	Upadacitinib[311]: https://www.nice.org.uk/guidance/ta856 Ozanimod[296]: https://www.nice.org.uk/guidance/ta828 Filgotinib[312]: https://www.nice.org.uk/guidance/ta792 Ustekinumab[313]: https://www.nice.org.uk/guidance/ta633 Tofacitinib[314]: https://www.nice.org.uk/guidance/ta547 Vedolizumab[315]: https://www.nice.org.uk/guidance/ta342 Infliximab, adalimumab, golimumab[316]: https://www.nice.org.uk/guidance/ta329
Search dates	30 January – 9 February 2023
Limits	Past 10 years (January 2013 – January 2023)

A total of thirteen citations were selected (Table 24). Seven of these citations were also identified in the usual practice search (original + update).[274,275,295,296,298,303] One study (Arsenau *et al*. 2006) was not identified in the usual practice search, although it was used in the original model.[317] In the usual practice search, this source was identified outside the literature review where it was adopted into the Assessment Group model from one of the manufacturer models (Abbvie model).[272] Five citations that were identified in the iterative search were not identified in the usual practice search. Three of these citations were identified in a literature review of existing economic evaluations.[302,318,319] Two of the sources were NICE technology appraisals that included *post hoc* analyses of clinical trial data to produce utility values.[312,315] TA342 reported utility values from the GEMINI 1 trial by disease severity. In TA792, utility values from the SELECTION clinical study were marked confidential, and therefore the actual values were not available to be tested in the model. A total of 12 sources were identified- These are shown on Table 24, alongside the equivalent results for the usual practice searches. The utility values identified are summarised in Table 25.

Table 24. Overview of utility sources – Iterative search, iteration 1

Study	Usual Practice		Iterative search – iteration 1	
	Report (n=28)	Included in model (n=6)	Report (n=13)	Included in model (n=5)
ACT1 & ACT2 2014[33,34274,275]	✓	✓	✓	✓
Alrubaiy 2015[276]	✓			
Armuzzi 2020[277]	✓			
[a]Arsenau 2006[317]			✓	
Assche 2016[279]	✓			
Beilman 2016[278]	✓			
Biedermann 2022[280]	✓			
Burisch 2022[281]	✓			
CADTH 2019[282]	✓			
Casellas 2003[283]	✓			
Chaudhary 2013[318]			✓	
Dulai 2021[284]	✓			
GEMINI 1 2015[315]			✓	
Gherardi 2018[285]	✓			
Gibson 2014[286]	✓			
Hagelund 2020[287]	✓			
Hernandez 2020[288]	✓			
Kawalec 2018[289]	✓			
Kuruvilla 2012[290]	✓			
McLeod 1991[300]	✓			
Punekar 2010[319]			✓	
PURSUIT 2014[274,275]	✓	✓	✓	✓
Richards 2001[301]	✓			

Sardesai 2021[293]	✓	✓		
Scott 2020[294]	✓			
[b]SELECTION 2022[312]			✓	
[c]Swinburn 2012[295]	✓	✓	✓	✓
TRUENORTH 2022[296]	✓	✓	✓	✓
Tsai 2008[302]	✓		✓	
Vaizey 2013[303]	✓		✓	
Van d. Valk 2015[297]	✓			
Woehl 2008[298]	✓	✓	✓	✓
Yue Min Ho 2019[304]	✓		✓	

Footnote: a The utility weights derived by the time-trade-off method are reported, b Utility values are marked confidential in the submission, c Approximate estimate based on graph reported in Swinburn *et al*.[295].

Utility Values Used in the Model from Iteration 1

The health state utility values from Woehl *et al*. and Swinburn *et al*. were the most frequently cited utility sources among sources reviewed.[295,298] These two studies are both UK-based studies and included a relatively large number of patients compared to the other sources identified. Most importantly, they also have the greatest coverage of health states included in the model. This is important for the combinability of the utility values, as they are derived from the same group of respondents. The valuation for the surgery state in Woehl *et al*. (0.71 to 0.72) was considered more consistent with the other post-surgery valuations identified as compared with the Swinburn *et al*. by the authors of the Infliximab, adalimumab, golimumab MTA.[316] Woehl *et al*. was most often selected as the base case utility source in the models submitted to NICE. In addition to the base case, five further sources were chosen for deterministic sensitivity analyses for this iterative search: ACT1 & ACT2[274,275], PURSUIT[274,275], Swinburn *et al*[295], TRUENORTH[295], and Vaizey.[303] The utility values tested in the model are shown on Table 25. Many of the health economic models valued the post-surgery health state using values reported by Arsenau *et al*.[317] The post-surgery complication health decrement was estimated using the difference between the surgery and chronic pouchitis.

Values from Chaudhar *et al*.[318], GEMINI[315], Panekar & Hawkins[319], SELECTION[312], or Tsai *et al*.[302] were not tested in the model. Two of the publications (Chaudhar *et al*.[318] and Tsai *et al*.[302]) cited Woehl *et al*.[27] as a source for the utility values. Punekar & Hawkins cited a source for utility that appears to be for Crohn's disease.[319] GEMINI study utility values are not analysed according to response, but by disease severity so they do not fit the health economic model structure, although approximation could have been made to assume that Active UC = no response and Mild UC = response and/or remission.[315] However, as other sources were available that did not require this assumption, the source was de-prioritised. The utility values derived from the SELECTION study were marked as confidential and therefore were not available.[312] Table 25 reports the sources for the base case and scenarios for iteration 1. All sources identified in the iteration 1 of the iterative search were also identified in the usual practice search. Further, the usual practice search included a source that was not identified in this iterative search; Sardasei *et al* 2021.[293] This is a German study that provides utility values for the three main health states. This study

used the German EQ-5D tariff which may not reflect societal preferences for UK patients. The scenario results from the usual practice search were similar to those from TRUENORTH scenario, which was identified by this alternative search method.[296]

Table 25. Health state utility values in the model - Iterative search, iteration 1

Analysis	Source	Remission	Response	Active UC
Base case	Woehl *et al*.	0.87	0.76	0.41
SA1: Remission and response from ACT1	ACT1[274,275]	0.88	0.82	NR
SA 2: Remission and response from PURSUIT	PURSUIT[274,275]	0.89	0.80	NR
SA3: All utilities except post-surgical complications from Swinburn	Swinburn[295]	0.91	0.80	0.55
SA4: Remission, response and active UC from TRUENORTH	TRUENORTH[296]	0.90	0.84	0.68

Health Economic Model Results

Colectomy dominates all other treatment options in the reference case, sensitivity analysis 1 and sensitivity analysis 2. Sensitivity analyses 3 and 4 reverse the results, so that colectomy becomes the least effective option, instead of the most effective. The results are shown on Table 26.

Table 26. Health economic model results – Iterative search, iteration 1

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Base case	Dominated	Dominated	Dominated	Dominated	Dominating
Sensitivity analysis 1: Remission and response from ACT1	Dominated	Dominated	Dominated	Dominated	Dominating
Sensitivity analysis 2: Remission and response from PURSUIT	Dominated	Dominated	Dominated	Dominated	Dominating
Sensitivity analysis 3: All utilities except post-surgical complications from Swinburn	£178,982	£79,714	Dominated	Ext dom	-
Sensitivity analysis 4: Remission, response and active UC from TRUENORTH	£282,898	£128,997	Dominated	Ext dom	-

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab, SA: sensitivity analysis, SLR: systematic literature review, Ext. Dom: extended dominance.

Iteration 2

In the first iteration the search was started as a focused search, and this was then expanded in the second iteration to assess the marginal benefits of performing further searching. Iteration 1 was based on a combination of information found on 1) HTA appraisal documents (no time limit but latest HTA appraisal was from January 2022) and 2) a supplementary Medline literature review (from January 2022 - present). The HTA reviews may not have captured all relevant utility information prior to January 2022, and therefore gaps in information may exist from that period. There might be further health economic models or utility studies with unique information that are relevant for the model. This second iteration expands the Medline search from iteration 1 to uncover further relevant health state utility information, from a time period not previously covered by the Medline search. In iteration 1, Medline was searched from January 2022 to present. In this second iteration, Medline was searched using the same search terms but without any time limit. The details of the search are shown in Table 27.

Table 27. Iterative search: 2nd iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title only but without time limit, to expand on the search done in iteration 1
Type of studies	No limitation
Sources	MEDLINE (via P ProQuest Dialog)
Limits	No Limit
Terms used	(ti(ulcerative colitis) OR ti(inflammatory bowel)) and (ti(health utility))
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

In addition to the citations identified in iteration 1, a further six sources were identified in iteration 2 (Table 28), making the total identified sources 18 for the iterative search. Of the newly identified sources, four had also been identified in the usual practice search.[277,278,289,291] The remaining two sources had not been previously identified either in the usual practice search or iteration 1.[300,320 300,320]

Table 28. Overview of utility sources considered – Iterative search, iteration 2

Study	Usual Practice		Iterative search – iteration 1 + 2	
	Report (n=28)	Model (n=6)	Report (n=19)	Model (n=6)
ACT1 & ACT2 2014[33,34274,275]	✓	✓	✓	✓
Alrubaiy 2015[276]	✓			
Armuzzi 2020[277]	✓		✓	
[a]Arsenau 2006[317]			✓	
Assche 2016[279]	✓			
Beilman 2016[278]	✓		✓	
Biedermann 2022[280]	✓			
Burisch 2022[281]	✓			
CADTH 2019[282]	✓			
Casellas 2003[283]	✓			
Chaudhary 2013[318]			✓	
Dulai 2021[284]	✓			
GEMINI 1 2015[315]			✓	
Gherardi 2018[285]	✓			
Gibson 2014[286]	✓			
Hagelund 2020[287]	✓			
Hernandez 2020[288]	✓			
Kawalec 2018[289]	✓		✓	
Kuruvilla 2012[290]	✓			
Leidl 2012[291]			✓	✓
McLeod 1991[300]	✓		✓	
Poole 2010[321]				
Punekar 2010[319]			✓	
PURSUIT 2014[274,275]	✓	✓	✓	✓
Richards 2001[301]	✓			
Sardesai 2021[293]	✓	✓		
Scott 2020[294]	✓			
[b]SELECTION 2022[312]			✓	

[c]Swinburn 2012[295]	✓	✓	✓	✓
TRUENORTH 2022[296]	✓	✓	✓	✓
Tsai 2008[302]	✓		✓	
Vaizey 2013[303]	✓		✓	
Van d. Valk 2015[297]	✓			
Waljee 2011[320]			✓	
Woehl 2008[298]	✓	✓	✓	✓
Yue Min Ho 2019[304]	✓		✓	

Footnote: a The utility weights derived by the TTO method are reported, b Utility values are marked confidential in the submission, c Approximate estimate based on graph reported in Swinburn *et al*.[295].

Utility Values Used in the Model from Iteration 2

Initially, several key word combinations were tested to be used in searching the titles. The patches were browsed to see if the retrieval appears relevant or should be discarded. This was repeated six times until a patch with suitably high relevance was generated. The search string from iteration 1 provided the highest relevance, and in this second iteration it was applied without time limits. In this final patch, the focused Medline search identified 62 records. Those results were browsed and only relevant information was selected for downloading. Several records could be excluded quickly as they were not related to the topic. However, several publications were checked at full text level, including multiple cost-effectiveness and cost-utility analyses that may contain unique sources of utility values. Most model publications referred to the same sources that had already been found during iteration 1 and were not extracted again.[288,322-335] Some of the health economic models included new utility data sources, but the values were not compatible with the model health states and were therefore not included.[285,335]

One of the new data sources from 2012 (Leidl *et al*.) provided UK and German data for all three main health states (remission, response and active UC).[291] The utility values were 0.91, 0.74 and 0.63 for remission, response and active UC, respectively. The model result show that colectomy was the least effective option (see Table 29).

Table 29. Health economic model results – Iterative search, iteration 2

Analysis	IFX	ADA	GOL	Conventional management	Colectomy
Leidl[291]	£278,628	£100,718	Dominated	Ext dom	-

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab, Ext. Dom: extended dominance.

Decision to Stop Searching

Six new sources were identified during iteration 2 for consideration to be used in the model (Table 28). None of these sources provided more complete coverage of utility data for the health states than what had already been found in iteration 1. Further, these new references from iteration 2 were from countries other than the UK or used methods less preferred by NICE, and were therefore not as suitable as the sources identified in iteration 1 for our case study, since the reference country is England. Further, running the model using values reported by Leidl *et al*. resulted in very similar results to sensitivity analysis 4 (TRUENORTH) from iteration 1. The search iteration 2 did not identify any further relevant sources that would be more preferable to the values identified during iteration 1 and that would result in notably different model results. Therefore, a decision was made to stop searching after iteration 2, based on the marginal relevance of further studies retrieved in iteration 2.

Summary of Iterative Search Results and Efficiency

In total, 19 sources were included for consideration in the model from iterative searching (iteration 1 + 2), and six were actually included in the model analyses (Figure 9). These publications were listed in Table 25. All the publications used in the model from iterative searching were also identified during the usual practice search. The iterative search output had a precision of 23.17% in terms of items cited in the report and 7.32% for items used in the model. The number needed to read was therefore 4 and 14 for sources cited in report and sources used in the model, respectively.

Figure 9. Summary of results iterative searching UC

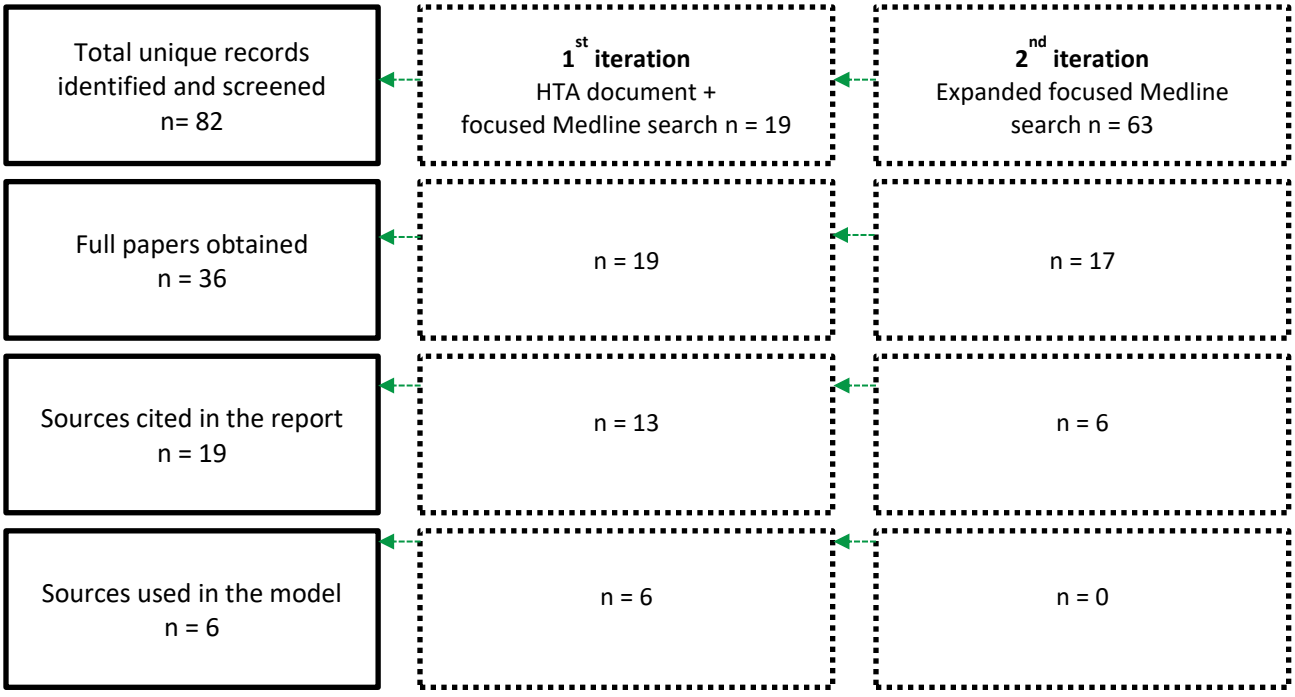


Table 30 displays the estimated search burden, separately for iteration 1 and 2 as well as the total. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening for the two iterations took 380 minutes (6 hours and 20 minutes).

Table 30. Search tracker – Iterative search

Item	Iterative search							
	Number of studies identified			Time in minutes			Precision	NNR
	Iteration 1	Iteration 2	Total	Iteration 1	Iteration 2	Total		
Search protocol development				30	30	60		
Running searches				15	20	35		
Identified citation	19	63	82					
Duplicate removal	0		0	0	0	0		
Citations at title/abstract level	19	63	82	15	50	65		
Citations excluded	0	44	44					
Full papers to retrieve + review	19	17	36	120	100	220		
Sources cited in the report	13	6	19				23%	4
Sources used in model	5	1	6				7%	14
TOTAL	**5**	**1**	**6**			**380**		

Abbreviations: NNR: number needed to read

5.3.3 Alternative Search Method 2

Rapid Review

The second experiment in the UC case study is using rapid review methods. The key difference to the first experiment (iterative searching) is that in iterative searching, 'patches' of information are evaluated one at a time and further searching is initiated as judged relevant. In rapid review, the method is similar to systematic literature review in that the search is run only once using a pre-defined search protocol. However, limits are added to manage the scope of the literature review.

For this rapid review, Medline (via ProQuest Dialog) was searched from inception. The search strategy combined free-text and MeSH terms relating to UC with terms for specific utility measures and more general utility terms. An overview of the search strategy is shown in Table 31, with further details of the specific search terms shown in Table 32. The rapid review limitations applied included using only one

database (Medline), restricting synonyms, using MeSH terms and searching utility methods only in the titles and abstracts. Literature searches were conducted during January and February 2023. References were collected in a bibliographic management database. The reference lists were checked for further relevant sources, especially to trace back the sources that have original utility values. The search results are summarised in Figure 10.

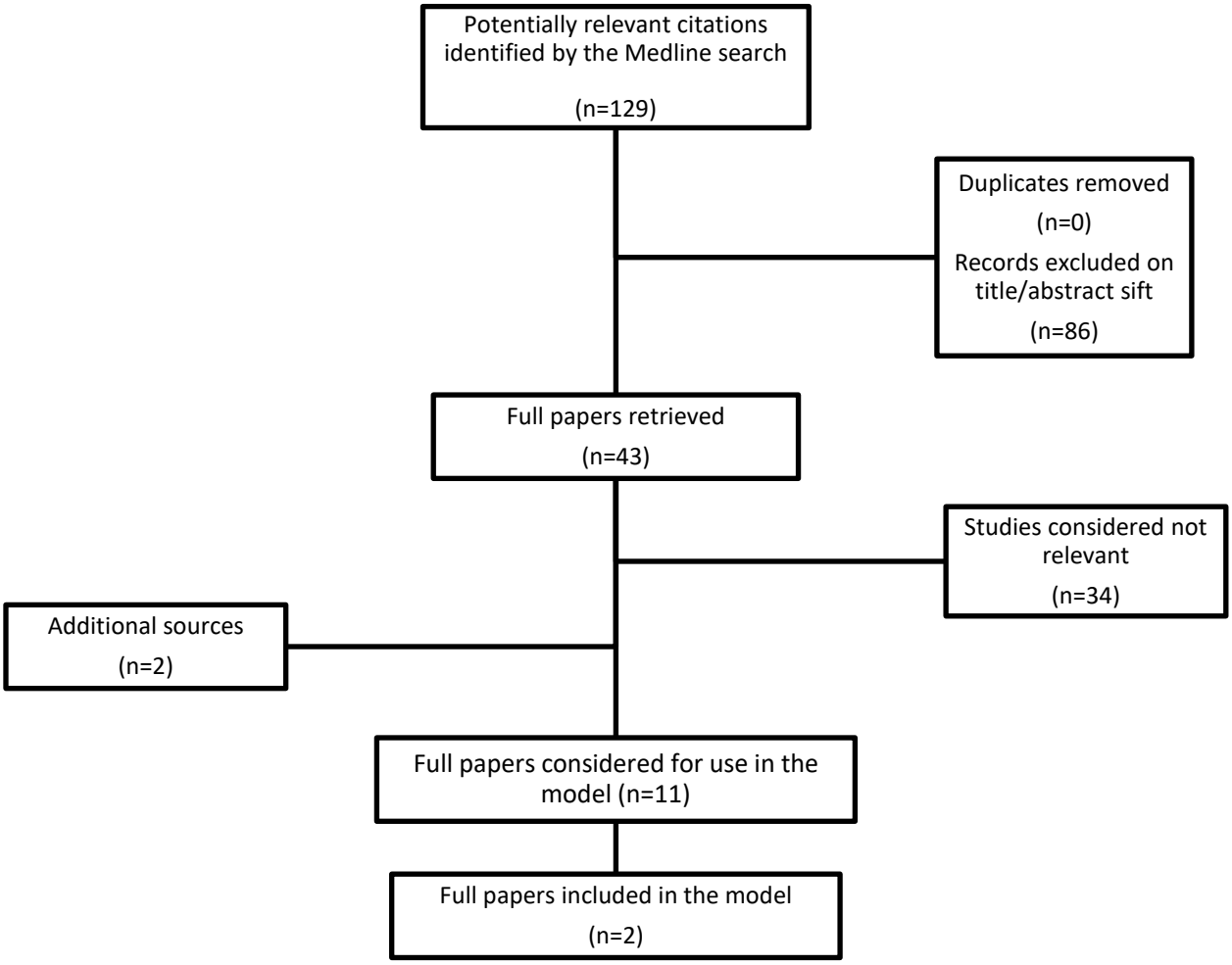
Table 31. Rapid Review

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms, limiting part of the search to abstract only; utility method keyword should be included in the title/abstract
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	No time limit
Terms used	Quality of life AND ulcerative colitis AND utility method (see Table 32 for details)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Table 32. Medline (via ProQuest Dialog) search strategy – Rapid Review

Search Strategy	Count
1. (MESH.EXACT("Quality of Life")) OR (MESH.EXACT("Quality-Adjusted Life Years")) OR (qaly) OR (quality AND (life OR wellbeing)) OR (health gain) OR ti,ab(disutility) OR ti,ab(utility) OR ti,ab(utilities)	824,943
2. (MESH.EXACT("Colitis, Ulcerative")) OR (MESH.EXACT("Inflammatory Bowel Diseases")) OR (Inflammatory Bowel Disease) OR (Ulcerative Colitis) OR (Colitis Gravis) OR (Inflammatory Bowel Disease) OR (Inflammatory Bowel Diseases)	91,864
3. ti,ab(hui) OR ab(standard gamble) OR ti,ab(euro qol) OR ti,ab(eq-5d) OR ti,ab(eq5d) OR ti,ab(eq 5d) OR ti,ab(euroqol) OR ti,ab(tto) OR ti,ab(time AND (trade off OR tradeoff)) OR ti,ab(person AND (trade off OR tradeoff))	26,497
1 AND 2 AND 3	129

Figure 10. Study selection results – Rapid review



Eleven studies were selected as potential sources to use in the model. The results of the searches are summarised in Table 33. Most of the sources identified in this rapid review method, were also identified in the usual practice search.[277,279,286,291,298,303,317] Three sources selected during the rapid review, were not identified in the usual practice search.[302,319,321] Importantly, several scenario analyses from the usual practice search were not identified using this rapid review method: including Swinburn *et al*. 2012, ACT1, PURSUIT, maintenance trial, Sardesei *et al*., and TRUENORTH.[274,275,293,295,296] The main reason for missing out many of the key publications in the rapid review was related to the simplified way of searching for utility/quality of life related terms in the titles/abstracts only, and not searching in HTA publications.

Table 33. UC case study: Studies included – Rapid Review

Study	Usual Practice		Iterative search		Rapid Review	
	Report (n=28)	Included in model (n=6)	Report (n=18)	Included in model (n=6)	Report (n=11)	Included in model (n=2)
ACT1 & ACT2 2014[33,34274,275]	✓	✓	✓	✓		
Alrubaiy 2015[276]	✓					
Armuzzi 2020[277]	✓		✓		✓	
[a]Arsenau 2006[317]			✓		✓	
Assche 2016[279]	✓				✓	
Beilman 2016[278]	✓		✓			
Biedermann 2022[280]	✓					
Burisch 2022[281]	✓					
CADTH 2019[282]	✓					
Casellas 2003[283]	✓					
Chaudhary 2013[318]			✓			
Dulai 2021[284]	✓					
GEMINI 1 2015[315]			✓			
Gherardi 2018[285]	✓					
Gibson 2014[286]	✓				✓	
Hagelund 2020[287]	✓					
Hernandez 2020[288]	✓					
Kawalec 2018[289]	✓		✓			
Kuruvilla 2012[290]	✓				✓	
Leidl 2012[291]			✓	✓	✓	✓
McLeod 1991[300]	✓		✓			
Poole 2010[321]					✓	
Punekar 2010[319]			✓		✓	
PURSUIT 2014[274,275]	✓	✓	✓	✓		
Richards 2001[301]	✓					
Sardesai 2021[293]	✓	✓				
Scott 2020[294]	✓					
[b]SELECTION 2022[312]			✓			
[c]Swinburn 2012[295]	✓	✓	✓	✓		
TRUENORTH 2022[296]	✓	✓	✓	✓		
Tsai 2008[302]	✓		✓		✓	
Vaizey 2013[303]	✓		✓		✓	

Van d. Valk 2015[297]	✓					
Waljee 2011[320]			✓			
Woehl 2008[298]	✓	✓	✓	✓	✓	✓
Yue Min Ho 2019[304]	✓		✓			

Footnote: a The utility weights derived by the TTO method are reported, b Utility values are marked confidential in the submission, c Approximate estimate based on graph reported in Swinburn *et al*.[295].

Utility Values Used in the Model from Rapid Review

The health state utility values from Woehl *et al*. and Arsenau provided the most complete sets of utilities.[298,317] Woehl *et al*. is a UK based study and therefore suitable as a base case for this case study is England. Leidl was included as sensitivity analyses.[303] All utility values included in the model are listed in Table 34. Values from Armuzzi[277], Assche[279] , Gibson[286], Kuruvilla[290], Poole[321], Punekar[319] and Tsai[302] were not tested in the model. The sample sizes were smaller than in Woehl and/or the values did not provide more comprehensive set of utilities to test in the model. One of the publications (Tsai *et al*.[302]) cited Woehl *et al*.[27] and Arsenau[317] as a sources for the utility values, and therefore provided no original utility data. As noted in Section 5.3.2, Punekar & Hawkins cited a source for utility that appears to be for Crohn's disease.[319]

Table 34. Health state utility values – Rapid Review

Analysis	Source	Remission	Response	Active UC
Base case	Woehl[298]	0.87	0.76	0.41
Sensitivity analysis 1: Remission, response, and active UC from Leidl	Leidl[291]	0.90	0.74	0.63

Health Economic Model Results

The results are shown in Table 35. Colectomy dominated all other treatments in the base case. Sensitivity analysis 1 reverses the results, so that colectomy becomes the least effective option, instead of the most effective.

Table 35. Deterministic results – Rapid Review

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Base case	Dominated	Dominated	Dominated	Dominated	Dominating
Sensitivity analysis 1: Leidl[291]	£278,628	£100,718	Dominated	Ext dom	-

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab, Ext. Dom: extended dominance.

Summary of Rapid Review Results and Efficiency

In total, 11 full texts were included for consideration in the model, and values reported in two studies were included in the model analyses. Table 36 displays the number of sources found by the search and

the estimated search burden. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening took 440 minutes (7 hours and 20 minutes).

The search output had a precision of 8.53% in terms of items cited in the report and 1.55% for items used in the model. The number needed to read was therefore 12 and 65 for sources cited in report and sources used in the model, respectively.

Table 36. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		25		
Running searches		15		
Identified citations	129			
Duplicate removal	0	0		
Citations to sift at title/abstract level	129	100		
Citations excluded	86			
Full papers to retrieve + review	43	300		
Sources cited in the report	11		8.53%	12
Sources used in the model	2		1.55%	65
TOTAL	**2**	**440**		

Abbreviations: NNR: number needed to read

5.3.4 Case Study Health State Utility in UC – Summary

The key observations from the UC utility case study were:

- Usual practice identified the most publications and took the longest to complete (Table 37).
- Iterative searching identified 12 (out of 28) of the as the usual practice search sources, and importantly, all the key sources were found.
- The burden required for the iterative search was much less (380 minutes) than what was needed for the usual practice search (1,440 minutes). This represents a 74% reduction in search and study selection time.
- No major differences were observed in the resulting health economic model results, either for the reference case or the sensitivity analyses.
- Rapid review was also more quicker than the usual practice search (440 minutes versus 1,440 minutes, which represents 69% reduction in search time), but this method did not identify all the key sources. For example, Swinburn *et al*. 2012[295] was not identified.
- The sources that were additionally identified by the alternative search methods did not make a major difference in model results. Both base case and scenario analysis results were of the same magnitude.
- In conclusion, the iterative search method identified all the relevant utility values, and it was also the most efficient search method in this UC case study and did not impact on model results.

Table 37. Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search	28	6	4.98%	1.07%	20	94	1,440	-
Iterative search	19	6	23.17%	7.32%	4	14	380	-74%
Rapid review	11	2	8.53%	1.55%	12	65	440	-69%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

5.4 Baseline Risk

5.4.1 Description of the Usual Practice (Control)

The UC model includes surgery-related inputs, including the probability of elective surgery, the probability of complications of surgery, the probability of complications requiring further surgery, the probability that patients develop pouchitis and the probability of death due to surgery. The probability of having surgery was identified in a focused Medline search. No information retrieval steps were reported for the rest of the surgery-related parameters. Therefore, the usual practice search definitions that were utilised in this search included one search utilising "minimum search recommendation" and for the rest "no specified search method" was recorded. Table 38 provides an overview of the baseline risk of clinical event parameters.

Table 38. Overview of baseline risk of clinical event parameters in the ulcerative colitis model

Model Input – Probability of	Search method	Search dates in MTA	Search strategy
Elective surgery	Focused Medline search	Inception to April 2014	"ulcerative colitis/ exp" and "colectomy rate.tw."
Complications of surgery	Not reported	N/A	N/A
Complications requiring further surgery	Not reported	N/A	N/A
Patients develops pouchitis	Not reported	N/A	N/A
Death due to surgery	Not reported	N/A	N/A

Footnote: MTA: multi technology assessment, N/A: not applicable

Identification - Surgery-related Model Inputs

The assessment group undertook a focused Medline search was undertaken to identify studies reporting long-term rates of colectomy in patients with moderate to severe UC. Medline was searched from inception to April 2014 using a simple search comprising two search terms: 'ulcerative colitis/exp' and 'colectomy rate.tw.' Studies were considered for inclusion in the economic model if they reported on long-term colectomy rates and if they either related to the moderate-to-severe population as a collective group of patients, or if they reported on colectomy rates in moderate and severe UC populations separately. The search was updated in August 2023 as part of this project. Table 39 provides the Medline search strategy. More details on the original search can be found in Archer *et al*.[272]

Table 39. Medline search strategy for baseline risk values for probability of surgery

Search Strategy
1. ulcerative colitis/exp 2. colectomy rate.tw. 3. 1 and 2

The search identified 198 citations (original 70 + update 128). Of these, ten citations related to the relevant population (Table 40). Solberg *et al*. was selected for inclusion in the model, as the estimate related to population who had not been hospitalised for UC flare, therefore the estimate is less likely to over-estimate the true colectomy rate. In addition, the Solberg *et al*. study was relatively large, including 423 patients who have completed 10-year follow-up. The rest of the surgery-related parameters in the model were taken from Arai *et al*., except for mortality.[336] Arai *et al*. was selected as the source in the Assessment Group model because it had been used in the other manufacturer-developed models. Further, the model developers did not view this to be an impactful model input that would warrant further searching. The surgery-related parameters used in the model are shown on Table 41.

Table 40. Overview of baseline risk sources considered for long-term probability of colectomy (original + update)

Study	Usual Practice (n=10)	Reported rate
Actis 2007[337]	✓	24/34 (65%)
Al-Darmaki *et al*. 2017[338]	✓	282/489 (57.7%)
Constant *et al*. 2022[339]	✓	Approximately 50% (Kaplan-Meier estimate)
Gower-Rousseau *et al*. 2009[340]	✓	Approximately 25% (Kaplan-Meier estimate)
Gustavsson *et al*. 2010[341]	✓	All UC: approximately 50% Mild UC: approximately 40% Moderate UC: approximately 50% Severe UC: approximately 62%
Manetti *et al*. 2016[342]	✓	49/837 (5.5%) (Kaplan-Meier estimate)
Misra *et al*. 2015[343]	✓	3,157/43,917 (7.2%)
Molnar *et al*. 2011[344]	✓	16/110 (14.5%) steroid-responders 29/73 (39.7%) steroid-refractory Overall: 24.6%
Mocciaro *et al*. 2012[345]	✓	Infliximab group: 60% Ciclosporin group: 30%
Solberg *et al*. 2009[346]	✓	Cumulative colectomy rate after 10 years: 9.8% (95% confidence interval 7.4 – 12.4%)

Table 41. Overview of baseline risk sources considered for other baseline risk inputs

Model Input – Probability of	Value used in the model	Source
Elective surgery	0.0051 6-month rate	Solberg *et al*. 2009[346]
Complications of surgery	47.3% (140/296)	Arai *et al*. 2009[336]
Complications requiring further surgery	19%	Arai *et al*. 2009[336]
Patients develops pouchitis	5%	Arai *et al*. 2009[336]
Death due to surgery	0.03 per model cycle	UK IBD Audit[347]

Baseline Risk Health Economic Model Results – Probability of Surgery

Colectomy dominates all other treatment options in all model analyses carried out by the assessment group (Table 42).

Table 42. Deterministic health economic model results

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Base case	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 1: Probability of pouchitis doubled	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 2: Probability of chronic pouchitis halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 3: Probability of surgery halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 4: Probability of surgery based on Gower-Rousseau	Dominated	Dominated	Dominated	Dominated	Dominating

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab.

Search Efficiency

Table 43 displays the estimated burden to identify one of the surgery-related model inputs. In the focused Medline search 198 studies were identified, out of which one was used in the model and it took 240 minutes. The search output had a precision of 3.03% in terms of items cited in the report and 0.51% for items used in the model. The NNR was therefore 33 and 198 for sources cited in report and sources used in the model, respectively. The burden to identify the rest of the surgery-related model inputs is unknown.

Table 43. Search tracker - Usual practice

Item	Original Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Original	Update	Total	Original	Update	Total		
Search protocol					10	10		
Running searches					10	10		
Identified citations	70	128	198					
Duplicate removal				0	0	0		
Title/abstract	70	128	198	60	90	150		
Citations excluded	64	124	188					
Full papers	6	4	10	40	30	70		
Sources in the report	6		6				3.03%	33
Sources in model	1		1				0.51%	198
TOTAL			**1**			**240**		

Abbreviations: NNR: number needed to read

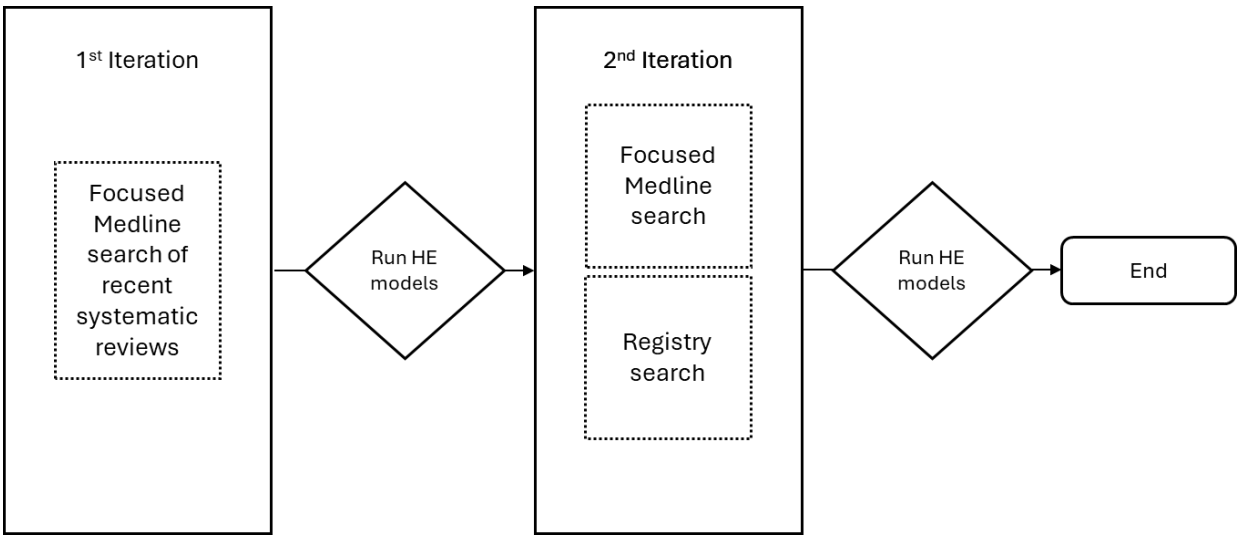
5.4.2 Alternative Search Method 1

Iterative Searching

The first iteration included a precise bibliographic search in titles only. The first iteration aimed to identify some evidence for each of the surgery-related parameters in the model. This approach was more extensive than usual practice, where no search was recorded for four out of the five inputs. The subsequent iteration aimed at understanding whether additional searching would provide further relevant information. For utility inputs (Section 5.3.2), HTA documents were searched in the first iteration. This was not done for baseline risk of clinical events. This is because utility values derived from clinical studies (i.e., those that are most likely to be the most relevant for the population assessed) are often only found as post-hoc analyses in HTA submissions. Therefore, the importance of HTA documents for utility inputs is high. This was not the case for the surgery-related inputs, as that relate to surgeries more general and are likely to come from a variety of sources (RCTs, observational studies, registries) rather than the pivotal phase 3 study. Registry search was additionally done to capture data outside what is reported in scientific journals.

Figure 11 shows the concept of the iterative search. The health economic model was run between each search iteration to assess the marginal impact of identifying additional evidence on the health economic model results, and to assess whether continuing searching is likely to identify further relevant evidence. Testing in the model was done for studies identified from bibliographic databases where input values could be extracted.

Figure 11. Summary of alternative search method 1: Iterative search (UC case study)



Iteration 1

In the first search iteration, a focused bibliographic database search was run. Details of these searches are summarised in Table 44. The bibliographic searching was done in July and August 2023. References were collected in a bibliographic management database. A total of 429 studies were identified in the focused search. Of these, 15 studies were identified that reported data for one of the surgery parameters of interest. The colectomy rates identified from these sources are summarised in Table 45, and the rest of the surgery-related values in Table 46. These two tables also contain comparison to sources identified during the usual practice search.

Table 44. Iterative search: 1st iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title/abstract only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit from January 2003 – present
Terms used	To search for elective surgery rates: (ti(ulcerative colitis) OR ti(inflammatory bowel)) AND (ti,ab("colectomy rate") OR ti,ab("surgery rate")) To search for complications of surgery or complications requiring further surgery or patients developing late pouchitis: (ti(ulcerative colitis) OR ti(inflammatory bowel) OR ti(colectomy)) AND ti,ab("outcome of complications") OR (ti,ab(complications requir*) AND ti,ab(surgery)) OR ti,ab(pouch*) To search for mortality: (ti(ulcerative colitis) OR ti(inflammatory bowel)) AND ti(mortality)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Table 45. Overview of sources considered for probability of colectomy

Study	Usual Practice (n=10)	Iteration 1 (n=11)	Reported rate
Probability of elective surgery			
Actis 2007[337]	✓		24/34 (65%)
Al-Darmaki *et al*. 2017[338]	✓		282/489 (57.7%)
Burisch *et al* 2022[348]		✓	Cumulative colectomy rate after 10 years: 22% (95% confidence interval 20 – 25%)
Constant *et al*. 2022[339]	✓		Approximately 50% (Kaplan-Meier estimate)
Dai *et al*. 2023[349]		✓	Approximately 10% (Kaplan-Meier estimate)
Eriksson *et al*. 2017[350]		✓	Cumulative colectomy rate after 10 years: 13.5% (95% CI 11.1%-15.8%)
Gower-Rousseau *et al*. 2009[340]	✓		Approximately 25% (Kaplan-Meier estimate)
Gustavsson *et al*. 2010[341]	✓		All UC: approximately 50% Mild UC: approximately 40% Moderate UC: approximately 50% Severe UC: approximately 62%
Hoie *et al*. 2007[351]		✓	Cumulative colectomy rate after 10 years: 8.7%
Kurti *et al*. 2023[352]		✓	Approximately 4.4.% (Kaplan-Meier analysis)
Manetti *et al*. 2016[342]	✓	✓	49/837 (5.5%) (Kaplan-Meier estimate)
Misra *et al*. 2015[343]	✓	✓	3,157/43,917 (7.2%)
Molnar *et al*. 2011[344]	✓	✓	16/110 (14.5%) steroid-responders 29/73 (39.7%) steroid-refractory Overall: 24.6%
Mocciaro *et al*. 2012[345]	✓	✓	Infliximab group: 60% Ciclosporin group: 30%
Parragi *et al*. 2018		✓	Cumulative colectomy rate after 10 years: 6.4%
Solberg *et al*. 2009[346]	✓	✓	Cumulative colectomy rate after 10 years: 9.8% (95% confidence interval 7.4 – 12.4%)

Table 46. Overview of sources considered for probability of surgery complications, complications requiring further surgery, development of (late) pouchitis and death from surgery

Study	Usual Practice (n=4)	Iteration 1 (n=6)	Reported rate
Surgery complications			
Arai *et al*. 2009[336]	✓	✓	47.50%
Feuerstein *et al*. 2018[353]		✓	44%
De Silva *et al*. 2011[354]		✓	27%
Complications requiring further surgery			
Arai *et al*. 2009[336]	✓	✓	19%
Development of (late) pouchitis			
Arai *et al*. 2009[336]	✓	✓	5%
Death from surgery			
Ordas *et al*. 2018[355]		✓	0.06 per model cycle
UK IBD Audit[347]	✓		0.03 per model cycle

Baseline Risk Values Used in the Model from Iteration 1

Several studies reporting probabilities of undergoing colectomy for UC were identified. The values ranged from 4.4% in a Hungarian population-based study to 22% in a Danish population-based study.[348,352] Many studies reported values around 10% at 10 years of follow-up.[343,346,349-351] Solberg *et al*. was chosen as the base case because the study reported long-term colectomy rates in moderate to severe patients, and was not primarily for patients who experienced UC flare.[346] The lowest (4.4%) and highest (22%) among the identified sources were tested as scenarios.[348,352]

Very few studies were identified for the rest of the surgery related inputs (Table 46). Three sources were identified that reported surgery complications.[336,353,354] Feuerstein *et al*. reported value of 44%, that was specific for elective colectomy population, and used in the base case instead of Arai *et al*. as it is was a newer publication (2018 versus 2005).[93,110] These studies were of similar size, Arai *et al*. from Japan and Feuerstein *et al*. from the US. Only one study was identified for each of these inputs: complications requiring further surgery, development of pouchitis and death from surgery.[336,355]

Health Economic Model Results

In the reference case, colectomy dominates all other treatment options (Table 47). All the surgery-related inputs were tested in sensitivity analyses. The model was run with the minimum/maximum values found in literature, or base case values were doubled or halved, if no alternative values were available. In all the sensitivity analyses, colectomy continues to dominate all other treatment options (Table 47).

Table 47. Deterministic results – Probability of surgery

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Reference case	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 1: Probability of surgery 4.4%[352]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 2: Probability of surgery 22%[348]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 3: Probability of surgery complications doubled	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 4: Probability of surgery complications halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 4: Probability of complications requiring further surgery doubled	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 5: Probability of complications requiring further surgery halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 6: Probability of chronic pouchitis doubled	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 7: Probability of chronic pouchitis halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 8: Probability of death halved	Dominated	Dominated	Dominated	Dominated	Dominating

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab, SA: sensitivity analysis

Iteration 2

The first iteration showed that the model was not sensitive to surgery-related parameters. The decision was made to stop searching further for these, with the exception of death from surgery. The model base case is for the UK, and only a Spanish mortality study was identified in the initial search iteration. The second iteration aimed to find further UK-specific data to inform the death from surgery model input by searching further databases, as well as using revised search terms. The details of the search are shown in Table 48. The bibliographic searching was done in August 2023. References were collected in a bibliographic management database. A total of 89 studies were identified in this search. Of these, two studies were identified that reported death from surgery rates for the UK. The death rates are summarised in Table 49.

In addition to searching for UK-specific mortality data, a decision was made to search for UC-specific registries as these have potential to provide baseline risk of clinical event data that is not currently reported in RCTs. The details of the registry search are given in Section 5.4.2.1.

Table 48. Iterative search: 2nd iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling
Type of studies	No limitation
Sources	MEDLINE, BIOSIS Previews, Embase (via ProQuest Dialog)
Limits	No limitation
Terms used	<u>To search for mortality:</u> ab("the UK") OR ab(United Kingdom) OR ab(England) AND ti(ulcerative colitis) OR ti(inflammatory bowel) AND mortality
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Table 49. Overview of sources for death from surgery – Iteration 2

Study	Usual Practice	Iteration 1	Iteration 2	Rate per model cycle
Ordas *et al*. 2018[355]		✓		0.06
UK IBD Audit[347]	✓		✓	0.03
Shawihdi *et al*. 2019[356]			✓	0.05

Baseline Risk Values Used in the Model from Iteration 2

The base case death rate after surgery was changed to a UK-specific source. Shawihdi *et al*. was selected because it was recent (2019) and because it reported UK-specific death rates after surgery.[356] The results from the model are shown in Table 50. Both of the newly identified data sources return the same model conclusions i.e., colectomy dominates all other treatment options.

Table 50. Deterministic results – Probability of surgery

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Base case[356]	Dominated	Dominated	Dominated	Dominated	Dominating
SA: Probability of death from surgery[347]	Dominated	Dominated	Dominated	Dominated	Dominating

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab, SA: sensitivity analysis

Decision to Stop Searching

As all data tested in the model gave the same conclusion - that colectomy is the dominating treatment option - a decision was made to stop searching for further studies, as the likelihood that additional data that would alter the model results was low.

Summary of Iterative Search Results and Efficiency

In total, 17 studies were included for consideration in the model from iterative searching, and values from seven studies were included in the model analyses (Figure 12). The usual practice search identified five of the same studies for colectomy rates as iterative searching.[342-346] In the usual practice search, Solberg *et al*. was used in the base case.[346] In the iterative search, the same study was still a good choice to be used in the base case, as many other sources included patients with acute UC flare or were smaller. For the other surgery-related parameters, only a few studies were identified in the iterative search. It is unknown if the usual practice search identified more studies than just Arai *et al*., which was used in the model. Both the usual practice search and the iterative search identified the study by Arai *et al*.[336] This one study provided data for three of the surgery-related inputs in the model (probability of surgery complications, complications requiring further surgery, and development of pouchitis). The iterative search identified an additional source for probability of surgery complications.[353] Feuerstein *et al*. provided an alternative, albeit similar estimate, to use in the model than Arai *et al*.[353] The first search iteration did not identify UK-specific mortality studies to estimate death rates after surgery. The second search iteration focused on the identification of UK studies. Two further studies were found.[347,356] One of these had been identified by the usual practice, an earlier UK IBD Audit publication by Lynch *et al*.[14] The second publication is from the same registry, from a later date.[25] It is likely that this later publication could have also been found by the usual practice search. However, this cannot be confirmed as the steps to identify the study were not described, and therefore the search could not be updated for this model input.

The iterative search output had a precision of 3.28% in terms of items cited in the report and 1.35% for items used in the model. The NNR was therefore 30 and 74 for sources cited in the report and sources used in the model, respectively. Table 51 displays the estimated search burden, separately for iteration 1

and 2, as well as the total. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening for the two iterations took 460 minutes (7 hours and 40 minutes).

Figure 12. Summary of results – Iterative searching

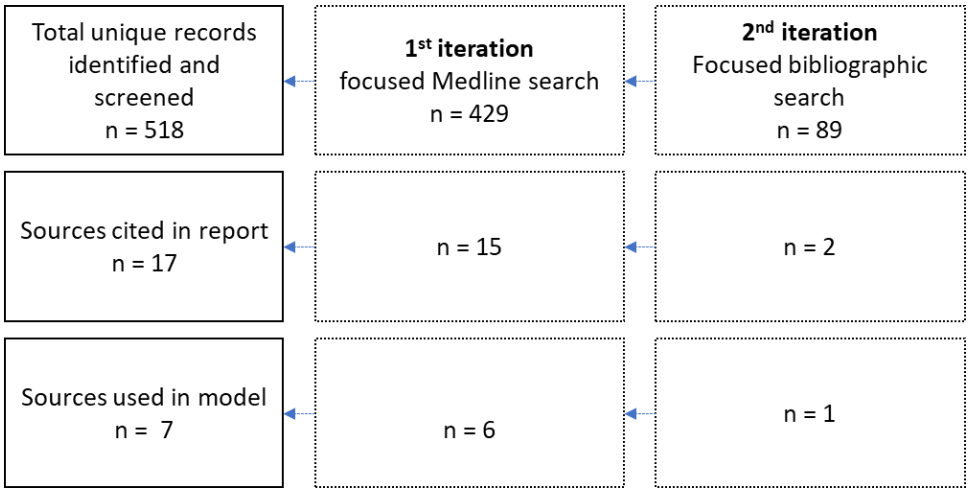


Table 51. Search tracker – iterative search

Item	Iterative search							
	Number of studies identified			Time in minutes			Precision	NNR
	Iteration 1	Iteration 2	Total	Iteration 1	Iteration 2	Total		
Search protocol				15	10			
Running searches				10	5			
Identified citation	429	89	518					
Duplicate removal								
Title/abstract level	429	89	518	260	50			
Citations excluded	410	87	497					
Full papers	19	2	21	95	10			
Sources cited in the report	15	2	17				3.28%	30
Sources used in model	6	1	7				1.35%	74
TOTAL	6	1	**7**	380	80	**460**		

Abbreviations: NNR: number needed to read

5.4.2.1 Registry Search

When assessing cost-effectiveness, relying solely on RCT data may not offer enough insights for decision makers.[357] In such situations, alternative sources of evidence, such as patient registries, can be important. Registries can provide valuable real-world evidence related to clinical practice, patient outcomes, safety, and comparative effectiveness, and can therefore be of special importance when searching input data for baseline risk of clinical events.[358] NICE also states in the methods manual that evidence from registries is considered.[44]

A targeted web search was done to identify UC registries that may contain useful information for estimating the baseline risk of clinical events. One registry was identified from clinicaltrials.gov (CorEvitas-IBD). One registry was identified through google.com (UK Inflammatory bowel disease [IBD] registry). The rest were identified through google scholar (Competency network, ENEIDA, EPIMAD, SWIBREG and SIBDC). No input values could be extracted, as access to registries typically requires subscription. Therefore, for registry search potential sources are listed but not tested in the model as the databases could not be accessed. The search details are provided on Table 52, and the identified registries are shown on Table 53. The search time was 320 minutes.

Table 52. Registry search

Bibliographic database search elements	Considerations
Sampling	None
Type of studies	None
Sources	None
Limits	None
Terms used	None
Conceptual limitations	None
Non-database search elements	**Considerations**
Approaches	Targeted searches were used to identify publications/websites for registries. National and international European registries were searched. Search terms related to ulcerative colitis/irritable bowel syndrome and registry/database/cohort study were used.
Source names	https://clinicaltrials.gov/ https://scholar.google.com/ https://www.google.com/
Search dates	August 2023
Limits	Not applicable

Table 53. Overview of registries identified for UC surgery-related input

Registry name	Location	Year established	Current status	Sample size	Reference
Competence Network	Germany	1999	Active	4,000	Competency Network IBD[359]
CorEvitas-IBD	US	2017	Active	1,000	NCT03162549[360]
ENEIDA	Spain	2016	Final report expected Sep 2023	3,200	ENEIDA[361]
EPIMAD	France	2015	Completed in 2019	966	EPIMAD[362]
SWIBREG	Sweden	2005	Active	61,153	SWIBREG[363]
SIBDC	Switzerland	2005	Active	3,000	SIBDC[364]
UK IBD registry	UK	2012	Active	4,500	UK IBD[365]

5.4.3 Alternative Search Method 2

Rapid Review

For this rapid review, Medline (via ProQuest Dialog) was searched from inception. The search strategy is given below in Table 54, with more details provided in Table 55. The rapid review limitations included searching only one database, using free-text search and searching in titles/abstracts. The literature search was conducted in August 2023. References were collected in a bibliographic management database. The reference lists were checked for further relevant sources. The search results are summarised in Figure 13.

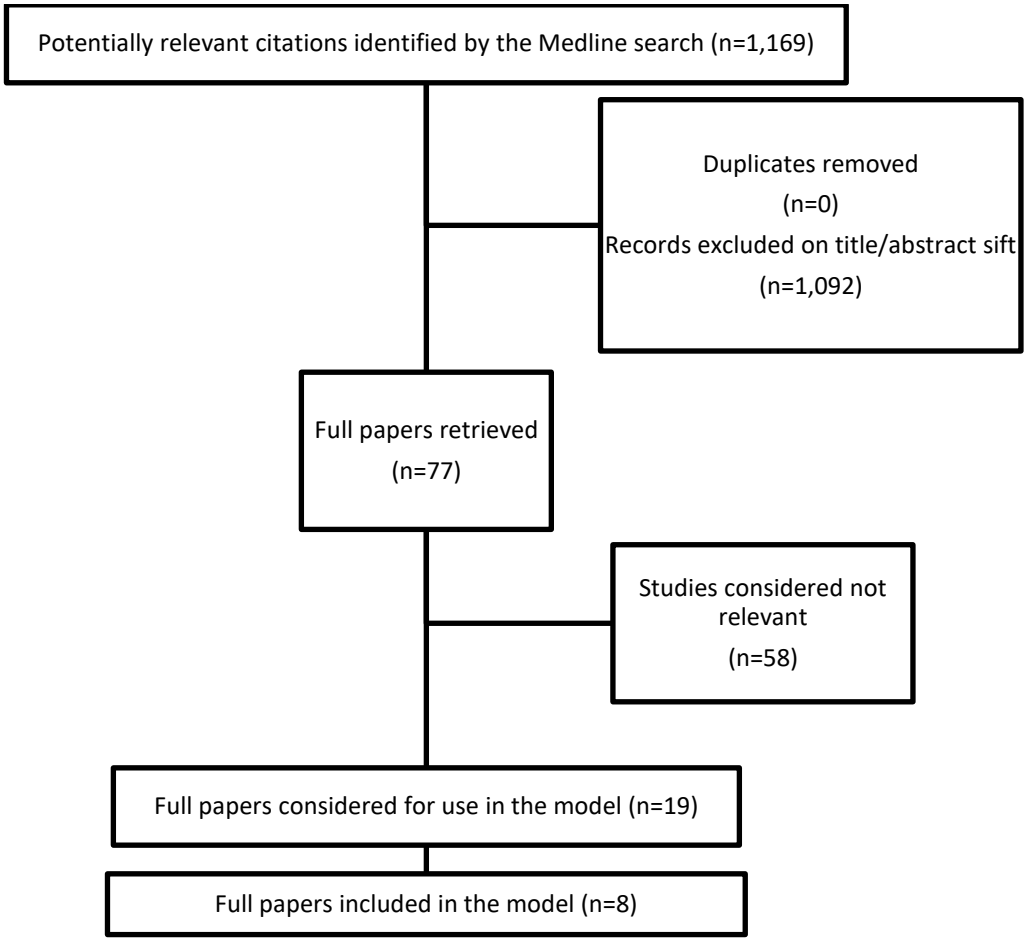
Table 54. UC Case study: Alternative search – Rapid Review

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms, limiting part of the search to abstract only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	20 years i.e., 2003 – Aug 2023
Terms used	(colectomy AND complication) OR (colectomy AND ulcerative colitis) *See Table 55 for detailed search strategy*
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Table 55. Medline search strategy – Rapid Review

Search Strategy
(ti(colectomy OR proctocolectomy) AND ti(complication*)) OR (ti(ulcerative colitis) AND ti(colectomy OR protocolectomy OR outcome* OR pouchitis OR disease progression OR disease course OR clinical course OR clinical progression))

Figure 13. Study selection results – Rapid review



The search identified 19 potential sources to be used in the model.[336,338,341,342,346,348-352,354,356,366-371] Six of these were only found in this rapid review,[366-371] while the rest had also been identified in either the usual practice search or iterative search. The colectomy rates from the newly identified sources were similar to those that had been identified by other search methods as well. Previously only Arai *et al*. had been identified as a source for complications requiring further surgery. This rapid review identified two additional sources in addition to Arai.[368,371] Both Nunez *et al*. and Worley *et al*. reported higher rates of complications than Arai *et al*. (around 30% compared to about 20%).[368,371] The newly identified sources for the rest of the inputs (rates of surgery complications, development of pouchitis and death from surgery) were similar to those identified by other search methods. The results of the searches are summarised in Table 56 and Table 57.

Table 56. Overview of sources considered for probability of colectomy

Study	Usual Practice (n=10)	Iterative search (n=11)	Rapid Review (n=14)	Reported rate
Probability of elective surgery				
Actis 2007[337]	✓			24/34 (65%)
Al-Darmaki *et al*. 2017[338]	✓		✓	282/489 (57.7%)
Burisch *et al* 2022[348]		✓	✓	Cumulative colectomy rate after 10 years: 22% (95% confidence interval 20 – 25%)
Constant *et al*. 2022[339]	✓			Approximately 50% (Kaplan-Meier estimate)
Dai *et al*. 2023[349]		✓	✓	Approximately 10% (Kaplan-Meier estimate)
Eriksson *et al*. 2017[350]		✓	✓	Cumulative colectomy rate after 10 years: 13.5% (95% CI 11.1%-15.8%)
Eronen *et al*. 2023[366]			✓	Approximately 45%
Gower-Rousseau *et al*. 2009[340]	✓			Approximately 25% (Kaplan-Meier estimate)
Gustavsson *et al*. 2010[341]	✓		✓	All UC: approximately 50% Mild UC: approximately 40% Moderate UC: approximately 50% Severe UC: approximately 62%
Hoie *et al*. 2007[351]		✓	✓	Cumulative colectomy rate after 10 years: 8.7%
Kurti *et al*. 2023[352]		✓	✓	Approximately 4.4.% (Kaplan-Meier analysis)
Manetti *et al*. 2016[342]	✓	✓	✓	49/837 (5.5%) (Kaplan-Meier estimate)
Misra *et al*. 2015[343]	✓	✓		3,157/43,917 (7.2%)
Molnar *et al*. 2011[344]	✓	✓		16/110 (14.5%) steroid-responders 29/73 (39.7%) steroid-refractory Overall: 24.6%
Mocciaro *et al*. 2012[345]	✓	✓		Infliximab group: 60% Ciclosporin group: 30%
Monstad *et al*. 2021[367]			✓	10 year cumulative rate 5% (Kaplan-Meier estimate)
Parragi *et al*. 2018		✓	✓	Cumulative colectomy rate after 10 years: 6.4%
Senanayake *et al*. 2013[369]			✓	Cumulative colectomy rate after 10 years: 5.4%
Solberg *et al*. 2009[346]	✓	✓	✓	Cumulative colectomy rate after 10 years: 9.8% (95% confidence interval 7.4 – 12.4%)
Targownik *et al*. 2012[370]			✓	Cumulative colectomy rate after 10 years: 10.4%

Table 57. Overview of sources considered for probability of surgery complications, complications requiring further surgery, development of (late) pouchitis and death from surgery

Study	Usual Practice (n=4)	Iterative search (n=7)	Rapid Review (n=7)	Reported rate
Surgery complications				
Arai *et al*. 2009[336]	✓	✓	✓	47.50%
Feuerstein *et al*. 2018[353]		✓		44%
De Silva *et al*. 2011[354]		✓	✓	27%
Complications requiring further surgery				
Arai *et al*. 2009[336]	✓	✓	✓	19%
Nunez *et al*. 2023[368]			✓	32%
Worley *et al*. 2018[371]			✓	30%
Development of (late) pouchitis				
Arai *et al*. 2009[336]	✓	✓	✓	5%
Death from surgery				
Ordas *et al*. 2018[355]		✓		0.06 per model cycle
UK IBD Audit[347]	✓			0.03 per model cycle
Shawihdi *et al*. 2019[356]		✓	✓	0.05 per model cycle

Footnote: Abbreviations: NR: not reported.

Baseline Risk Values Used in the Model from Rapid Review

Fourteen studies reporting values for colectomy were identified. The values ranged from 4.4% in a Hungarian population-based study to 62% in a Swedish-Danish population-based study.[341,352] Solberg *et al*. was chosen as the base case, because the study reported long-term colectomy rates in moderate-to-severe patients, and was not primarily for patients who experienced UC flare.[346] The lowest (4.4.%) and highest (62%) among the identified sources were tested as scenarios.[341,352] As with the other search methods, not many studies were identified for the rest of the surgery-related inputs (Table 57). Two sources were identified that reported surgery complications.[336,353,354] De Silva *et al*. reported a value of 27%, that was based on a study population of 666 patients, as compared to only 296 patients in the study

by Arai *et al*.[336,354] The study by de Silva is also more recent, and therefore was used in the base case. The alternative source provides much higher rate and it was included as a scenario analysis.

Three sources were identified as containing inputs for complication rates requiring further surgery.[336,368,371] Worley *et al*. was selected as a base case. The study was conducted in English-Swedish populations, and it is fairly large and recent. The lower estimate by Arai *et al*. is tested as a scenario analysis. For two of the model inputs (development of pouchitis and death from surgery), only one source was identified for each; Arai *et al*. and Shawihdi *et al*. respectively.[336,356] These were therefore used in the model.

Health Economic Model Results

In the reference case, colectomy dominates all other treatment options (Table 58). All the surgery related parameters were tested in sensitivity analyses. The model was run either with the minimum and maximum values found in literature, or base case values were doubled or halved if no alternative values were available. In all the sensitivity analyses, colectomy continues to dominate all other treatment options (Table 58).

Table 58. Deterministic results – Rapid Review

Analysis	IFX	ADA	GOL	Conventio-nal mgmt.	Colectomy
Reference case	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 1: Probability of surgery 4.4%[352]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 2: Probability of surgery 62%[369]	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 3: Probability of surgery complications 47.5%	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 4: Probability of complications requiring further surgery 19%	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 5: Probability of chronic pouchitis doubled	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 6: Probability of chronic pouchitis halved	Dominated	Dominated	Dominated	Dominated	Dominating
Scenario 7: Probability of death halved	Dominated	Dominated	Dominated	Dominated	Dominating

Abbreviations: IFX: infliximab, ADA: adalimumab, GOL: golimumab.

Summary of Rapid Review Results and Efficiency

Table 59 displays the estimated effort to identify surgery-related model inputs. In the focused Medline search 1,169 studies were identified, out of which eight were used in the model. This effort was estimated to take 1,535 minutes. The search output had a precision of 1.63% in terms of items cited in the report and 0.68% for items used in the model (base case and sensitivity analyses). The NNR was therefore 62 and 146 for sources cited in the report and sources used in the model, respectively.

Table 59. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		35		
Running searches		20		
Identified citations	1,169			
Duplicate removal	0			
Citations to sift at title/abstract level	1,169	935		
Citations excluded	1,092			
Full papers to retrieve + review	77	540		
Full papers selected for consideration	19			
Sources cited in the report	19		1.63%	62
Sources used in the model	8		0.68%	146
TOTAL	**8**	**1,530**		

Abbreviations: NNR: number needed to read

5.4.4 Case Study Baseline Risk of Clinical Events in UC – Summary

The key observations from the UC baseline risk of clinical events case study were:

- The UC model included five surgery-related baseline risk of clinical event parameters.
- In the usual practice search, only one of the baseline risk parameters was associated with a description of information retrieval steps. For the other four, no details were provided on how they came to be incorporated in the model.
- The alternative search methods (iterative searching and rapid review) recorded information retrieval steps for all five model inputs, thereby increasing transparency.
- The precision and NNR were similar for the usual practice and the iterative search. Rapid review had lower precision, and higher NNR (i.e., more search effort was associated with each identified source) compared with the other two methods.
- The time that it took to carry out the rapid review was three times more than to carry out the iterative search, and six times more than the usual practice search (Table 60). In rapid review several surgery-related inputs were searched in one search whereas iterative search techniques were able to utilise a 'bit-at-the-time' technique, making it faster to search and select relevant sources. Usual practice took less time, mostly because only one (out of five) surgery-related inputs was associated with a search.
- Usual practice was the fastest search method but least transparent. Iterative searching was more transparent, and yet the efficiency measures (precision and NNR) were similar to usual practice. Rapid review returned the most sources but did not provide further increases in transparency and took considerably longer than both usual practice and iterative searching.
- All identified data resulted in little differences in model results. In fact, the conclusions that could be drawn from the model remained the same for all analyses conducted.

- In summary, with relatively little additional effort (3 hours and 40 minutes), improvements to model transparency could be made by using iterative searching instead of usual practice. As the actual time taken to develop the usual practice is not known, and the method of estimating it for the purposes of this case study were conservative, it is possible that there is no significant time difference. In this case study, rapid review required considerably more effort without additional gains in efficiency or model results.
- Identifying six registries potentially containing useful data for baseline risk of clinical events took 320 minutes. Number of inputs, precision or NNR could not be estimated for the registry search.

Table 60. UC case study: Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision Report	Precision Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search – Only one input	6	1	3.03%	0.51%	33	198	240	-
Iterative search – Five inputs	17	6	3.28%	1.35%	30	74	460	+90%
Rapid review - Five inputs	19	8	1.63%	0.68%	62	146	1,530	+539%
Registry search	N/A	N/A	N/A	N/A	N/A	N/A	320	+33%

Abbreviations: N/A not applicable. **Footnote:** [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

5.5 Case Study in UC – Summary

The key observations from the UC case study were:

- Iterative searching has emerged as a promising approach for utility inputs. In this case study, it offered an advantage in terms of speed compared to usual practice.
- Moreover, the most relevant sources were identified, and the model outputs obtained through iterative searching remain comparable to those achieved using usual practice.
- For baseline risk for clinical events, this was not the case. The time taken for the search was longer with iterative searching compared with usual practice, although time could only be estimated for one input instead of all five inputs. Using iterative searching, the relevant sources were identified and model results were comparable also for baseline risk of clinical events.
- One major drawback of the usual practice for baseline risk of clinical events is that the steps to identify the model inputs were not recorded for all the inputs. Using iterative searching in combination with the search reporting framework provides more transparent process for model input identification.
- The results (in terms of which sources were identified) for rapid review were similar to those of iterative searching for baseline risk of clinical events, but the search took significantly longer than the iterative search.
- Additionally, a registry search was performed. For this case study registries were searched with targeted web-based searches. Most of the registries identified were identified using google scholar searches, and few additional ones were found through google.com and clinicaltrials.gov. The data on registries could not be accessed, and therefore the impact of being able to include data from registries is unknown. Empirical research would be needed to understand the impact of including/excluding registry data.

6 Case Study 2: Thyroid Cancer

6.1 Chapter Overview

This chapter reports the findings of the thyroid cancer case study. This Chapter provides an introduction to the case study, and then reports the searching for utility (Section 6.3) and baseline risk of clinical events (Section 6.4), separately. For both model input types, usual practice search, iterative search and rapid review methods and results are reported. This includes details of how the search was carried out, what the results were, how studies were selected and used in the model. This is followed by an assessment of impact of the sources on the health economic model results, and then an assessment of using the performance measures outlines in the methods chapter (Section 4.2.2.4).

Although the different search methods employed in this case study are reported in the same order as in case studies 1 and 3, they were actually carried out in a different order, as reported in the methods section. Iterative search was carried out first, then rapid review and finally usual practice.

6.2 Case Study Introduction

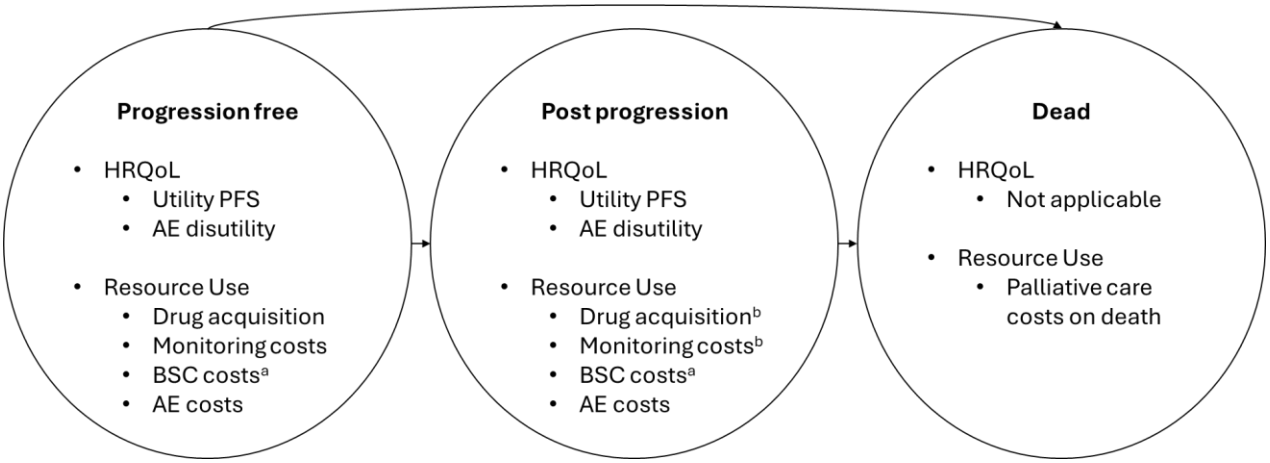
In 2019, the Health Technology Assessment journal published a systematic review and economic model assessing cabozantinib and vandetanib versus best supportive care (BSC), for the treatment of unresectable locally advanced or metastatic medullary thyroid cancer (MTC).[372] Medullary thyroid cancer is a very rare type of cancer, and is characterised by the presence of tumour masses within the thyroid gland located in the neck.[373] MTC affects both patients' HRQoL and survival.[374,375] This independent research was funded by National Institute for Health Research (NIHR) to inform a NICE Multiple Technology Appraisal of cabozantinib and vandetanib. A health economic model was developed by an independent Assessment Group to estimate the incremental cost-effectiveness of cabozantinib and vandetanib compared with each other and with BSC. The population within the Assessment Group economic analysis related to two populations of patients: (1) patients with symptomatic and progressive disease ("EU label for vandetanib") and (2) patients with symptomatic and progressive disease with carcinoembryonic antigen (CEA) and calcitonin ("restricted EU label for vandetanib"). In this case study, only the "EU label for vandetanib" population is included.

In the absence of direct head-to-head evidence comparing cabozantinib with vandetanib, an indirect comparison using a network meta-analysis (NMA) was considered by the Assessment Group. Due to significant differences in the intention-to-treat (ITT) populations of the pivotal EXAM (cabozantinib) and ZETA (vandetanib) trials, NMA was not considered suitable. The validity of the NMA relies on the assumption that there are no substantial differences in the distribution of treatment effect modifiers at the trial level between the two populations. However, this assumption is unlikely to hold true for the ITT populations of the ZETA and EXAM trials. Specifically, participants in the EXAM trial had confirmed disease progression, whereas the ZETA trial included a broader population without a requirement for established disease progression. In the health economic model, two pairwise economic comparisons were made for cabozantinib versus BSC based on the ITT population of the EXAM trial[376] and for vandetanib versus BSC

based on the *post hoc* EU-label (symptomatic and progressive) subgroup of the ZETA trial.[377,378] The Assessment Group report also included fully incremental analyses, but those alternative analyses exploring the relative treatment effect of vandetanib versus cabozantinib are not explored in this case study. The model was run for two pairwise comparisons: Cabozantinib versus BSC and vandetanib versus BSC.

The health economic model used a partitioned survival approach, based on three health states (see Figure 14). A modifiable version of the original executable Excel model was made available by one of my supervisors (PT) to test the identified utility data. Further details of the model can be found in Tappenden *et al*. 2019.[372]

Figure 14. Thyroid cancer health economic model diagram (Tappenden *et al*. 2019[372])



Footnote: a applies only to patients not receiving vandetanib or cabozantinib, b applies only to open label vandetanib costs in vandetanib vs. BSC comparison.

In this MTC case study, the type of data for baseline risk of clinical events is time-to-event data (OS and PFS). Compared to more simple constant estimates (i.e., underlying exponential model) in the previous case study e.g., the probability of undergoing surgery, more steps are needed to incorporate this type of survival data in the model. These steps have the potential to add a significant amount of time to the process that is not related to the search methods. Further, there are choices that need to be made during the incorporation of survival data that are likely to impact the model results (ICER), for example the choice of survival curve type. Another example of a choice that might need to be made is whether to estimate the risk of an event over time using a single fitted parametric survival curve or using a hybrid approach which combines the empirical Kaplan-Meier estimate and a parametric extrapolation. If the latter approach is adopted, choices also need to be made about when to switch from the KM to the fitted model estimates. OS and PFS are not only baseline risk of clinical event parameters, but they also inform relative treatment effect estimates (hazard ratios or acceleration factors), and it is therefore important to consider how new data would be incorporated in the model. For example, it should be considered whether it is necessary to perform an indirect treatment comparison, even if with the original set of evidence it was not feasible. It is also possible that registries are identified as potential data sources, and often registry data cannot be accessed without paying a fee.

6.3 Utility Values

6.3.1 Description of the Usual Practice (Control)

Utility Value Identification

In the original search reported in the Assessment Report, broad HRQoL literature searches were undertaken to identify utility studies in the literature relating to locally advanced or metastatic thyroid cancer, including MTC as well as other, more common types of thyroid cancer. This approach was taken due to the rarity of MTC, and the anticipated lack of evidence ('proxy relevance'). In the health economic model, health utility is assumed to be dependent on the presence/absence of disease progression and whether patients experience AEs. This case study 2 in the PhD project focuses on the identification of relevant health state utilities for the progression-free and progressed disease/post-progression states. The usual practice search strategies included MeSH or Emtree Thesaurus terms and free-text synonyms for 'thyroid cancer'. The authors in this case study have used more sensitive search filter for quality of life than in case study 1 (see Table 18). The Medline search strategy is presented in Table 61. The search strategy was translated across all databases searched. No date or language restrictions were applied. Literature searches were originally conducted on the 3 November 2019 by the Assessment Group. As part of this project, I updated the searches in April 2024. Full details of the original usual search methods can be found in Tappenden *et al*. 2019.[372]

Table 61. Medline (via Ovid) search strategy for utility values - Usual practice (source: Tappenden *et al*.[372])

Search Strategy
1. exp Thyroid Neoplasms/
2. exp Goiter, Nodular/
3. (thyr?oid* adj5 (cancer* or neoplas* or carcinoma* or malignan* or tumor* or tumour* or adenocarcinoma*)).mp.
4. Thyroid Gland/
5. exp Neoplasms/
6. 4 and 5
7. or/1-3,6
8. "Quality of Life"/
9. (qol or (quality adj2 life)).ab,ti.
10. (value adj2 (money or monetary)).tw.
11. value of life/
12. quality adjusted life year/
13. quality adjusted life.tw.
14. (qaly$ or qald$ or qale$ or qtime$).tw.
15. disability adjusted life.tw.
16. daly$.tw.

17. health status indicators/
18. (sf36 or sf 36 or short form 36 or shortform 36 or sf thirtysix or sf thirty six or shorform thirtysix or shortform thirty six or short form thirtysix or short form thirty six).tw.
19. (sf 6 or sf6 or short form 6 or shortform 6 or sf six or sfsix or shortform six or short form six).tw.
20. (sf12 or sf 12 or short form 12 or shortform 12 or sf twelve or sftwelve or shortform twelve or short form twelve).tw.
21. (sf6D or sf 6D or short form 6D or shortform 6D or sf six D or sfsixD or shortform six D or short form six D).tw.
22. (sf20 or sf 20 or short form 20 or shortform 20 or sf twenty or sftwenty or shortform twenty or short form twenty).tw.
23. (euroqol or euro qol or eq5d or eq 5d).tw.
24. (hql or hqol or h qol or hrqol or hr qol).tw.
25. (hye or hyes).tw.
26. health$ year$ equivalent$.tw.
27. health utilit$.tw.
28. (hui or hui1 or hui2 or hui3).tw.
29. disutilit$.tw.
30. rosser.tw.
31. (quality adj2 wellbeing).tw.
32. qwb.tw.
33. (willingness adj2 pay).tw.
34. standard gamble$.tw.
35. time trade off.tw.
36. time tradeoff.tw.
37. tto.tw.
38. letter.pt.
39. editorial.pt
40. comment.pt.
41. 38 or 39 or 40
42. or/8-37
43. 42 not 41
44. 7 and 43

The inclusion criteria were defined broadly and the sifting process followed an inclusive approach in order to maximise the range of candidate options (Table 62). HRQoL studies in MTC, or other types of thyroid cancer (papillary, follicular, Hürthle cell carcinoma) were included.

Table 62. Assessment group's inclusion criteria for review of published health utility data

Eligibility criteria
Inclusion criteria • Studies reporting preference-based health utilities relating to any type of thyroid cancer **Exclusion criteria** • Studies evaluation diagnostic/staging interventions • Editorials • Reviews • Clinical studies • Letters and commentaries • Non-English language

The Assessment Group carried out two systematic reviews: (1) studies reporting economic evaluations of cabozantinib and/or vandetanib and (2) HRQoL studies. The authors explain that due to the cost-effectiveness search having also identified studies related to health utilities (e.g., those used within the health economic models), the results of both searches were looked at together. In the update search, only the HRQoL search was updated, although there is a possibility that some of the utility studies came through from the economic evaluation search.

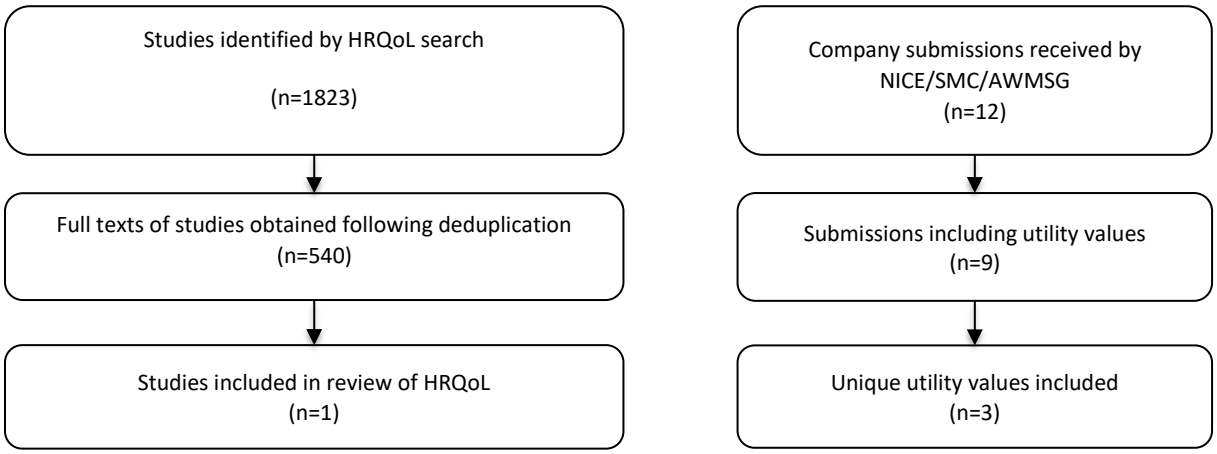
The study selection process for the combined Assessment Group original search plus the update is summarised in Figure 15. In total, there were 1,823 citations identified in the HRQoL search (original plus update). References were collected in a bibliographic management database (EndNote), and duplicates were removed. After titles/abstracts were screened, 1,283 citations were excluded and therefore 540 citations remained. One study was selected as potentially includable in the model (Fordham *et al*.).[379] The selected study reported health utilities for patients with radioactive iodine-refractory differentiated thyroid cancer by heath state (PFS and OS). No MTC specific utilities were identified. Table 63 shows details of the included studies, including the one used for the base case. Fordham *et al*. is a seven health-state vignette study (base state – stable/no response, response to therapy, progressive disease, diarrhoea, fatigue, hand and foot syndrome, alopecia).[379] A total of 100 members of the UK public participated in time-trade-off (TTO) interviews to value the health states.

The Assessment Group also ran scenario analyses that used health state utility values from previous thyroid cancer drug submissions. To identify further utility values to use in scenario analyses, the Assessment Group explored health utility values within previous thyroid cancer submissions to NICE, Scottish Medicines Consortium (SMC) and All Wales Therapeutics and Toxicology Centre (AWMSG). This HTA website search was updated in April 2024. In total, 12 HTA submissions were identified, seven in the original search (two NICE company submissions, two SMC reports and two AWMSG reports) and five in the update search, including two full NICE reports of the company submissions identified in the original search.[377,380-390] Most economic evaluations in the HTA reports were referring to the same study (Fordham

et al.). Only three of the 12 documents contained utility values that were at least partly based on sources other than Fordham *et al*. and were therefore used in model.[380,388,389]

In case study 1 (UC), there were almost 30 utility sources included in the report. Of those around 20% were included in the model. In this case study, all values that were identified were included both in the report and the model. This is due to scarcity of data in thyroid cancer, and especially MTC.

Figure 15. Study selection results UC utility search – Usual practice



Abbreviations: AWMSG: All Wales Medicines Strategy Group, NICE: National Institute for Health and Care Excellence, SMC: Scottish Medicines Consortium.

Table 63. Overview of utility sources considered for and used in model – Usual practice

Study	Report (n=4)	Included in model (n=4)
Fordham *et al*.[379] (base case)	✓	✓
Sanofi company submission[380]	✓	✓
DECISION (sorafenib SMC)[388]	✓	✓
Cabozantinib SMC[389]	✓	✓

Abbreviations: SMC: Scottish Medicines Consortium.

Utility Values Used in the Model

This Section provides the utility values selected for use in the model. The health state utility values from Fordham *et al*. were used as the base case.[379] The health state utility values in this study were associated with the absence/presence of disease progression, were specific to thyroid cancer, and health utilities were derived using the preference-based time to trade-off (TTO) elicitation approach. The alternative utility values based on the current and previous HTA submissions were: the Sanofi company submission to NICE (reported in Tappenden *et al*. 2019[1]), the DECISION trial, and the cabozantinib SMC submission 2015.[372,388,389] The utility values that were applied in the health economic model are summarised in Table 64.

In the Sanofi company model, the utility value for the progression-free health state utility estimate was derived by mapping ZETA trial FACT-G data to EQ-5D. The post-progression utility estimate was derived by using the progression-free health state utility and a multiplier from Beusterien *et al*.[391] The sorefenib SMC submission reported utilities that were derived from EQ-5D data from the DECISION study. The cabozantinib SMC submission included utilities in which SF-36 outcomes had been converted to utilities by mapping to the EQ-5D, and converting these to SF-36 values for the non-progressed and progressed health states.

Table 64. Health state utility values used in the model - Usual practice

	Progression free	Post progression
Base case: Fordham *et al*.[379]	0.80	0.50
Scenario 1: Sanofi company submission[372]	0.84	0.64
Scenario 2: DECISION (sorafenib SMC)[388]	0.80	0.64
Scenario 3: Cabozantinib SMC[389]	0.80	0.62

Abbreviations: SMC: Scottish Medicines Consortium.

Health Economic Model Results

In the base case pairwise comparison of cabozantinib versus BSC, cabozantinib produces more QALYs at higher cost than BSC. The ICER for cabozantinib versus BSC is expected to be £148,169 per QALY gained. When running the model with the alternative utility values, the results are similar in all scenarios. The results indicate that the ICER remains in excess of £140,000 per QALY gained across all scenarios. Table 65 presents these results. In the base case pairwise comparison of vandetanib versus BSC, vandetanib also produces more QALYs at higher cost than BSC. The ICER for vandetanib versus BSC is expected to be £336,896 per QALY gained. The scenario analyses produce significantly higher ICER, up to £1,532,109 per QALY gained. Table 66 presents these results.

Table 65. Health economic model results – Usual practice, pairwise comparison of cabozantinib vs BSC

	ICER (£)
Base case: Fordham *et al*.[379]	148,169
Scenario 1: Sanofi company submission[372]	154,582
Scenario 2: DECISION (sorafenib SMC)[388]	166,890
Scenario 3: Cabozantinib SMC[389]	165,816

Abbreviations: ICER: incremental cost-effectiveness ratio, SMC: Scottish Medicines Consortium.

Table 66. Health economic model results – Usual practice, pairwise comparison of vandetanib vs BSC

	ICER (£)
Base case: Fordham *et al*.[379]	336,896
Scenario 1: Sanofi company submission[372]	822,117
Scenario 2: DECISION (sorafenib SMC)[388]	1,532,109
Scenario 3: Cabozantinib SMC[389]	1,161,487

Abbreviations: ICER: incremental cost-effectiveness ratio, SMC: Scottish Medicines Consortium.

Search Efficiency

The original and update usual practice search took 1,200 minutes (approx. 20 hours). This included the running of the searches, duplicate removal, citation identification, and citation and full-text screening against the inclusion criteria. Table 67 presents the estimated effort that went into searching, using the usual practice search method, for both the published search (covering time from database inception to November 2019) and an update of that search (covering the time period from November 2019 to April 2024). The time spent doing the update search was recorded and rounded to the nearest 5-minute increments. The time for the published original search needed to be estimated, as this was not available

from the publication. The estimation was based on using the times from the update search per record screened. These methods are explained in the Methods Section (4.2.2) and a worked-up example is given in case study 1. As with case study 1, it was not possible to estimate the time for the original search protocol development or running of the searches. No additional time was assumed for search protocol development and running of those original searches. The time for running the search update was recorded, and included in the estimate. It is likely to be an underestimation of the effort, and this represents a limitation of the study.

Table 67. Search tracker - Usual practice

Item	Original and Update Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Original	Update	Total	Original	Update	Total		
Search protocol development						0		
Running searches					120	120		
Identified citations	1282	541	1823					
Duplicate removal	1178	105	1283	590	50	640		
Title/abstract level	104	436	540	80	350	430		
Citations excluded	103	436	539					
Full papers to retrieve + review	1	0	1	10	0	10		
Sources identified through UK HTA web search	7[a]	5	10[b]					
Sources cited in the report	4	0	4				0.22%	456
Sources used in model	4		4				0.22%	456
TOTAL			**4**	**670**	**400**	**1,200**		

Abbreviations: NNR: number needed to read**. Footnotes:** a: includes five SMC and AWMSG HTA reports plus two NICE company submissions. B: Two company submissions in the original search were identified as full NICE reports in the update search, and therefore were considered replacements.

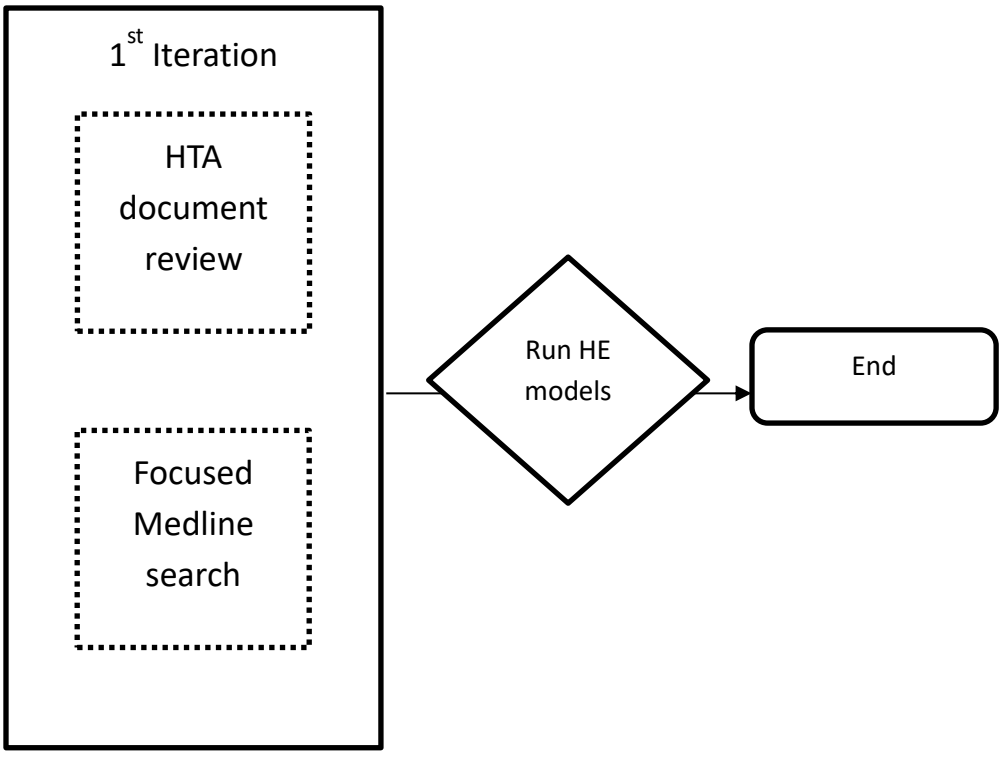
In total, one source was included for consideration in the model from the usual practice bibliographic search, including both the original and update searches. Additionally, the Assessment Group had included three sources from other UK HTA agencies. These sources are listed in Table 63. The update search did not identify further relevant sources in the bibliographic database search, but five more recent NICE reports were identified and reviewed. These reports did not contain any new unique utility values and so no further scenarios were added. The search output had a precision of 0.22% in terms of items cited in the report and 0.22% for items used in the model. The NNR was therefore 456 and 456 for sources cited in the report and sources used in the model, respectively.

6.3.2 Alternative Search Method 1

Iterative Searching

Similar to case study 1, this iterative search aimed to maximise the rate of gain for relevant sources by using techniques associated with berry-picking and information foraging models.[309,310] This first iteration aimed to identify those utility sources that were already identified through literature reviews by others as relevant for use in health economic models, as well as any new information that may have become available since the most recent HTA review. The first iteration therefore included a review of HTA documents that include literature reviews of utility values for health economic models, and a highly precise bibliographic search of the most recent evidence (searched in the titles only). The most recent evidence in the title only search was limited to the period since the last published HTA report. Any subsequent search iteration(s) aimed to understand whether further bibliographic searching would provide additional, relevant evidence for the health economic model. Figure 16 shows the concept of the iterative search. The health economic model was run between each search iteration to assess the marginal impact of identifying additional evidence on the health economic model results, and to assess whether continuing searching is likely to identify further, relevant evidence.

Figure 16. Summary of iterative search



Abbreviations: HE: Health economic

Iteration 1

In the first search iteration, the past NICE submissions that included health economic models were identified and browsed through for any systematic literature reviews of utility values. The NICE Technology Appraisal Guidance (TAG) documents on thyroid cancer treatments were reviewed (TA928, TA742, TA550, TA535, TA516, see Table 68). The documents were reviewed to collect information on which utility studies had been identified in past literature reviews, and subsequently selected for inclusion in health economic models. The TAG documents were also reviewed for any *post hoc* analyses of clinical study data to derive utilities for health economic models. Such *post hoc* analyses are often not published outside the appraisal documents and would not be identifiable through bibliographic database searches.

To ensure that no recent sources were missed since the most recent TAG document, a bibliographic database search was run in Medline (via ProQuest Dialog). No other databases were searched. The most recent NICE submission was cabozantinib (TA928) for previously treated advanced differentiated thyroid cancer (TA856). The literature search for utilities for that appraisal was performed on the 14th October 2021.[387] The literature search identified utility values for all types of thyroid cancer, not only differentiated, and was therefore deemed suitable for this case study in MTC. The committee papers state that the search strategy is given in an appendix to the company submission, but these are not included in the online document. Therefore, the original search strategy cannot be accessed. A simple search strategy (see Table 68) was constructed for retrieving the latest studies for the purpose of this case study. An update (title-focused) search from October 2021 to present (April 2024) was carried out to retrieve any records that may not have been captured by the literature reviews in the TAG documents. Both the non-bibliographic and bibliographic searching was performed in April 2024. References were collected in a bibliographic management database (EndNote).

Table 68. Iterative search: 1st iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit from October 2021 – present (April 2024)
Terms used	ti(thyroid cancer) AND ti(utility)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	NICE Technology Appraisal Guidance in thyroid cancer were reviewed: https://www.nice.org.uk/guidance/conditions-and-diseases/cancer/thyroid-cancer/products?GuidanceProgramme=TA
Source names	Cabozantinib (TA928)[387]: https://www.nice.org.uk/guidance/ta928 Selpercatinib[386]: https://www.nice.org.uk/guidance/ta742 Vandetanib[383]: https://www.nice.org.uk/guidance/ta550 Lenvatinib[384]: https://www.nice.org.uk/guidance/ta535 Cabozantinib (TA516)[385]: https://www.nice.org.uk/guidance/ta516
Search dates	18 - 21 April 2024
Limits	No further limits

The NICE TAG documents were reviewed for systematic literature reviews of utility values, as well as any utility values reported from clinical trials or other sources. The oldest NICE appraisal was TA516 for cabozantinib for treating MTC (March 2018).[385] This submission contained a table with health utility values applied in previous UK thyroid cancer submissions (lenvatinib SMC, sorafenib SMC, cabozantinib SMC, vandetanib AWMSG, cabozantinib AWMSG).[380,381,388-390] These submissions contained utility values from the DECISION trial (sorafenib SMC), published and mapped utility values (cabozantinib SMC), and 'Sanofi model' utility values (vandetanib AWMSG). The utility value for the progression-free health state in Sanofi's model was based on the ZETA trial.[380] The post-progression health state utility value in Sanofi's model was estimated by applying a multiplier (0.766) reported by Beusterien *et al*.[391] This submission also referred to an Assessment Group's literature review, where Fordham *et al*. was identified during a bibliographic database search.[379] In August 2018, a NICE appraisal for lenvatinib and sorafenib for treating differentiated thyroid cancer after radioactive iodine (TA535) was published.[384] In the submission, the company initially used utilities from the DECISION trial, that are specific to treatment and include 'response' as well as disease progression. However, due to differences between the sorafenib and lenvatinib trials, the response health state could not be modelled equally for both treatments and

therefore committee preferred a model structure without a 'response' health state. The preferred source for decision-making was Fordham *et al*.[379] In December 2018, a NICE appraisal was published for vandetanib for treating MTC (TA550).[383] The Fordham *et al*. TTO study was used for decision making, and no other utility values were available from the documents. A NICE recommendation on selpercatinib for treating advanced thyroid cancer with RET alterations (TA742) was published in November 2021.[386] The utility sources reported in this appraisal included Fordham utilities as the base case utility source, and DECISION utilities values from TA535 and the sorafenib SMC submission.[384,388] The latest NICE appraisal from November 2023 was TA928 which assessed cabozantinib for previously treated advanced differentiated thyroid cancer unsuitable for or refractory to radioactive iodine. In this submission, the utility value for progression-free health state was reported from COSMIC-311 study. The company argued that Fordham *et al*. is a more relevant source, due to limited follow-up in the COSMIC-311 study following disease progression, as well as missing values. However, the committee preferred to use COSMIC-311 utility value for the progression-free health state and the utility value from Fordham *et al*. for the progressed disease health state. The progression-free utility value from COSMIC-311 is marked as confidential and therefore cannot be included in this case study as a scenario analysis. The latest literature search in TA928 document identified two sources from bibliographic databases: Kerr *et al*.[392] and Fordham *et al*.[379] Kerr *et al*. is a poster publication whereas Fordham *et al*. is a peer reviewed full manuscript, and therefore Fordham *et al*. is preferred.[379,392]

The second element of the first iteration, the title-focused update literature search of a single database was run on the 18th April 2024, and 17 records were identified and screened. Sixteen citations were excluded at title level. One full-text was checked for potential utility value sources, as it was a cost-utility model publication, but no new, unique studies were found.[393] Therefore, this bibliographic search update did not identify any new sources. A total of five sources were identified (see Table 63), all from the NICE TAG documents.

Utility Values Used in the Model from Iteration 1

Fordham *et al*. was the most frequently cited utility source, and was found in many in of the NICE TAG documents reviewed.[379] Additionally, other utility values were identified that were used in scenario analyses. Table 64 in the previous section reports the sources for the base case and scenarios. All sources identified in the iteration 1 of the iterative search, were also identified in the usual practice search. No additional sources were identified.

Health Economic Model Results

As the utility sources identified in the iterative search were identical to those identified in usual practice, the health economic model results were also identical. These are shown on Table 65 and Table 66 (reported in the previous Section detailing the usual practice search).

Decision to Stop Searching

The five NICE TAG documents contained a limited evidence base in terms of potential utility sources that can be used in health economic modelling of MTC.[383-387] The bibliographic database search that was done as part of this case study to update the most recent literature search in TA928 did not uncover any further

references that contained utility values. It is unlikely that further searching would find additional relevant sources of information. The health economic model results for cabozantinib versus BSC show that the ICER for all utility scenarios is approximately between £148,000 and £167,000 per QALY gained. Therefore, cabozantinib is not a cost-effective treatment option compared to BSC at threshold of £36,000 in any scenario. In the vandetanib versus BSC comparison, the ICER varied between approximately £337,000 and £1,500,000, also demonstrating cost-ineffectiveness at a threshold of £36,000 per QALY gained. Documents are judged to have a high marginal relevance if they are both relevant to the information requirement, and they contain minimal similarity to previously selected sources. In this case, even if further sources with relevant information were to be found, it is unlikely that the model conclusion would change i.e. the marginal relevance of any new information would be low. Due to the limited evidence base and a low likelihood of finding further evidence that would change the conclusion from the health economic model, a decision was made to stop searching after iteration 1.

Summary of Iterative Search Results and Efficiency

In total, five sources were included for consideration in the model from iterative searching, and four were actually included in the model analyses (Figure 17). These publications were listed in Table 63. All the publications used in the model from iterative searching were also identified during the usual practice search. The iterative search output had a precision of 29% in terms of items cited in the report and 24% for items used in the model. The NNR was therefore 3 and 4 for sources cited in the report and sources used in the model, respectively.

Figure 17. Summary of results – iterative searching thyroid cancer

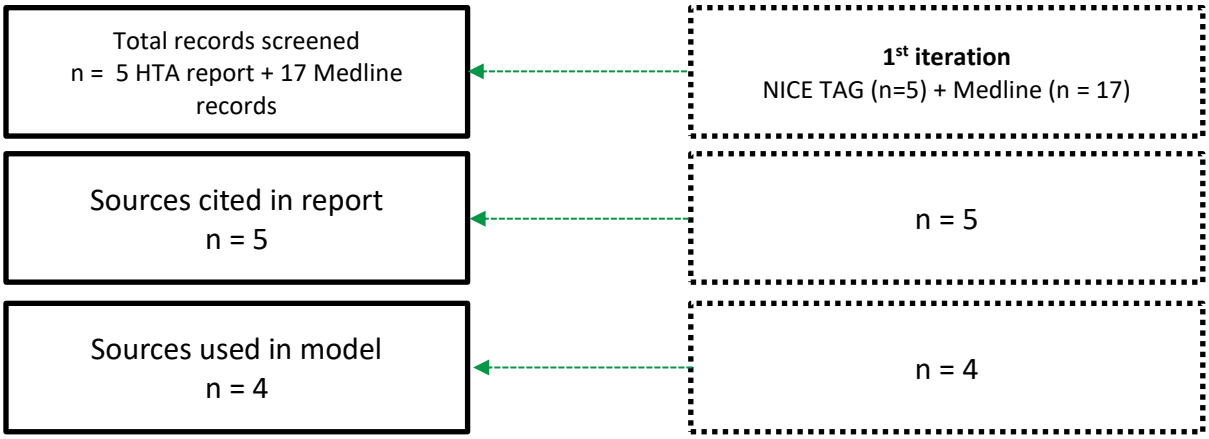


Table 69 displays the estimated search effort. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening took 130 minutes (2 hours and 10 minutes). This time also included running the web searches specified in Table 68.

Table 69. Search tracker – Iterative search

Item	Iterative search			
	Number of studies	Time in minutes	Precision	NNR
Medline search protocol development + identification of HTA websites		70		
Running searches, including HTA report identification		40		
Identified citation Medline	17			
Duplicate removal	0	0		
Citations at title/abstract level	17	10		
Citations excluded	16			
Full papers to retrieve + review	1	10		
Sources cited in the report	5		29%	3
Sources used in model	4		24%	4
TOTAL	**4**	**130**		

Abbreviations: NNR: number needed to read

6.3.3 Alternative Search Method 2

Rapid Review

The second experiment in the thyroid cancer case study is using rapid review methods. The key difference to the first experiment (iterative searching) is that in iterative searching, 'patches' of information are evaluated one at a time and further searching is initiated as judged relevant. In rapid review, the method is similar to systematic literature review in that the search is run only once using a pre-defined search protocol. However, limits are added to manage the scope of the literature review.

For this rapid review, Medline (via ProQuest Dialog) was searched from inception. The search strategy combined free-text and MeSH terms relating to thyroid cancer with terms for utility. An overview of the search strategy is shown in Table 70, with further details of the specific search terms shown in Table 71. The rapid review limitations applied included using only one database (Medline), restricting synonyms, and limiting publication dates to the last ten years (April 2014 – April 2024). Literature searches were conducted in April 2024. References were collected in a bibliographic management database (EndNote). The search results are summarised in Figure 18. The utility sources identified during the searches are summarised in Table 72.

Table 70. Rapid Review search framework

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms, limited to the last 10 years only.
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Last 10 years April 2014 – April 2024
Terms used	Quality of life AND thyroid cancer
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Table 71. Medline search strategy – Rapid Review

Search Strategy
1. (MESH.EXACT("Quality of Life")) OR (MESH.EXACT("Quality-Adjusted Life Years")) OR ti,ab(qaly) OR ti,ab(quality AND ti,ab(life OR wellbeing)) OR ti,ab(health gain) OR ti,ab(disutility) OR ti,ab(utility) OR ti,ab(utilities)
2. MESH.EXACT("Thyroid Neoplasms") OR ti(thyroid cancer)
1 AND 2

Figure 18. Study selection results - Rapid Review

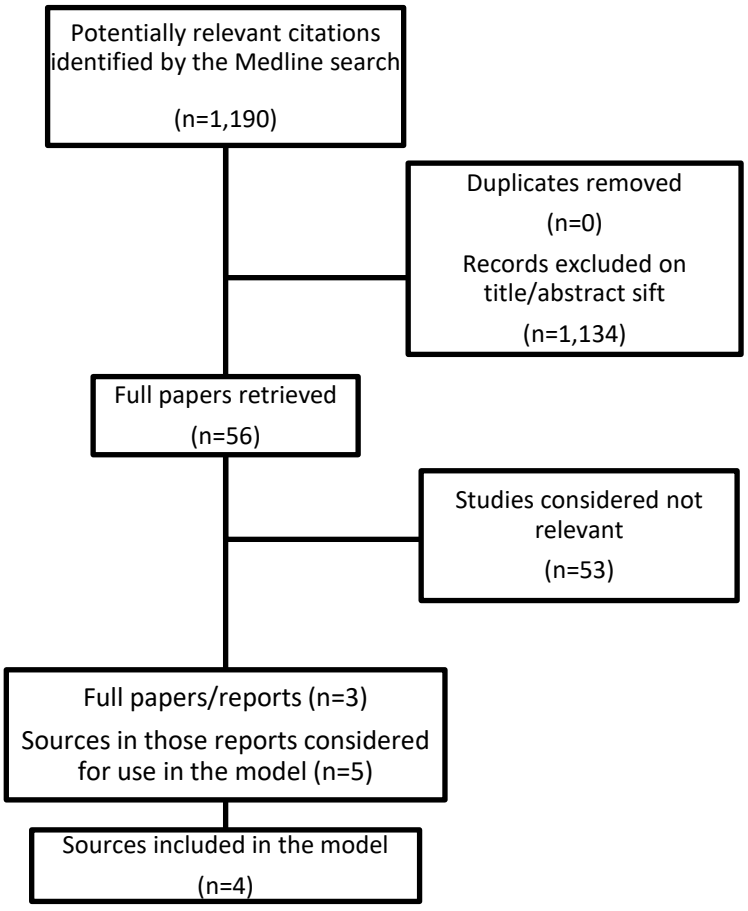


Table 72. UC case study: Studies included – Rapid Review

Study	Usual Practice		Iterative search		Rapid Review	
	Report (n=5)	Included in model (n=4)	Report (n=5)	Included in model (n=4)	Report (n=5)	Included in model (n=4)
Fordham *et al*.[379]	✓	✓	✓	✓	✓	✓
Kerr *et al*.[392]	✓		✓		✓	
Sanofi submission[372]	✓	✓	✓	✓	✓	✓
DECISION (sorafenib SMC)[388]	✓	✓	✓	✓	✓	✓
Cabozantinib SMC[389]	✓	✓	✓	✓	✓	✓

Abbreviations: SMC: Scottish Medicines Consortium.

Utility Values Used in the Model from Rapid Review

Table 64 reports the utility values for the base case and scenarios. All sources identified in the rapid review were also identified in the usual practice and iterative searches. No additional sources were identified.

Health Economic Model Results

As the utility sources identified in the rapid review were identical to those identified in usual practice, the health economic model results were also identical. These are shown in Table 65 and Table 66 (reported in the previous Section detailing the usual practice search).

Summary of Rapid Review Results and Efficiency

Five full texts were included for consideration in the model, and four were included in the model. Table 73 displays the number of sources found by the search and the estimated search effort. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening took 1,400 minutes (23 hours and 20 minutes). The search output had a precision of 0.42% in terms of items cited in the report and 0.34% for items used in the model. The NNR was therefore 238 and 298 for sources cited in the report and sources used in the model, respectively.

Table 73. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		40		
Running searches		15		
Identified citations	1190			
Duplicate removal	0	0		
Citations to sift at title/abstract level	1190	950		
Citations excluded	1134			
Full papers to retrieve + review	56	390		
Full papers for consideration	5			
Sources cited in the report	5		0.42%	238
Sources used in the model	**4**		0.34%	298
TOTAL	4	**1,400**		

Abbreviations: NNR: number needed to read

6.3.4 Case Study Utility Values in Thyroid Cancer – Summary

The key observations from the thyroid cancer utility case study were:

- All search methods found the same evidence base used in the model.
- Iterative searching was the most efficient approach in finding these utility values, representing an 89% reduction in search effort compared to usual practice (Table 74). Usual practice and rapid review required similar time effort at 1,200 and 1,400 minutes, respectively. Rapid review took 17% longer than usual practice. Due to methodological limitations associated with this study (i.e., original search protocol development time and search running time are not known), time for usual practice may be underestimated.
- Model results were identical, as all searches found a similar set of evidence.
- In conclusion, the iterative search method identified all the relevant utility values and it was also the most efficient search method in this thyroid cancer case study.

Table 74. Thyroid cancer case study: Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search	4	4	0.22%	0.22%	456	456	1,200	-
Iterative search	5	4	29.41%	23.53%	3	4	130	-89%
Rapid review	5	4	0.42%	0.34%	238	298	1,400	17%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

6.4 Baseline Risk of Clinical Events

6.4.1 Description of the Usual Practice (Control)

The thyroid cancer model includes parametric survival models fitted to time-to-event data on overall survival (OS) and progression-free survival (PFS) as baseline risks for BSC (Table 75. The baseline risk in this case study is defined as PFS and OS for the comparator group i.e., best supportive care (BSC).

Table 75. Overview of baseline risk of clinical event parameters in the thyroid cancer model

Model Input	Search method	Search dates	Search strategy
OS	Clinical efficacy systematic review	Inception to November 2016	See Table 39
PFS	Clinical efficacy systematic review	Inception to November 2016	See Table 39

Footnote: OS: overall survival, PFS: progression-free survival.

Identification – Overall Survival and Progression-free Survival Inputs

The Assessment Group's systematic review of clinical effectiveness evidence included OS and PFS for cabozantinib and vandetanib, but not for BSC unless it was a comparator arm in a cabozantinib/vandetanib study. The scope of the Assessment Group's systematic review allows estimation of the relative treatment effects of cabozantinib and vandetanib versus BSC in RCTs. No separate model input identification steps were recorded for the identification of the baseline risk of OS and PFS in patients not treated with cabozantinib or vandetanib, or outside RCTs. The health economic model requires the natural history of MTC to be modelled in terms of OS and PFS over a patient's lifetime. Therefore, the Assessment Group's systematic literature review for clinical effectiveness only partially covers this model input identification need.

As part of the Assessment Group analysis, a systematic search was undertaken to identify studies reporting OS and PFS in MTC patients receiving cabozantinib or vandetanib. Several electronic databases, including Medline, EMBASE, CINAHL, and others, were searched from inception to November 2016. The search was updated in April 2024 as part of this project. Table 76 provides the Medline search strategy. Key inclusion and exclusion criteria are given in Table 77. The full details on the original search can be found in Tappenden *et al*. 2019, Chapter 3 and Appendix 1.

Table 76. Medline search strategy for baseline risk values (OS, PFS) - Usual practice (source: Tappenden *et al*.[372])

Search Strategy
1 exp Thyroid Neoplasms/
2 exp Goiter, Nodular/
3 (thyr?oid* adj5 (cancer* or neoplas* or carcinoma* or malignan* or tumor* or tumour* or adenocarcinoma*)).mp.
4 Thyroid Gland/

5 exp Neoplasms/
6 4 and 5
7 or/1-3,6
8 exp Carcinoma, medullary/
9 (medullary or MTC).mp.
10 8 or 9
11 7 and 10
12 Randomized controlled trials as Topic/
13 Randomized controlled trial/
14 Random allocation/
15 randomized controlled trial.pt.
16 Double blind method/
17 Single blind method/
18 Clinical trial/
19 exp Clinical Trials as Topic/
20 controlled clinical trial.pt.
21 clinical trial$.pt.
22 multicenter study.pt.
23 or/12-22
24 (clinic$ adj25 trial$).ti,ab.
25 ((singl$ or doubl$ or treb$ or tripl$) adj (blind$ or mask$)).tw.
26 Placebos/
27 Placebo$.tw.
28 (allocated adj2 random).tw.
29 or/24-28
30 23 or 29
31 Case report.tw.
32 Letter/
33 Historical article/
34 31 or 32 or 33
35 exp Animals/
36 Humans/
37 35 not (35 and 36)
38 34 or 37
39 30 not 38
40 meta-analysis/
41 meta-analysis as topic/
42 (meta analy* or metanaly* or metaanaly*).ti,ab.
43 ((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
44 (reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.

45 (search strategy or search criteria or systematic search or study selection or data extraction).ab.
46 (search* adj4 literature).ab.
47 (medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or scie nce citation
index or bids or cancerlit).ab.
48 cochrane.jw.
49 ((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
50 or/40-49
51 39 or 50
52 11 and 51

Table 77. Key inclusion/exclusion criteria for baseline risk of OS and PFS in MTC (source: Tappenden *et al*.[372])

Eligibility criteria
Inclusion criteria • **Population:** Participants with unresectable locally advanced or metastatic MTC, aged ≥ 18 years. Studies with populations broader than unresectable locally advanced or metastatic MTC was considered only if data for the relevant study population are available and are reported separately • **Interventions:** Cabozantinib, vandetanib • **Comparators:** Interventions were compared with each other and against BSC • **Outcomes:** Overall survival, progression-free survival, response rates, adverse effects of treatment, health-related quality of life • **Study design:** Randomised controlled trial **Exclusion criteria** • Paediatric population • Pre-clinical studies, animal models, narrative reviews, systematic reviews, clinical guidelines, editorials, letters, opinion pieces and abstracts.

The study selection process including the original Assessment Group search plus the update search is summarised in Figure 19. In total, there were 3,641 citations identified in the systematic search (original plus update). References were collected in a bibliographic management database (EndNote), and duplicates were removed. After duplicates were removed, 2,958 citations remained. The original systematic review identified four publications that were included in the report as potential sources to be considered for the model. These sources related to two studies: the ZETA study which compared vandetanib against placebo and the EXAM study which compared cabozantinib against placebo.[376,394-396] The pivotal publications of the Phase III studies were: Wells *et al*. 2012 (ZETA study) and Elisei *et al*. (EXAM

study).[376,394] Two additional sources were identified from clinicaltrials.gov.[395,396] A later EXAM publication reported final OS and PFS data from the EXAM study, and these final data were used in the health economic model to model cabozantinib versus BSC.[397] The ZETA trial publication by Wells *et al*. did not include data for the vandetanib EU label population necessary for the model, and therefore the clinicaltrials.gov source and an additional unpublished clinical study report source, which the Assessment Group had access to, were used for modelling of the EU population PFS and OS.[377,395] Table 78 lists the sources identified during the usual practice search, and Table 79 provides details of those included in the report versus those included in the model.

Figure 19. Study selection results thyroid cancer baseline risk of clinical event – Usual practice

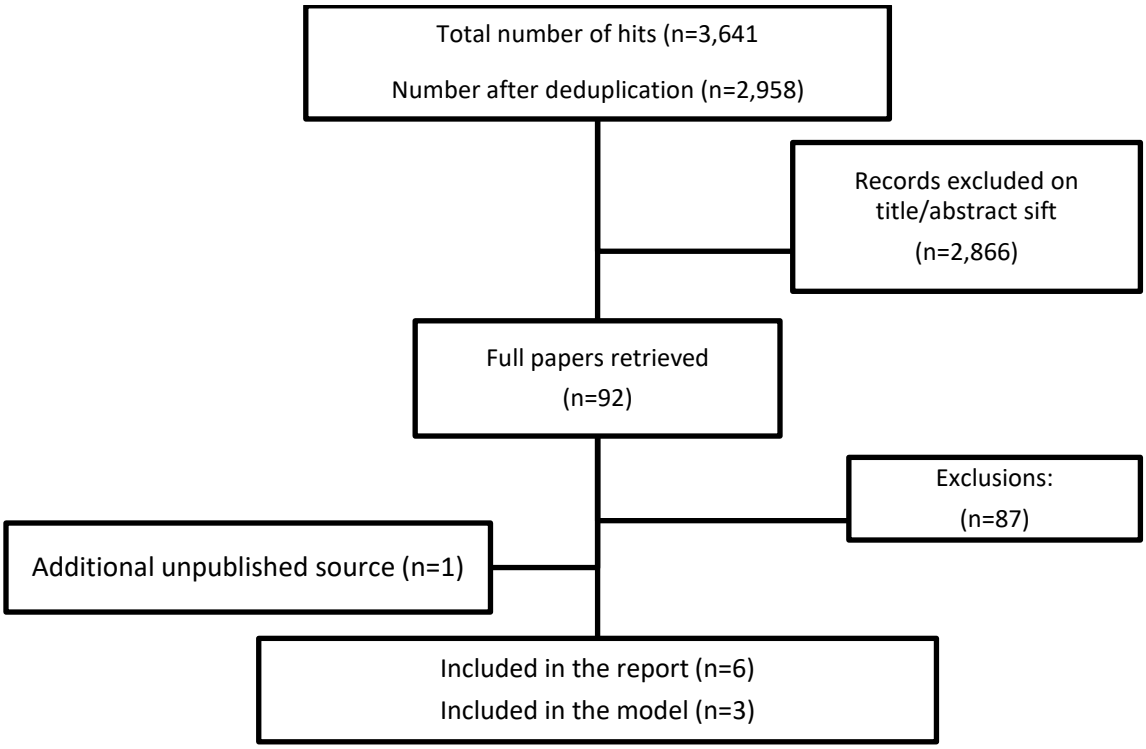


Table 78. Sources identified in the usual practice search – MTC baseline risk of clinical events

Search Strategy
ZETA: Wells SA, Robinson BG, Gagel RF, Dralle H, Fagin JA, Santoro M, *et al*. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. J Clin Oncol 2012;30:134–41. https://doi.org/10.1200/JCO.2011.35.5040[394]
ZETA: ClinicalTrials.gov. NCT00410761. An Efficacy Study Comparing ZD6474 to Placebo in Medullary Thyroid Cancer. 2006. URL: https://clinicaltrials.gov/ct2/show/NCT00410761 (accessed 25 April 2024).[395]
ZETA: Sanofi Genzyme. Vandetanib for Treating Unresectable Locally Advanced or Metastatic Medullary Thyroid Cancer (ID56): Evidence Submission to NICE. Unpublished. Oxford: Genzyme Therapeutics; 2017.[377]
EXAM: Elisei R, Schlumberger MJ, Müller SP, Schöffski P, Brose MS, Shah MH, *et al*. Cabozantinib in progressive medullary thyroid cancer. J Clin Oncol 2013;31:3639–46. https://doi.org/10.1200/JCO.2012.48.4659[376]
EXAM: ClinicalTrials.gov. NCT00704730. Efficacy of XL184 (Cabozantinib) in Advanced Medullary Thyroid Cancer. 2008. URL: https://clinicaltrials.gov/ct2/show/NCT00704730 (accessed25 April 2024).[396]
EXAM follow-up: Schlumberger M, Elisei R, Müller S, ***et al***. Overall survival analysis of EXAM, a phase III trial of cabozantinib in patients with radiographically progressive medullary thyroid carcinoma. Ann Oncol 2017; 28(11): 2813-9.[397]

Table 79. Overview of baseline risk of clinical event sources – Usual practice

Study	Report (n=6)	Included in model (n=2)
ZETA study[394]	✓	
ZETA clinicaltrials.gov[395]	✓	
ZETA unpublished EU label data[377]	✓	✓
EXAM study[376]	✓	
EXAM follow-up, final data cut[397]	✓	✓
EXAM clinicaltrials.gov[396]	✓	

Baseline Risk – Qualitative Assessment of Data Included

The median PFS in the EXAM trial was 11.2 months in the cabozantinib arm, compared to 4.0 in the BSC arm. The HR was 0.28; 95% confidence interval 0.19 – 0.40; p<0.001. The median PFS in the ZETA trial, in the EU label for vandetanib population was 28.0 months in the vandetanib arms, compared to 16.4

months in the BSC arm. The HR was 0.47, 95% confidence interval was 0.29 – 0.77, *p*=0.0024. In the EXAM study, the primary analysis of OS showed no difference between treatment arms (HR, 0.89; 95% CI, 0.63 – 1.52), when only 96 of the planned 217 deaths had occurred.[397] The final Kaplan-Meier analysis showed a 5.5-month increase in median OS with cabozantinib versus BSC (HR, 0.85; 95% confidence interval 0.64 – 1.12; *p*=0.24). The EU-label population median OS from ZETA trial is confidential. The HR was 0.89; 95% confidence interval 0.48 – 1.65; *p*-value not reported. In the base case, the ICER for cabozantinib versus BSC is expected to be £148,169 per QALY gained, and the ICER for vandetanib versus BSC is expected to be £336,896 per QALY gained. The Assessment Group report did not include scenario analyses with alternative data sources for the baseline risk of OS or PFS. The base case results are presented below in Table 80. The PFS and OS results are presented in Table 81. HRs are provided for information only.

Table 80. Deterministic health economic model results – Usual practice

Pairwise analysis	ICER (£)
Vandetanib vs BSC (ZETA)	336,896
Cabozantinib vs BSC (EXAM)	148,169

Abbreviations: ICER: incremental cost-effectiveness ratio.

Table 81. HRs and median OS and PFS from EXAM and ZETA studies

	PFS (months) - median	PFS HR; 95% CI; p-value	OS (months) - median	OS HR; 95% CI; p-value
ZETA – EU label[377]	28.0 vs. 16.4	0.47; 0.29 – 0.77; 0.0024	Confidential	Confidential
EXAM[397]	11.2 vs. 4.0	0.28; 0.19 – 0.40; <0.001	26.6 vs 21.1	0.85; 0.64 – 1.12; 0.241

Abbreviations: PFS: Progression free survival; HR: hazard ratio; CI; confidence interval; OS: overall survival.

Search Efficiency

Table 82 displays the estimated effort to identify OS and PFS. In the systematic review 3,641 studies were identified, out of which two were used in the model. This was estimated to take 60 hours and 10 minutes. The search output had a precision of 0.16% in terms of items cited in the report and 0.05% for items used in the model. The NNR was therefore 607 and 1,821 for sources cited in the report and sources used in the model, respectively. The usual practice searching was limited to RCTs with cabozantinib or vandetanib.

Table 82. Search tracker - Usual practice

Item	Original Search							
	Number of studies identified			Time in minutes			Precision	Number needed to read
	Original	Update	Total	Original	Update	Total		
Search protocol development						0		
Running searches					258	260		
Identified citation	2189	1452	3641					
Duplicate removal	608	75	683	300	40	340		
Citations to sift at title/abstract level	1581	1377	2958	1260	1100	2360		
Citations excluded	1516	1350	2866					
Full papers to retrieve + review	65	27	92	460	190	650		
Sources cited in the report	6	0	6				0.16%	607
Sources used in model	2	0	2				0.05%	1821
TOTAL			**2**	**1560**	**1140**	**3610**		

Abbreviations: NNR: number needed to read

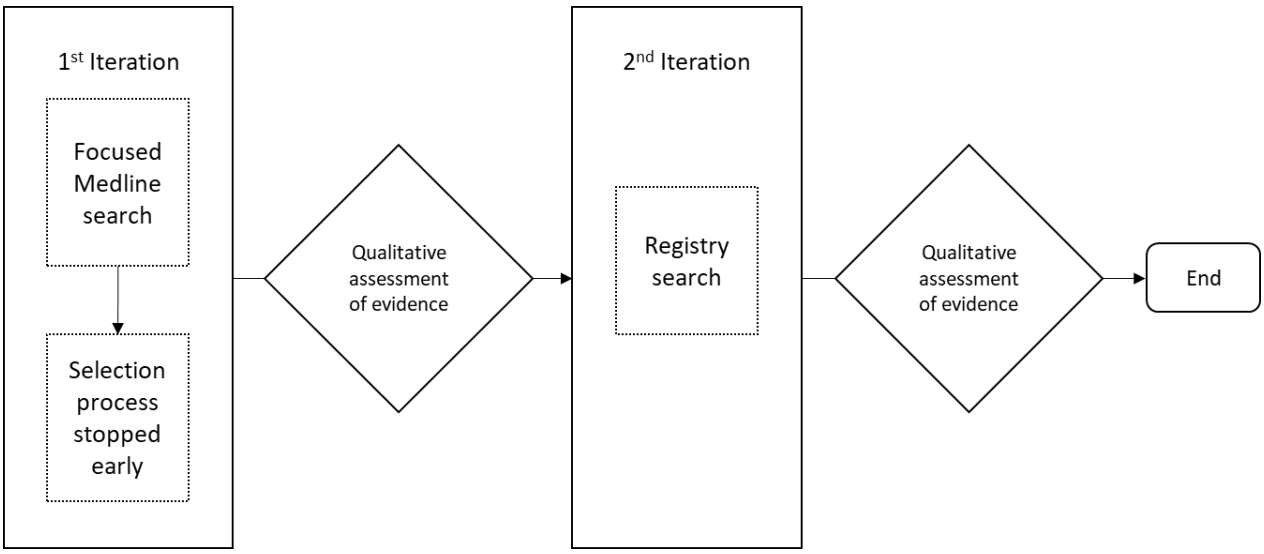
6.4.2 Alternative Search Method 1

Iterative Searching

The first iteration included a precise bibliographic search in titles only without treatment or study type limitations, with the aim of identifying key evidence for baseline risk of OS and PFS. The second iteration included a registry search only. This first iteration search was aimed at capturing all types of studies (not only RCTs). The biggest challenge observed in iteration 1 was that none of the publications matched the target population and/or included data in the correct format. Several of the "close-but-not-quite" publications were analyses from registries, and it was observed that at least in theory the registry data had potential to be analysed in a manner that would allow for it to be included in the model. Therefore, it was decided that iteration 2 should be a search for registries.

Figure 20 shows the concept of the iterative search. For this model input in this case study, the health economic model was not run between the iterations, as specified in the introduction to the case study. Instead, any identified data were assessed qualitatively.

Figure 20. Summary of alternative search method 1: Iterative search (thyroid cancer case study)



Iteration 1

In the first search iteration, a focused bibliographic database search was run. Details of these searches are summarised in Table 83. The bibliographic searching was done in May 2024. References were collected in a bibliographic management database (EndNote). A total of 131 studies were identified. Of these, 70 citations went through the selection process and 25 were selected for full text review. The title/abstract review process was stopped early, because there was uncertainty on whether the selected publications would contain time-to-event data for OS and PFS for the target population (advanced or metastatic MTC). Several of the studies were comparing survival in specific population, such as with/without amyloid

deposition.[398] It was unclear from the title/abstract whether the overall population survival was also reported and whether Kaplan-Meier estimates were included in the publication. Instead of going through the whole patch, including the older studies which seemed less relevant, a decision was made to review early the full texts of the already selected titles/abstracts before proceeding.

Stopping title/abstract review early is not usual practice, and the impact of doing so is currently unknown. Therefore, the remaining titles/abstracts were reviewed for verification purposes. No relevant publications were found during this verification review.

Table 83. Iterative search: 1st iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title/abstract only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	No limitation
Terms used	To search for MTC related sources: ti(medullary thyroid cancer) OR ti(medullary thyroid carcinoma) AND To search for PFS and OS: ti(survival) OR ti(mortality)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Twenty-five full texts were reviewed. Of these, two publications were included in the report. The first publication that was included was by Liu *et al*. 2024, where OS was reported for those locally advanced or metastatic patients who had received primary tumour resection versus those who had not.[399] In the EXAM study the majority of patients had received thyroidectomy; 92% and 94% in cabozantinib and placebo arms, respectively. The proportion was not reported for the ZETA study. The OS probabilities from Liu *et al*. of the no resection population were visually compared with BSC OS probabilities reported in Schlumberger *et al*. (see Figure 21 and Figure 22). The median OS in EXAM study is under 21.1 months in

the BSC arm, when in Liu *et al*. it seems to be under 20 months from visual inspection. Further examination, ideally with a clinical expert, would be needed to more closely understand the differences between the study populations. In the longer term (e.g., at 72 months) a higher proportion of patients are alive in the EXAM study compared to Liu *et al* publication (over 20% versus under 20%, estimated by visual inspection).Therefore, this publication was not included in the model.

The second publication that was included in the report was by Schlumberger *et al*. 2017.[397] This publication was also identified in the usual practice search, and it contained the final data cut for both cabozantinib and placebo arms of the EXAM study. A long-term vandetanib follow-up publication by Ramos *et al*. 2019 was also reviewed but excluded as it only contained survival data for vandetanib, and none for placebo/BSC.[400] Eight of the studies (including Ramos *et al*.) were excluded as they were specific to a population that had received a specific treatment (e.g., a type of surgery, vandetanib, etc.) or did not contain the correct type of data (e.g., only HRs).[400-407] A further 15 studies reported subpopulation results.[398,400,408-420] The subpopulations included presence/absence of amyloid depositions, only sporadic MTC, only hereditary MTC, by lymph node ratio, by calcitonin or C19.9 levels, or by age group. One study reported earlier results for the EXAM study that were superseded by Schlumberger *et al*. 2017, and so this was also therefore excluded.[421]

The PFS data sources identified are summarised in Table 84, and the OS data sources in Table 85. These two tables also contain comparison to sources identified during usual practice search.

Figure 21. Overall survival reported in Liu *et al*. 2024 (in Figure 2A)

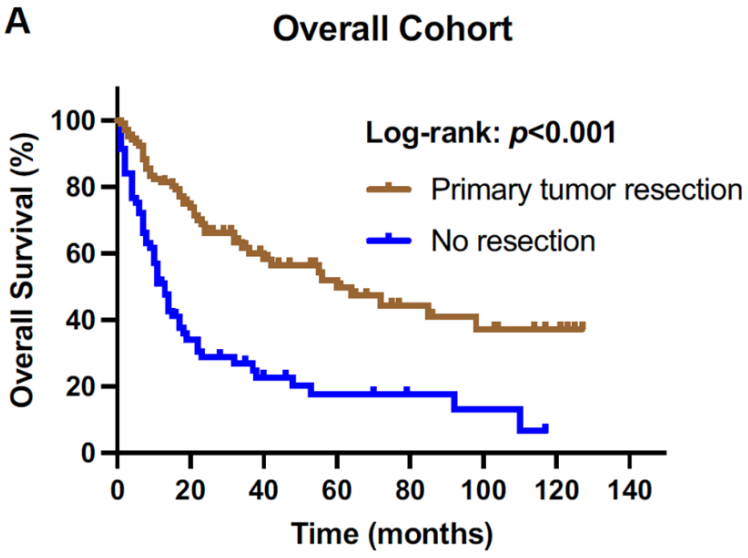


Figure 22. Overall survival reported in Schlumberger *et al*. 2017 (in Figure 1A)

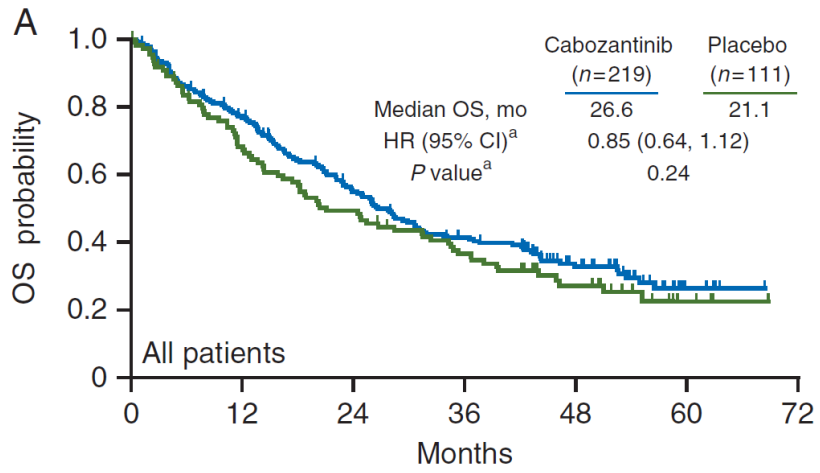


Table 84. Overview of sources considered for PFS

Study	Usual Practice (n=6)	Iteration 1 (n=1)
ZETA study[394]	✓	
ZETA clinicaltrials.gov[395]	✓	
ZETA unpublished EU label data[377]	✓	
EXAM study[376]	✓	
EXAM follow-up, final data cut[397]	✓	✓
EXAM clinicaltrials.gov[396]	✓	

Table 85. Overview of sources considered for OS

Study	Usual Practice (n=6)	Iteration 1 (n=2)
ZETA study[394]	✓	
ZETA clinicaltrials.gov[395]	✓	
ZETA unpublished EU label data[377]	✓	
EXAM study[376]	✓	
EXAM follow-up, final data cut[397]	✓	✓
EXAM clinicaltrials.gov[396]	✓	
Liu *et al*. 2024[399]		✓

Baseline Risk Values Used in the Model from Iteration 1

In iteration 1, several publications were selected and reviewed as full texts, but only one publication was finally included to be used in the model. This was a publication reporting the final data cut from the EXAM study.[397] This publication was also identified by usual practice search, as presented in Table 45 and Table 46. None of the other sources from usual practice were identified in this iteration. The usual practice search was not aimed at identifying baseline risks of clinical events, but rather to capture relative treatment effects of cabozantinib and vandetanib versus BSC in RCTs. The usual practice search therefore captured the pivotal Phase III trials, as well as entries to a clinical trial registry (clinicaltrials.gov). It also contained confidential, unpublished data from ZETA study that would not be possible to find in publicly available sources.[377] This first iteration focused search was aimed at capturing all types of studies (not only RCTs) that report PFS or OS in MTC. Although several potential publications were reviewed in iteration 1 of the iterative search, none were suitable for the locally advanced or metastatic population and included time-to-event data for OS or PFS for BSC, except Liu *et al*. 2024 which is discussed above.[399] Several of these publications were analyses from registries, and therefore it was decided that iteration 2 should be a search for registries.

Iteration 2

The initial iteration did not uncover any relevant baseline risk data beyond those identified in the usual practice clinical effectiveness systematic review. However, some of the excluded publications included data from registries. Although these publications did not present the necessary breakdown/population/etc. for the health economic model, the registries could still serve as potential sources for estimating the OS and PFS for BSC if the analyses were defined differently. The registries in these publications identified in iteration 1 were reviewed and listed. A further targeted web search was performed to identify registries that may contain useful information. The search details are provided in Table 86, the identified registries that contain data for MTC patients are presented in Table 87. Table 88 lists identified registries for broader disease (e.g., thyroid cancer, endocrine diseases, cancer, national rare disease registries) that may contain MTC data. It was not possible to confirm whether these registries have MTC specific data due to language limitations, as well as data not being available without contact with the registry. Similarly, no input values could be extracted, as access to registries typically requires subscription and/or access fees. Therefore, for registry search potential sources are listed but not tested in the model as the databases could not be accessed. The search time for iteration 2 was 230 minutes.

Table 86. Registry search

Bibliographic database search elements	Considerations
Sampling	None
Type of studies	None
Sources	None
Limits	None
Terms used	None
Conceptual limitations	None
Non-database search elements	**Considerations**
Approaches	Targeted searches were used to identify publications/websites for registries. National and international European registries were searched.
Source names	https://clinicaltrials.gov/ https://www.orpha.net/ https://scholar.google.com/ https://www.google.com/
Search dates	May 2024
Limits	Not applicable

Table 87. Overview of registries identified for MTC

Registry name	Location	Year established	Current status	Sample size (MTC)	Reference
SEER	US	1973	Active	3,833[a]	SEER[422]
National Cancer Database	US	1988	Active	2,776[b]	NCDB[423]
MTC Registry Consortium	US	2010	Active	Not reported	MTC Registry Consortium[424]
The Danish Thyroid Cancer Database	Denmark	1996	Active	476[c]	DATHYRCA[425]
GPOH-MET Register	Germany	2022	Closed	N/A Paediatric	GPOH-MET[426]
RARECAREnet	Europe[d]	1978	Closed	2,223[e]	RARECAREnet[427]
Korea NHIS database	South Korea	2002	Active	1,790[f]	Korea NHIS[428]

Abbreviations: SEER: Surveillance, epidemiology and end results program **Footnote**: a Reported for 2000 – 2019 by Tao *et al*. 2024.[408] b Reported for 2004 – 2014 by Al-Qurayshi *et al*. 2018.[402]c Reported for 1960 – 2014 by Mathiesen *et al*. 2019.[401] d Austria, Belgium, Bulgaria, Croatia, Czech Republic, Estonia, Finland, France, Germany, Iceland, Ireland, Italy, Latvia, Lithuania, Malta, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Switzerland, The Netherlands and United Kingdom. e Reported for 2000 – 2007 by Locati *et al*. 2020.[428] f Reported for 2004 – 2016 by Ahn *et al*. 2020.[429]

Table 88. Overview of registries identified for broader diseases

Registry name	Location	Year established	Current status	Reference
Central Registry Rare Diseases	Belgium	2012	Active	CRRD[428]
EuRRECa - HUDERF	Belgium	2018	Active	HUDERF[428]
French National Registry for Rare Diseases	France	2017	Unknown	BNDMR[430]
Registry for Rare Thyroid and Parathyroid Carcinoma	Germany	2013	Active	Seltene Tumoren der Schilddrüse[431]
Italian National Rare Diseases Registry	Italy	2001	Active	RNMR[432]
Rare Diseases Patient Registry	Spain	2008	Active	RePER[433]
Swiss Rare Diseases Registry	Switzerland	2017	Active	SRSK[434]
Coordination of Rare Diseases at Sanford University	US	2010	Active	CORDS Registry[435]
EuRRECa – European Registries for Rare endocrine conditions	UK/EU	2018	Active	ERN: Endo-ERN[436]

Abbreviations: All abbreviations in the references section are spelled out in respective references.

Baseline Risk Values Used in the Model from Iteration 2

Several registries were identified that either contained MTC specific data (Table 87) or had potential to contain such data (Table 88). These registries have potential to hold relevant data that would need to be accessed and analysed for use in the health economic model.

Decision to Stop Searching

None of the publications identified in iteration 1 reported analyses of OS/PFS that were suitable for use in the health economic model, apart from the Phase III study EXAM that had already been identified in the systematic review of clinical effectiveness. The likelihood of finding relevant evidence was deemed low and therefore a registry search was initiated in iteration 2. Several registries were identified. The next logical step would be to analyse registry data for the target population. However, this falls outside the scope of this thesis. Consequently, the search was stopped, as it is unlikely that suitable data, already analysed for the target population in a format that is usable in a health economic model, can be found in a publication.

Summary of Iterative Search Results and Efficiency

In total, 70 studies were reviewed, two were included in the report and one was included in the model from iteration 1 of iterative searching. In iteration 2, 16 registries were identified that have potential to include relevant data. A summary of the iterative searching is presented in Figure 23.

With this baseline risk of clinical event search in this case study, there are two possibilities how to define usual practice. One is to consider the systematic review of clinical effectiveness as the 'usual practice', as has been done here. A systematic review of clinical effectiveness should at a minimum capture any placebo/BSC data in the target population from pivotal Phase III studies that can be used to model OS/PFS. However, it can also be argued that the clinical effectiveness review was designed for another purpose (to identify relative treatment effects from RCTs), and searching for baseline risk should be an additional activity that had not been performed for this health economic model in the usual practice search. Therefore, another potential definition for usual practice is 'no search'. In this case study, I decided to update and consider the systematic review as 'usual practice'. This was for several reasons. Firstly, the pivotal study data is in fact used to populate the baseline risk of OS and PFS in the model, even if no further sources have been searched additionally (as could be done in a separate, dedicated search). Secondly, using this approach also allowed a comparison of two search practices (usual practice and alternative search methods), in terms of sources identified and time taken. Without a recorded search, it cannot be estimated how long it took to conduct. Finally, other case studies (case study 1 and 3) include usual practice searches where no search steps were recorded for some or all of the baseline risk of clinical event values. Therefore there are other examples in this PhD study where usual practice is 'no search'. I also included reflections on the fact that this search is a conventional relative treatment effects search, that is likely to increase resources needed. Searching for evidence beyond Phase III clinical studies has potential to improve modelling of OS and PFS, resulting in more realistic long-term estimations of OS and PFS, when compared to solely relying on company Phase III data that tends to be available for shorter time only. Even if no additional data are identified, there is also potential to improve transparency of the model by being explicit about which (if any) additional searches were performed to check if such data exists in literature/registries.

The iterative search output had a precision of 1.53% in terms of items cited in the report and 0.76% for items used in the model. The NNR was therefore 66 for the report and 131 for the model. These numbers do not include the registries identified, as the data could not be accessed. Table 89 displays the estimated search effort, separately for iteration 1 and 2, as well as the total. The development of the search protocol, running of the searches, citation identification, and publication screening, as well as the web searches for registries in the two iterations took 500 minutes (8 hours and 20 minutes).

Figure 23. Summary of results – iterative searching

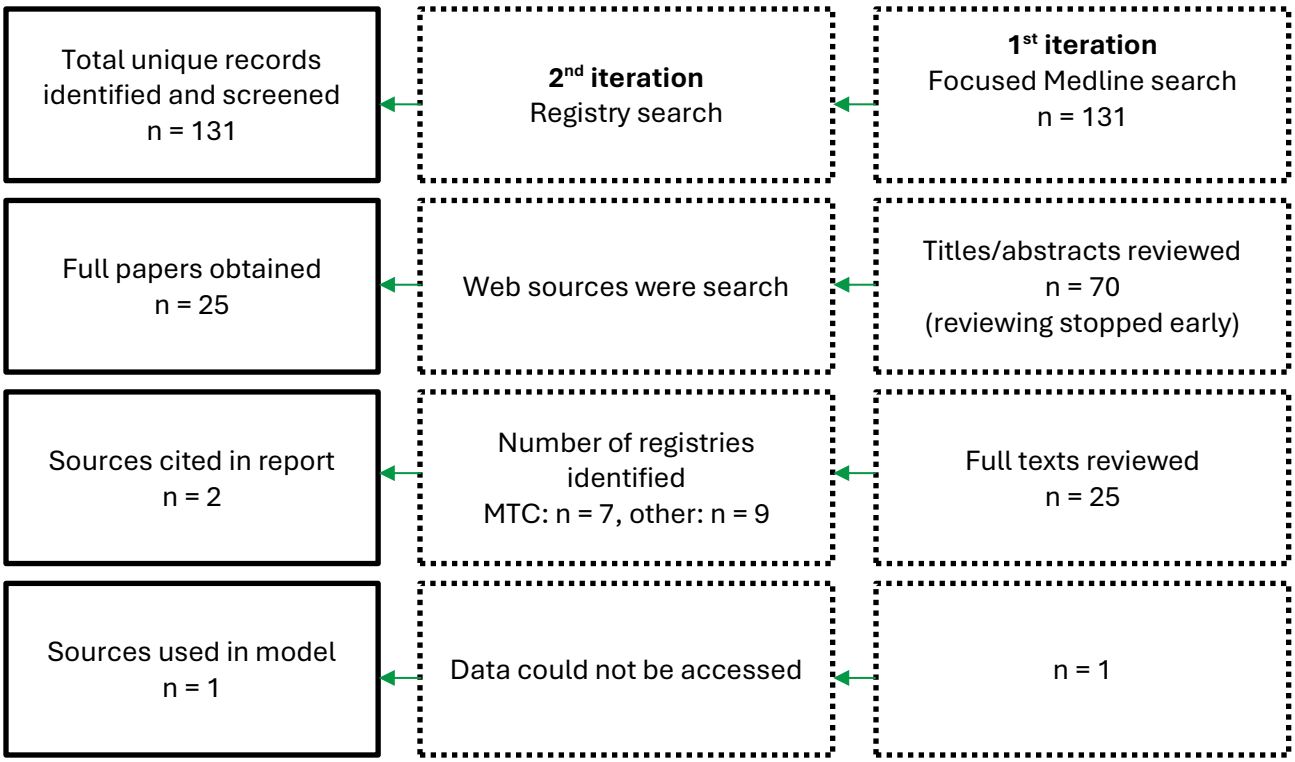


Table 89. Search tracker – iterative search

Item	Iteration 1		Iteration 2		Total		Precision	NNR
	Number	Time (mins)	Number	Time (mins)	Number	Time (mins)		
Search protocol development		20		190		210		
Running searches		10		40		50		
Identified citation	131				131	0		
Duplicate removal	0	0			0	0		
Title/abstract level	70	60			70	60		
Citations excluded	45				45	0		
Full papers to retrieve + review	25	180			25	180	1.53%	66
Sources cited in the report	2		16 registries		2		0.76%	131
Sources used in model	1		N/A		1			
TOTAL		**270**		**230**		**500**		

Abbreviations: mins: minutes

6.4.3 Alternative Search Method 2

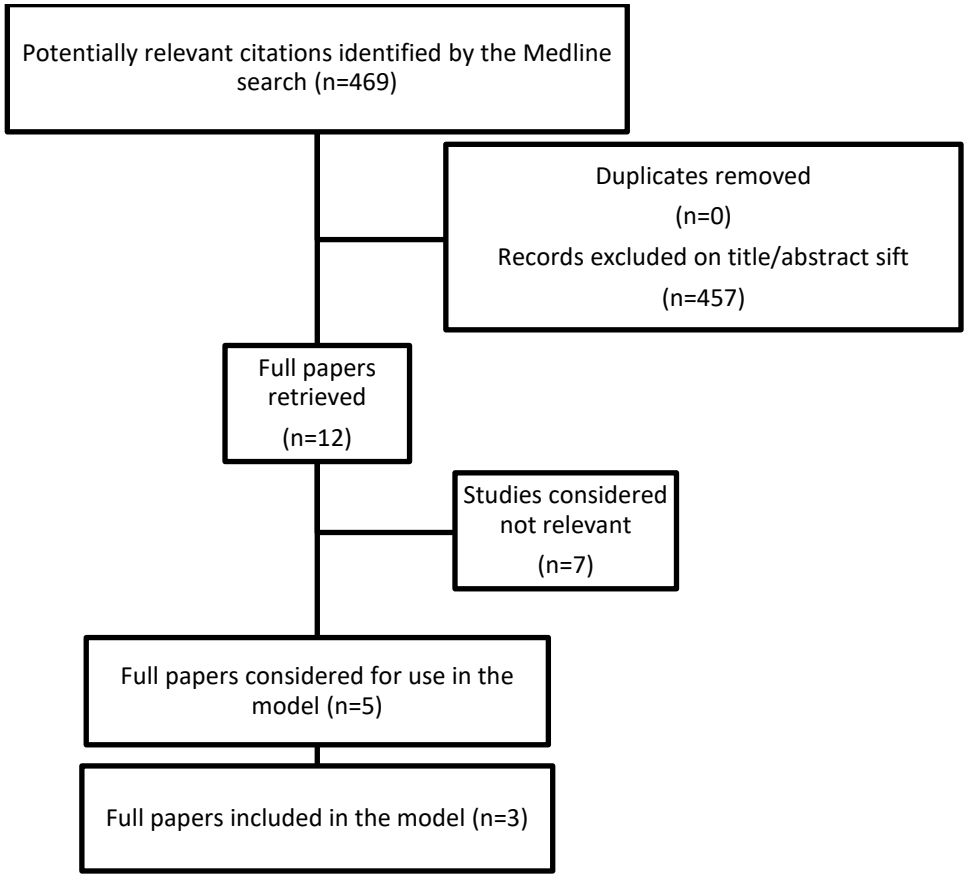
Rapid Review

For this rapid review, Medline (via ProQuest Dialog) was searched from January 2004. The search strategy is given below in Table 90. The rapid review limitations included searching only one database, using free-text search and searching in titles instead of full text. This literature search was conducted in May 2024. References were collected in a bibliographic management database (EndNote). The results are summarised in Figure 24.

Table 90. UC Case study: Alternative search – Rapid Review

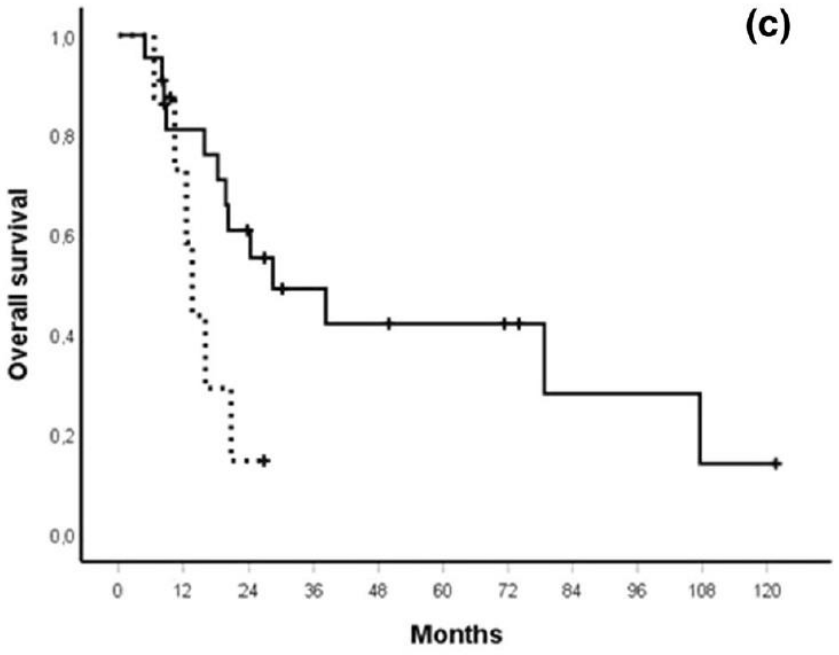
Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms, limiting the search to titles only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	20 years i.e., Jan 2004 – May 2024 Not a case study
Terms used	To search for MTC related sources: ti(medullary thyroid cancer) OR ti(medullary thyroid carcinoma) AND To search in locally advanced or metastatic population: ti(locally advanced) OR ti(metastatic) OR ti(progressive)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Figure 24. Study selection results – Rapid review



The search identified five potential sources to be used in the model for PFS or OS.[376,394,397,399,437] One of the OS sources was only found in this rapid review,[437] while the rest had also been identified in either the usual practice search or iterative search. This publication reported OS in a Slovakian locally advanced or metastatic MTC population. OS was reported for those patients who have not had surgery (see Figure 25, dashed line). When compared to BSC results from EXAM study, patients in in the Slovakian study had considerably higher probability of surviving. PFS was also reported in Kuhar *et al*. but not for the (approximate) target population. It is unclear why differences exist between these single country OS findings and the relevance of them for this health economic model, that uses England as the base case country. It is unlikely that this publication is suitable to be included in the model. Liu *et al*. 2024 was also identified as a source for OS in this rapid review. It was also identified in the iterative search, and discussed in more detail in the previous Section. The reasons for excluding it from modelling are similar to Kuhar *et al* paper: the OS seems considerably higher in the study population compared to OS in EXAM study. Both Kuhar *et al*. and Liu *et al*. were included in the report but not in the health economic model. Three studies were included in both the report and model.[376,394,397] All of these were identified in the usual practice search and one of them was identified in the iterative search. The results of the searches are summarised in Table 91 for PFS and Table 92 for OS.

Figure 25. Overall survival reported in Kuhar *et al*. 2021 (in Figure 2C)



Legend: Dashed line: no surgery (approximately comparable to EXAM study), solid line: surgery.
Copyright license number: 5878840193953

Table 91. Overview of sources considered for PFS

Study	Usual Practice (n=6)	Iterative search (n=1)	Rapid Review (n=4)
ZETA study[394]	✓		✓
ZETA clinicaltrials.gov[395]	✓		
ZETA unpublished EU label data[377]	✓		
EXAM study[376]	✓		✓
EXAM follow-up, final data cut[397]	✓	✓	✓
EXAM clinicaltrials.gov[396]	✓		
Kuhar *et al*. 2021[437]			✓

Table 92. Overview of sources considered for OS

Study	Usual Practice (n=6)	Iterative search (n=2)	Rapid Review (n=5)
ZETA study[394]	✓		✓
ZETA clinicaltrials.gov[395]	✓		
ZETA unpublished EU label data[377]	✓		
EXAM study[376]	✓		✓
EXAM follow-up, final data cut[397]	✓	✓	✓
EXAM clinicaltrials.gov[396]	✓		
Liu *et al*. 2024[399]		✓	✓
Kuhar *et al*. 2021[437]			✓

Summary of Rapid Review Results and Efficiency

Table 93 displays the estimated effort to identify the model inputs. In the focused Medline search, 469 studies were identified, of which three were used in the model. This effort was estimated to take 590 minutes. The search output had a precision of 1.07% in terms of items cited in the report and 0.64% for items used in the model. The NNR was therefore 94 and 156 for sources cited in the report and sources used in the model, respectively.

Table 93. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		110		
Running searches		20		
Identified citations	469			
Duplicate removal	0	0		
Citations to sift at title/abstract level	469	380		
Citations excluded	457			
Full papers to retrieve + review	12	80		
Sources cited in the report	5		1.07%	94
Sources used in the model	3		0.64%	156
TOTAL	**3**	**590**		

Abbreviations: NNR: number needed to read

6.4.4 Case Study Baseline Risk of Clinical Events in Thyroid Cancer – Summary

The key observations from the MTC baseline risk of clinical events case study were:

- The MTC model included two baseline risk of clinical event parameters: OS and PFS for the comparator group (BSC).
- The clinical effectiveness systematic review was considered to be usual practice, although it was aimed at identifying relative treatment effects from RCTs rather than OS and PFS in MTC more widely i.e., in BSC, in observational studies, etc. Therefore, it could also be argued that usual practice is 'no search'. In this sense, the alternative search methods can be considered supplementary to the usual practice, rather than a pure replacement.
- Usual practice, defined as the clinical effectiveness systematic review conducted by the Assessment Group, identified two Phase III studies - EXAM (cabozantinib versus BSC) and ZETA (vandetanib versus BSC). The EXAM study was associated with three sources: pivotal publication, an OS follow-up publication and clinicaltrials.gov record.[376,396,397] The ZETA study was associated with three sources also: publication of the main data-cut, unpublished data from a NICE submission, and a clinicaltrials.gov record.[377,394,395]
- The alternative search methods (iterative searching, rapid review) showed that it is challenging to identify OS and PFS data from publications that are not reports of the pivotal clinical trials. This is because of population differences. Considerable differences in median OS were observed and it is difficult to judge, without clinical knowledge, whether the population in the publication sufficiently matches that of the clinical study. There were even significant differences between the EXAM and ZETA study ITT populations, that prevented a formal indirect treatment comparison and led to the Assessment Group health economic model to focus on pairwise comparisons.[372]
- Iterative searching identified one source for PFS (EXAM study), and two for OS (EXAM study and the publication by Liu *et al*. 2024).[397,399] Iterative searching did not identify the rest of the pivotal Phase III publications. However, if we consider the iterative search as a supplementary rather than a replacement approach, it can be assumed that these would have already been identified in the clinical effectiveness systematic review.
- Many of the full texts reviewed in iteration 1 of the iterative search were publications that reported data from registries. The review of iteration 1 was stopped early because it became clear that likelihood of finding relevant data was very low. Only 70 of the 131 identified records were sifted. The verification step showed that no relevant publications were missed by stopping the review early.
- As none of the publications reported the registry data in a format that was readily useable in the health economic model, a decision was made to search for registries in iteration 2. Sixteen registries containing potentially relevant data were identified.
- A rapid review was conducted, and it resulted in identification of pivotal Phase III publications, as well as two alternative sources for OS. The two alternative sources were included in the health economic model report but were not used in the model. Liu *et al*. 2024 was also identified in the iterative search, and in addition a publication reporting OS in Slovakian locally advanced or metastatic MTC population

was identified. In both publications, the median OS was considerably higher than in the EXAM study and therefore they were only included in the report and not in the model.

- In terms of effort, the usual practice search took much longer than the iterative search or the rapid review. Usual practice took 3,610 minutes, and iterative search and rapid review took 500 and 590 minutes, respectively (Table 94). These represent reductions of 86% and 84%, respectively. Usual practice can also be seen as the most efficient search as it was part of the clinical efficacy search that needed to be done anyway for the clinical element of the HTA submission. There is, therefore, a judgement in whether to start searching additionally.
- The precision and NNR were similar for the iterative search and rapid review. Usual practice had a lower precision and higher NNR (Table 94).
- If we assume that the usual practice search needs to take place, irrespective of any alternative search method, in order to identify relative treatment effect estimates, then it is useful to compare iterative searching and rapid review to better understand the situations where these are helpful. The nature of the iterative search allowed a fast identification of the fact that there is a low likelihood of finding relevant evidence, and search time could be dedicated for identification of registries instead of sifting through the whole patch of titles/abstracts. Rapid review was slightly better at identifying the observational studies as well as the pivotal RCT publications. However, we can assume that the Phase III trial publications would be identified in the clinical effectiveness search and so the ability to identify those is of limited value, unless this is not conducted/available for some reason.
- This case study suggests that rapid review might be a useful alternative when no clinical effectiveness systematic review is available for modelling. Iterative searching seems to work better in cases where clinical effectiveness systematic review is available, and additional search effort is needed to supplement it, in order to identify potential data and sources for OS and PFS in the comparator group (BSC).

Table 94. Thyroid cancer case study: Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision Report	Precision Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search	6	2	0.16%	0.05%	607	1821	3,610	-
Iterative search	2	1	1.53%	0.76%	66	131	500	-86%
Rapid review	5	3	1.07%	0.64%	94	156	590	-84%

Abbreviations: NNR: Number needed to read. **Footnote:** [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

6.5 Case Study in Thyroid Cancer – Summary

The key observations from the thyroid cancer case study were:

- Iterative searching has emerged as a promising approach for utility input parameters. In this case study, it offered an advantage in terms of speed compared to usual practice. Moreover, the most relevant sources were identified, and the model outputs obtained through iterative searching remain comparable to those achieved using usual practice. Rapid review took longer than both iterative searching and usual practice, and identified fewer references. However, the impact on the model results was minor.
- For the baseline risk of clinical events, usual practice can be defined either as 'no search' or the systematic review of clinical effectiveness can be considered usual practice. Considering these two different usual practice definitions, two potential uses for iterative search and rapid review emerged from this case study. If there is no clinical effectiveness systematic review available to provide 'usual practice' data, then rapid review was more efficient at identifying pivotal Phase III RCTs as well as observational studies. However, the registries were not identified as part of it, and would require approximately an additional 230 minutes.

7 Case Study 3: Breast Cancer Tumour Profiling Risk Stratification

7.1 Chapter Overview

This chapter reports the findings of the breast cancer tumour profiling risk stratification case study. This Chapter provides an introduction to the case study, and then reports the searching for utility (Section 7.3) and baseline risk of clinical events (Section 7.4), separately. For both model input types, usual practice search, iterative search and rapid review methods and results are reported. This includes details of how the search was carried out, what the results were, how studies were selected and used in the model. This is followed by an assessment of impact of the sources on the health economic model results, and then an assessment of using the performance measures outlines in the methods chapter (Section 4.2.2.4). Although the different search methods employed in this case study are reported in the same order as in case studies 1 and 2, they were actually carried out in a different order, as reported in the methods section. Rapid review was carried out first, then usual practice and finally iterative searching.

7.2 Case Study Introduction

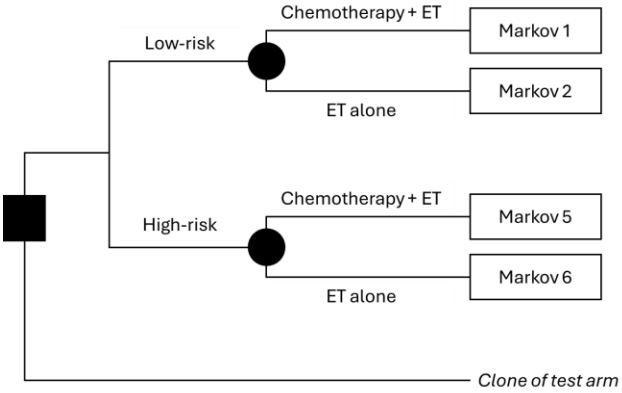
During 2023 and 2024, NICE undertook an appraisal of four tumour profiling tests (Oncotype DX, Prosigna, EPclin and MammaPrint), reported in NICE Diagnostic Guidance 58 (DG58).[438] ScHARR was the External Assessment Group (EAG). In 2018, NICE had published an earlier Diagnostics Guidance 34 (DG34), that recommended the use of three tests (Oncotype DX, Prosigna and EndoPredict (EPclin score)) for guiding chemotherapy decisions in people with oestrogen receptor (ER)-positive, HER2-negative, lymph node negative (LN0) early breast cancer, including those with micrometastases. DG58 represents an update to the systematic review and cost-effectiveness analysis which informed considerations for the LN+ subgroup within NICE DG34.[438,439]

The aim of the technology assessment was to determine whether tumour profiling tests used for guiding adjuvant chemotherapy decisions in patients with ER-positive (and/or PR-positive), HER2-negative, early-stage breast cancer with 1 to 3 positive lymph nodes represent an effective and cost-effective use of NHS resources.[440] Breast cancer is the most commonly diagnosed cancer and the second most common cause of death in women in the UK.[441] Initial treatment typically involves surgery to remove the primary tumour and affected lymph nodes. Subsequent therapies may include radiotherapy, endocrine therapy, targeted therapy, bisphosphonates, and chemotherapy. Chemotherapy can reduce recurrence risk and mortality in early-stage breast cancer, but it comes with adverse effects. To optimise treatment, tumour profiling tests aim to categorise patients by risk and to identify those who would benefit most from chemotherapy.

The EAG developed a health economic model to estimate the incremental cost-effectiveness of Oncotype DX, Prosigna, EPclin and MammaPrint for guiding adjuvant chemotherapy decisions in women with ER+, HER2-, LN+ early breast cancer. Each decision option was compared against current decision-making (which may or may not include the use of risk prediction tools). The model includes seven base case

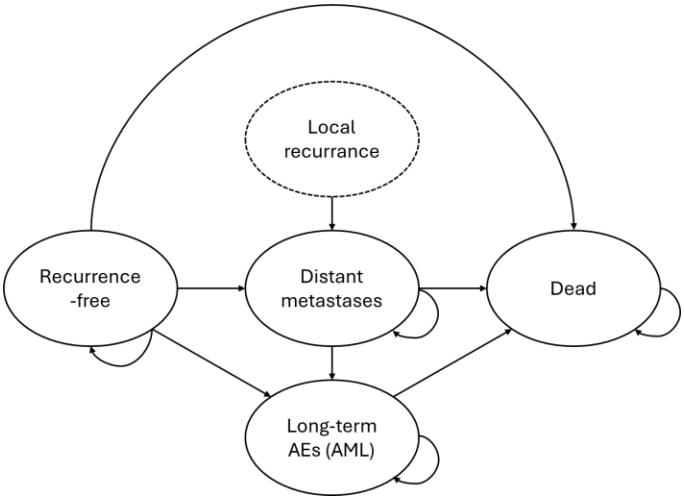
analyses. This case study focuses on base case 2. This relates to Oncotype DX versus current decision-making, using data from post-menopausal LN+ subgroup of the RxPONDER trial,[442] supplemented using external data on women with an Breast Recurrence Score (RS) of >25 (thereby assuming predictive benefit).[443,444] Oncotype Dx is both prognostic and predictive test. It provides information on how likely the breast cancer is to come back, as well as predicting the likelihood of benefit from chemotherapy. This base case was chosen because decision impact data (the probability of receiving chemotherapy with and without the test) in a LN+ population were available for Oncotype DX, but not for the other three tests, and because updated RCT data on the prognostic and predictive value of Oncotype DX (the RxPONDER trial) were available, which triggered NICE's decision to update guidance on the use of tumour profiling tests in LN+ women. The health economic model adopts a hybrid decision tree and Markov structure, shown in Figure 26 and Figure 27, respectively. A modifiable version of the original executable Excel model was made available by one of my supervisors (PT) to test the identified utility data. Further details of the model can be found in the EAG report.[440]

Figure 26. Health economic model diagram – Decision tree component (Figure 5 in EAG report[440])



Abbreviations: ET: endocrine therapy

Figure 27. Health economic model diagram – Markov model component (Figure 5 in EAG report[440])



Abbreviations: AE: adverse events, AML: Acute myeloid leukaemia

7.3 Utility Values

7.3.1 Description of the Usual Practice (Control)

Utility Value Identification

The model includes utility values for health states shown in Figure 27 (recurrence free, local recurrence, distant metastases, long-term AEs (AML) and dead). The EAG report provides details of systematic searches that were undertaken to identify studies reporting on HRQoL associated with different health states for women with breast cancer (recurrence free and distant metastases). The searches focussed specifically on studies which report HRQoL estimates for health states measured and valued using the EQ-5D in the following databases:

- MedlineE Epub Ahead of Print, In-Process & Other Non-Indexed Citations: Ovid, 1946 to present
- EMBASE: Ovid, 1974 to 2017 July 07
- Science Citation Index Expanded (SCI-E): Web of Science, 1900 to present
- Conference Proceedings Citation Index – Science (CPCI): Web of Science, 1990 to present.

The search strategy comprised sensitive MeSH or Emtree Thesauri terms and free-text synonyms for 'breast cancer' combined with free-text synonyms for 'EQ-5D'. The Medline search strategy is presented in Table 95. The search strategy was translated across all databases searched (details found in Appendix 1 of the EAG report[440]). The original search was reported in two HTA reports, DG34 (2018) and the more recent updated DG58 (2024).[439,440] Literature searches were originally conducted in July 2017 (DG34) and four relevant references were identified in the search. In DG58, an update of the DG34 search is reported. I further updated this search in June 2024. Full details of the original usual search methods can be found in the EAG Report.[440]

Table 95. Medline search strategy for utility values - Usual practice

Search Strategy
1. exp Breast Neoplasms/
2. exp mammary neoplasms/
3. exp breast/
4. exp neoplasms/
5. 3 and 4
6. (breast* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullary)).ti.
7. (mammar* adj5 (neoplasm* or cancer* or tumo?r* or carcinoma* or adenocarcinoma* or sarcoma* or dcis or ductal or infiltrat* or intraductal* or lobular or medullar)).ti.
8. 1 or 2 or 5 or 6 or 7
9. (euroqol or euro qol or eq5d or "eq 5d" or eq-5d).tw.
10. 8 and 9
11. limit 10 to yr="2017 -Current"

Relevant studies were assessed based on specific criteria. First, studies reporting EQ-5D valuations for both non-metastatic/early breast cancer and distant metastasis (DM) states were considered, thus reflecting the model structure. The review process involved sifting through study titles/abstracts, followed by retrieving full texts for potentially relevant studies. The inclusion criteria are shown on Table 96.

Table 96. Inclusion criteria for review of published health utility data

Eligibility criteria
Inclusion criteria • English language • Population: early breast cancer population receiving ET • EQ-5D-3L values reported recurrence free patients on ET and for patients who have DM • Must reflect a similar patients group to the target population (either European or UK)

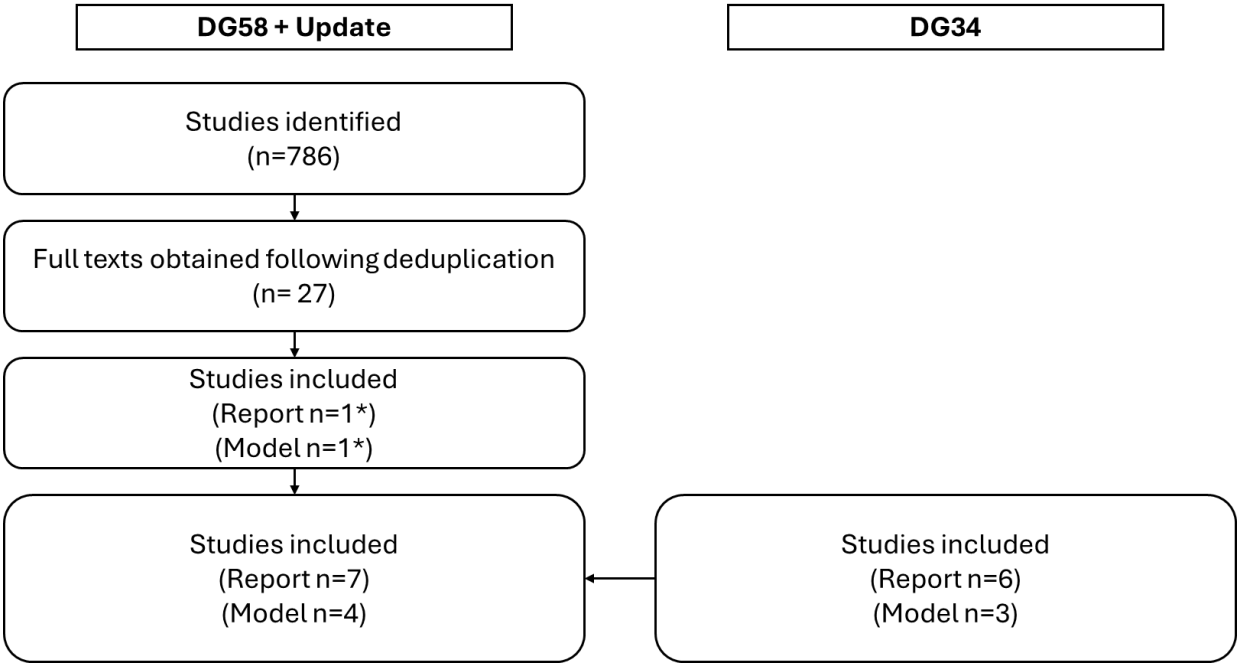
Abbreviations: DM : distant metastasis, ET: endocrine therapy.

The study selection process in DG58 and the update search is summarised in Figure 28. DG58 included six sources from the search reported in DG34. References were collected in a bibliographic management database (EndNote), and duplicates were removed. In the DG58 and my associated update search, 786 citations were identified. Three hundred and eighty-two citations were excluded after deduplication, and therefore 521 citations remained. After titles/abstracts were screened and full texts reviewed, neither the search in DG58 nor my update search identified any relevant new references in addition to those used in DG34. However, DG58 reports a 'near miss' in which the EQ-5D-3L utility values for patients with early and metastatic breast cancer were reported (Verrill *et al*.).[445] The study population was HER2+ (rather than the HER2-), and it was used in a sensitivity analysis. I count this study in 'included' numbers for DG58 in Figure 28 and elsewhere. Six studies from DG34 were included in the report: Lindgren *et al*.[446], Färkkilä *et al*.[447], Yousefi *et al*.[448], Naik *et al.*[449], Bewersdorf *et al*.[450] and Campbell *et al*.[451]. Lindgren *et al*. was the base case source for utility values in the model on the basis that this population was most likely to best reflect the ER+ women with breast cancer who are treated in England.

The model also includes acute myeloid leukaemia (AML) as a health state (Figure 27). AML is included in the model, because it is a long-term complication of chemotherapy resulting in HRQoL losses and costs, as well as having an impact on survival. The EAG model uses Bewersdorf *et al* as the utility source (calculated from rebuilt model).[450] No details are given about how this source was identified and selected.

Additionally, disutility values for two health states were also used in the model: the disutility of receiving chemotherapy and the disutility of local recurrence for those patients who have distant metastases. The EAG report states that Campbell *et al.* was used as a source.[451] The rationale was that it was also a source in the previous economic models for both model inputs: Exact Sciences model, Agendia model and NICE DG34.[439,452,453] Chemotherapy disutility was additionally searched for in the EQ-5D update bibliographic search, but no inputs were identified. No specific other steps of identification or selection were reported for identification of local recurrence utility value. Table 97 shows an overview of the utility sources identified by the usual practice search.

Figure 28. Study selection results for utility search – Usual practice for recurrence and distant metastasis health states



Abbreviations: DG: Diagnostic Guidance. **Footnote**: * Verrill *et al*. was reported as a '*near miss*' in the DG58 update search. Here it is counted in the numbers as it was included in the model sensitivity analysis.

Table 97. Overview of utility sources considered and used in model – Usual practice

Study	Report (n=7)	Included in model (n=4)
Lindgren *et al*.[446] (base case)	✓	✓
Verrill *et al*.[445]	✓	✓
Färkkilä *et al*.[447]	✓	
Yousefi *et al*.[448]	✓	
Naik *et al*.[449]	✓	
Bewersdorf *et al*.[450]	✓	✓
Campbell *et al*.[451]	✓	✓

Utility Values Used in the Model

This section provides an overview of the utility values selected for use in the model. The health state utility values from Lindgren *et al*. were used as the base case for the recurrence and distant metastasis health states.[446] The health state utility values in this study were associated with key model health states (recurrence free, distant recurrence) and the utility values were specific to an ER+, HER2- early breast cancer population. The alternative utility values for these health states were taken from Verrill *et al*.[445] For AML health state Bewersdorf *et al* was used.[450] Campbell *et al* was the source for chemotherapy disutility and local recurrence.[451] The utility values applied in the health economic model are summarised in Table 98.

Table 98. Health state utility values used in the model - Usual practice

	Recurrence free	Distant metastases	AML	Chemotherapy	Local recurrence
Base case: Lindgren *et al*.[446]	0.824	0.685	-	-	-
Base case: Bewersdorf *et al*.[450]	-	-	0.590[a]	-	-
Base case: Campbell *et al*.[451]	-	-	-	-0.038	-0.108
Scenario 1: Verrill *et al*.[445]	0.73	0.60	-	-	-

Abbreviations: AML: Acute myeloid leukaemia. **Footnote:** a: calculated based on the published, rebuilt cost-effectiveness model. The external assessment group estimated the utility by dividing mean quality adjusted life years by the mean life-years gained.

Health Economic Model Results

In the base case, Oncotype DX produces more QALYs at lower cost than current decision-making. When running the model with the alternative utility value, similar results were obtained i.e., Oncotype DX was the dominating strategy. Table 99 presents these results.

Table 99. Health economic model results – Usual practice

	ICER (£)
Base case: Lindgren *et al*.[446]	Dominating
Scenario 1: Verrill *et al*.[445]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio. **Footnotes:** ALM, chemotherapy and local reoccurrence health state utilities values were identical in the base case and scenario 1 (Bewersdorf *et al.*[450] and Campbell *et al.*[451])

Search Efficiency

The usual practice search took 850 minutes (approx. 15.5 hours). This included running the searches, citation identification, and citation and full-text screening against the inclusion criteria. Table 100 displays the estimated effort that went into searching, using the usual search method, for both the published search from DG58 (covering the time period from February 2017 to May 2023) and an update of that

search (covering the time period from May 2023 to June 2024). No search time could be estimated for identification of local recurrence, AML or chemotherapy disutility as no identification steps were reported.

The time spent doing the update search was recorded and rounded to the nearest 5-minute increment. The time for the published, original search needed to be estimated, as this was not available from the EAG report. The estimation was based on using the times from the update search per record screened. These methods are explained in the Methods Section and a worked-up example is given in case study 1. As with case study 1, it was not possible to estimate the time for the original search protocol development or running of the searches. No additional time was assumed for search protocol development and running of those original searches. The time for running the search update was recorded, and included in the estimate. It is likely to be an underestimate of the effort, and this represents a limitation of the study.

Table 100. Search tracker - Usual practice

Item	Original and Update Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Original	Update	Total	Original	Update	Total		
Search protocol development						Not known		
Running searches					80	80		
Identified citations	669	117	786					
Duplicate removal	265	117	382	130	60	190		
Title/abstract level	404	117	521	320	90	410		
Citations excluded	381	0	381					
Full papers to retrieve + review	23	4	27	160	30	190		
Sources outside bibliographic search	7	0						
Sources cited in the report	7	0	7				0.89%	112
Sources used in model	4	0	4				0.51%	197
TOTAL			**4**	**450**	**150**	**870**		

Abbreviations: NNR: number needed to read.

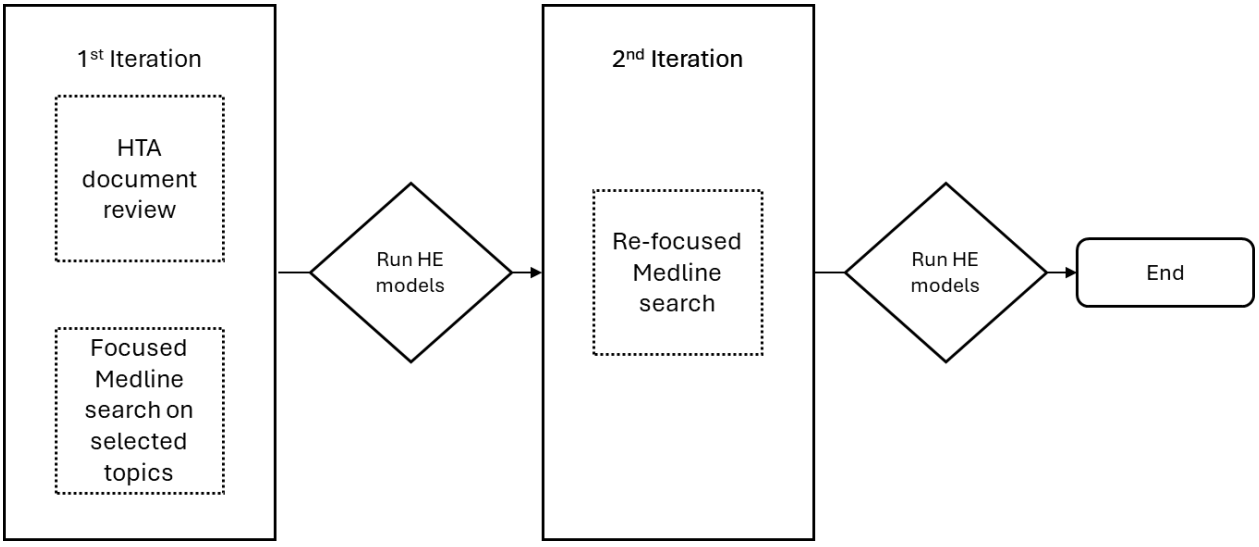
Seven sources were included for consideration in the model from the usual practice bibliographic search, including both the original and update searches. Eventually four sources from the bibliographic database search were used in the model (Table 97).[445,446,450,451] The update bibliographic database search did not identify further relevant sources. The search output had a precision of 0.89% in terms of items cited in the report and 0.51% for items used in the model. The NNR was therefore 112 and 197 for sources cited in the report and sources used in the model, respectively.

7.3.2 Alternative Search Method 1

Iterative Searching

Similar to case studies 1 and 2, this iterative search aimed to maximise the rate of gain for relevant sources by using techniques associated with berry-picking and information foraging models.[309,310] An important aim is also to ensure steps of input identification are recorded for all health state utility inputs. This first iteration aimed to identify those utility sources that were already identified through literature reviews by others as relevant for use in health economic models, as well as any new information that may have become available since the most recent HTA review. The first iteration therefore included a review of HTA documents that include literature reviews of utility values for health economic models, and a highly precise bibliographic search of the most recent evidence (searched in the titles only). The most recent evidence in the title only search was limited to the period since the last published HTA report. Any subsequent search iteration(s) aimed to understand whether further bibliographic searching would provide additional, relevant evidence for the health economic model, specifically for the heath states where only one source had been identified in the first iteration. Figure 29 shows the concept of the iterative search. The health economic model was run between each search iteration to assess the marginal impact of identifying additional evidence on the health economic model results, and to assess whether continuing searching would be likely to identify further, relevant evidence.

Figure 29. Summary of iterative search



Abbreviations: HTA: Health technology assessment

Iteration 1

In the first search iteration, the past NICE diagnostic guidance documents (DG58, DG34 and DG10) were identified and browsed through for any systematic literature reviews of utility values (see Table 101). The documents were reviewed to collect information on which utility studies had been identified in past

literature reviews, and subsequently selected for inclusion in health economic models. The most recent NICE diagnostic guidance was DG58 for tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer. Two other diagnostic guidance documents were identified (DG34 and DG10). DG10 covered tumour negative lymph nodes (LN0) only, and is no longer available online. DG34 covered LN0 and LN+.[440] DG58 is LN+ only, and therefore replaces part of DG34. In the DG58, the literature search for utilities was performed in May 2023.[438] The searches focussed on estimates derived by using EQ-5D. A focused update (title-focused) search from May 2023 to present (June 2024) was carried out to retrieve any records that may not have been captured by the literature reviews in the DG document. Both the non-bibliographic and bibliographic searching was performed in June 2024. References were collected in a bibliographic management database (EndNote).

Table 101. Iterative search: 1st iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit from May 2023 – present (June 2024)
Terms used	MESH.EXACT("Breast Neoplasms") AND ALL(EQ-5D)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	NICE Guidance in breast cancer were reviewed: https://www.nice.org.uk/guidance/conditions-and-diseases/cancer/breast-cancer/products?ProductType=Guidance&Status=Published
Source names	DG58: https://www.nice.org.uk/guidance/dg58 DG34: https://www.nice.org.uk/guidance/dg34 DG10: https://www.nice.org.uk/guidance/dg10
Search dates	12 - 15 June 2024
Limits	No further limits

The NICE DG documents were reviewed for systematic reviews of utility values, as well as any utility values reported from clinical trials or other sources. As described in the usual practice section previously, DG58 also contained utility values for further health states, but the identification and selection processes for those inputs were not reported. DG34 and DG10 were also reviewed for any additional information on the identification process, as well as any other values that may have been discarded as data sources from those submissions.

The second element of the first iteration, the title-focused update literature search of a single database was run on the 13th June 2024, and 19 records were identified. Sixteen citations were excluded at title level. Three full texts were checked, but no new studies were found. Therefore, this bibliographic search update did not identify any new sources. A total of five sources were identified (Table 102), all from the NICE DG documents. Four of those were also identified in the usual practice search.[445,446,450,451] One study was not included in DG58 (Younis *et al*.) but was used in the base case in DG34. Unlike in the usual practice search, it is included again here as a sensitivity analysis to better understand the model's sensitivity to this model input, and to determine whether further searching may yield more relevant information.

Table 102. Overview of utility sources considered for and used in model – Iterative search, iteration 1

Study	Usual Practice		Iterative search – iteration 1	
	Report (n=7)	Included in model (n=4)	Report (n=8)	Included in model (n=5)
Lindgren *et al*.[446] (base case)	✓	✓	✓	✓
Verrill *et al*.[445]	✓	✓	✓	✓
Färkkilä *et al*.[447]	✓		✓	
Yousefi *et al*.[448]	✓		✓	
Naik *et al*.[449]	✓		✓	
Bewersdorf *et al*.[450]	✓	✓	✓	✓
Campbell *et al*.[451]	✓	✓	✓	✓
Younis *et al*.[454]			✓	✓

Utility Values Used in the Model from Iteration 1

Lidgren *et al*. provided inputs for two key health states (recurrence and distant metastases) in the base case.[446] Two further sources were used for other health states in the base case, Bewersdorf *et al*. and Campbell *et al*. [450,451] Additionally, two other sources were identified for scenario analyses: Verrill *et al*. for recurrence and distant metastasis health states (scenario 1) and Younis *et al*. for AML the health state (scenario 2) (Table 103). All sources identified in the iteration 1 of the iterative search, were also identified in the usual practice search. Additionally, Younis *et al*. was identified from DG34 and used as a scenario.

Table 103. Health state utility values used in the model - Iterative search, iteration 1

	Recurrence free	Distant metastases	AML	Chemotherapy	Local recurrence
Base case: Lindgren *et al*.[446]	0.824	0.685	-	-	-
Base case: Bewersdorf *et al*.[450]	-	-	0.590[a]	-	-
Base case: Campbell *et al*.[451]	-	-	-	-0.038	-0.108
Scenario 1: Verrill *et al*.[445]	0.73	0.60	-	-	-
Scenario 2: Younis *et al*.[454]	-	-	0.26	-	-

Abbreviations: AML: Acute myeloid leukaemia.

Health Economic Model Results

The utility sources identified in the iterative search were identical to those identified in usual practice, with the addition of the Younis *et al*. utility value for the AML health state (scenario 2 below). In all cases (base case , scenario 1, and scenario 2) Oncotype DX was the dominating strategy. The new scenario 2 (AML utility value derived from Younis *et al.* instead of Bewersdorf *et al*.) resulted in very similar results as using the base case utility value for AML. Despite these values being considerably different from each other (0.26 in Younis *et al.* versus 0.588 in Bewertsdorf *et al.*), the results were almost unchanged. This is likely to be due to only small proportion of patients developing AML in the model. The results are shown in Table 104.

Table 104. Health economic model results – Iteration 1

	ICER (£)
Base case: Lindgren *et al*.[446]	Dominating
Scenario 1: Verrill *et al*.[445]	Dominating
Scenario 2: Younis *et al*.[454]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Iteration 2

After iteration 1, searching for three of the health states was stopped: recurrence, distant metastases and AML. This was done because changing the base case inputs to the scenario inputs, did not change the model results (marginal relevance).

It is not clear how chemotherapy and local recurrence utility values in DG58 had been identified and selected for inclusion in the model, as no information retrieval steps were reported. There are no alternative chemotherapy and local recurrence utility values to test as a scenarios to better understand the relevance of further evidence for these health states. Therefore, additional searching was carried out that aimed to specifically identify (dis)utility values for chemotherapy and/or local recurrence.

Focused Medline searching was done that combined search terms for breast cancer and chemotherapy/local recurrence and EQ-5D/utility. Further, economic evaluations in early breast cancer were reviewed. Table 105 shows the details for the search.

Table 105. Iterative search: 2nd iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	No limitation
Terms used	(MESH.EXACT(Breast Neoplasms)) AND ab(early) (MESH.EXACT.EXPLODE("Cost-Effectiveness Analysis")) OR ab(cost-effective*) S2 AND S1 chemotherapy OR recurrence S4 AND S3
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

The search identified 318 citations, of which 302 were excluded at title/abstract level. Sixteen full texts were reviewed. Of these, one publication that reported disutilities for both chemotherapy and local recurrence had already been identified in the first search iteration (Campbell *et al*.)[451] Two new references were identified that contained disutility values for chemotherapy (Table 106). Vaidya *et al*. 2017 contained utility values for inoperable early breast cancer, including disutility due to chemotherapy. The utilities were based on EQ-5D collected in UK general population. Saptaningsih *et al*. 2022[455] also reported disutility due to chemotherapy. In this study Indonesia Breast Cancer Health-Related Quality of Life (INA-BCHRQoL) was mapped to EQ-5D index using equations developed for Japan. Both Vaidya *et al*. and Saptaningsih *et al*. reported similar utility reduction, 0.078 and 0.07, respectively. Vaidya *et al*. was run in the model as a sensitivity analysis (scenario 3). Saptaningsih *et al*. was listed in the report but not included in the model. No new information for local recurrence was identified.

Table 106. Overview of utility sources considered for and used in model – Iterative search, iteration 1

Study	Usual Practice		Iterative search – iteration 1 + 2	
	Report (n=7)	Included in model (n=4)	Report (n=10)	Included in model (n=6)
Lindgren *et al*.[446] (base case)	✓	✓	✓	✓
Verrill *et al*.[445]	✓	✓	✓	✓
Färkkilä *et al*.[447]	✓		✓	
Yousefi *et al*.[448]	✓		✓	
Naik *et al*.[449]	✓		✓	
Bewersdorf *et al*.[450]	✓	✓	✓	✓
Campbell *et al*.[451]	✓	✓	✓	✓
Younis *et al*.[454]			✓	✓
Saptaningsih *et al*.[455]			✓	
Vaidya *et al*.[456]			✓	✓

Utility Values Used in the Model from Iteration 2

A sensitivity analysis with chemotherapy disutility from Vaidya *et al*. was added as a scenario 3 (Table 107).

Table 107. Health state utility values used in the model - Iterative search, iteration 2

	Recurrence free	Distant metastases	AML	Chemotherapy	Local recurrence
Base case: Lindgren *et al*.[446]	0.824	0.685	-	-	-
Base case: Bewersdorf *et al*.[450]	-	-	0.590[a]	-	-
Base case: Campbell *et al*.[451]	-	-	-	-0.038	-0.108
Scenario 1: Verrill *et al*.[445]	0.73	0.60	-	-	-
Scenario 2: Younis *et al*.[454]	-	-	0.26	-	-
Scenario 3: Vaidya *et al*.[456]	-	-	-	-0.078	-

Abbreviations: AML: Acute myeloid leukaemia.

Health Economic Model Results

The utility sources identified in the second iteration of iterative search were identical to those identified in usual practice, with the addition of the Vaidya *et al*. disutility value for chemotherapy (scenario 3 below). In all cases (base case , scenario 1 - 3) Oncotype DX was the dominating strategy. The new scenario 2 (AML utility value derived from Younis *et al.* instead of Bewersdorf *et al*.) resulted in very similar results as using the base case utility value for AML. Despite these values being considerably different from each other (0.26 in Younis *et al.* and 0.588 in Bewersdorf *et al.),* the results were almost unchanged. This is likely to be due to the very low risk of developing AML in the model. The results are shown in Table 108.

Table 108. Health economic model results – Iteration 2

	ICER (£)
Base case: Lindgren *et al*.[446]	Dominating
Scenario 1: Verrill *et al*.[445]	Dominating
Scenario 2: Younis *et al*. 2008[454]	Dominating
Scenario 3: Vaidya *et al*.[456]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Decision to Stop Searching

The second iteration, uncovered two further references that contained chemotherapy disutility values.[455,456] These values were both around 0.07, which is different from the base case value reported by Campbell *et al*. (0.038).[451] Despite this value being higher than the base case estimate, the model results were not substantially different, and therefore the conclusions that could be drawn from the model did not change. No values were identified for local recurrence. Due to the limited evidence base and low likelihood of finding further evidence that would change the conclusion from the health economic model, a decision was made to stop searching after iteration 2.

Summary of Iterative Search Results and Efficiency

In total, ten sources were included for consideration in the model from iterative searching, and six were actually included in the model analyses Figure 30). These publications are listed in Table 107. Iterative searching identified all the same sources as the usual practice search, plus three additional sources for the report and two for the model. The iterative search output had a precision of 3% in terms of items cited in the report and 2% for items used in the model. The NNR was therefore 34 and 56 for sources cited in the report and sources used in the model, respectively.

Figure 30. Summary of results – Iterative searching tumour profiling risk stratification

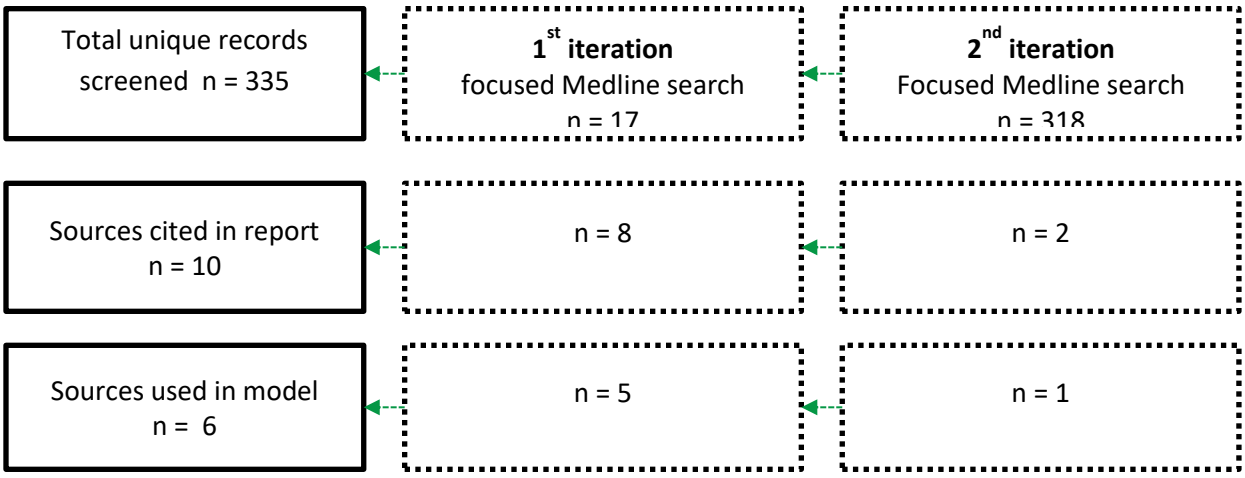


Table 109 displays the estimated search effort. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening took 530 minutes (8 hours and 50 minutes).

Table 109. Search tracker – Iterative search

Item	Iterative search							
	Number of studies identified			Time in minutes			Precision	NNR
	Iteration 1	Iteration 2	Total	Iteration 1	Iteration 2	Total		
Search protocol development				75	25	100		
Running searches				35	15	50		
Identified citations	17	318	335					
Duplicate removal	0		0	0	0	0		
Title/abstract level	17	318	335	10	250	260		
Citations excluded	14	302	316					
Full papers to retrieve + review	3	16	19	10	110	120		
Sources cited in the report	8	2	10				3%	34
Sources used in model	5	1	6				2%	56
TOTAL			**6**			**530**		

Abbreviations: NNR: number needed to read

7.3.3 Alternative Search Method 2

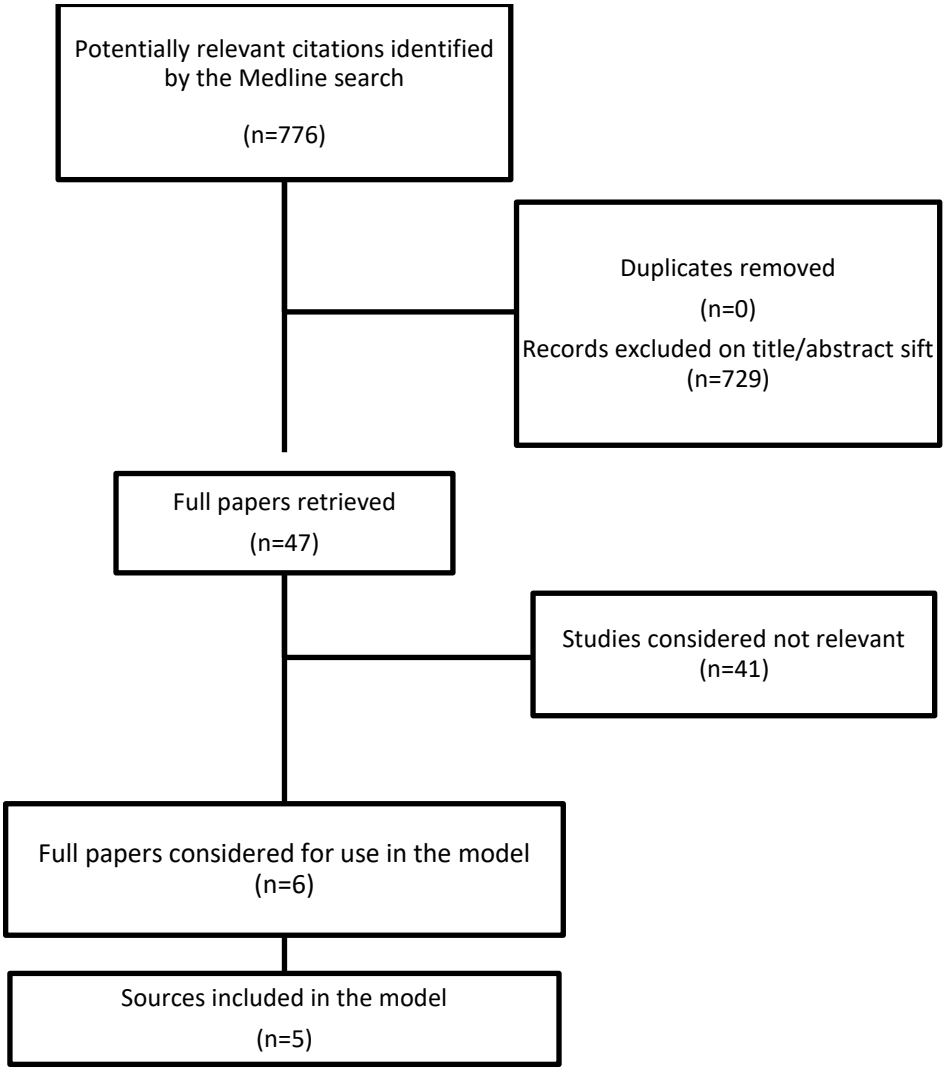
Rapid Review

Another experiment in the tumour profiling risk stratification case study was using rapid review methods. For this rapid review, Medline (via ProQuest Dialog) was searched from inception. The search strategy combined MeSH and free text terms relating to early breast cancer and acute myeloid leukemia (AML) combined with free text terms for EQ-5D and cost-effectiveness. All other health states, except AML, could be (at least in theory) identified using early breast cancer search terms. To ensure coverage of utility values across all health states, AML was therefore added as a separate search term. An overview of the search strategy is shown in Table 110. The rapid review limitations applied included using only one database (Medline), including publication for the last 20 years only and restricting synonyms. Literature searches were conducted in June 2024. References were collected in a bibliographic management database (EndNote). Figure 31 summarises search results.

Table 110. Rapid Review

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms.
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	No limitation
Terms used	S1 (MESH.EXACT(Breast Neoplasms) AND EQ-5D) OR (MESH.EXACT(Breast Neoplasms) AND cost-effectiveness) S2 (MESH.EXACT(Leukemia, Myeloid, Acute) AND EQ-5D) OR (MESH.EXACT(Leukemia, Myeloid, Acute) AND cost-effectiveness) S3 ab(early) OR ab(first) OR ab(primary) S4 S3 AND S1 S5 S4 OR S2
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Figure 31. Study selection results – Rapid review



In the search, 776 citations were identified. Of those, 729 were excluded at title/abstract screening and 47 full-text papers were retrieved. After reading the papers, six studies were included in the report and five were run in the model. The utility sources identified during the searches are summarised in Table 111.

Table 111. Breast cancer risk stratification case study: Studies included – Rapid Review

Study	Usual Practice		Iterative search		Rapid Review	
	Report (n=7)	Model (n=4)	Report (n=10)	Model (n=6)	Report (n=6)	Model (n=5)
Lindgren *et al*.[446] (base case)	✓	✓	✓	✓	✓	✓
Verrill *et al*.[445]	✓	✓	✓	✓	✓	✓
Färkkilä *et al*.[447]	✓		✓			
Yousefi *et al*.[448]	✓		✓			
Naik *et al*.[449]	✓		✓			
Bewersdorf *et al*.[450]	✓	✓	✓	✓	✓	✓
Campbell *et al*.[451]	✓	✓	✓	✓	✓	✓
Younis *et al*.[454]			✓	✓		
Saptaningsih *et al*.[455]			✓		✓	
Vaidya *et al*.[456]			✓	✓	✓	✓

Utility Values Used in the Model from Rapid Review

Table 112 reports the utility values for the base case and scenarios. All sources identified in the rapid review were also identified either in the usual practice or iterative searches. Rapid review did not identify as many sources as iterative search, but all key sources used in the base case were identified. [446,450,451]

Table 112. Health state utility values used in the model – Rapid review

	Recurrence free	Distant metastases	AML	Chemotherapy	Local recurrence
Base case: Lindgren *et al*.[446]	0.824	0.685	-	-	-
Base case: Bewersdorf *et al*.[450]	-	-	0.590[a]	-	-
Base case: Campbell *et al*.[451]	-	-	-	-0.038	-0.108
Scenario 1: Verrill *et al*.[445]	0.73	0.60	-	-	-
Scenario 2: Vaidya *et al*.[456]	-	-	-	-0.078	-

Health Economic Model Results

The utility sources identified in the rapid review were identical to those identified in usual practice and iterative search. The health economic model results were therefore identical as well, base case and both scenarios resulted in Oncotype DX being the dominating strategy. The results are shown in the previous iterative search Section (Table 104).

Summary of Rapid Review Results and Efficiency

Six full-texts were included for consideration in the model, and five were actually included. Table 113 displays the number of sources found by the search and the estimated search effort. The development of the search protocol, running of the searches, citation identification, duplicate removal, and publication screening took 1,020 minutes (17 hours). The search output had a precision of 0.77% in terms of items cited in the report and 0.64% for items used in the model. The NNR was therefore 129 and 155 for sources cited in the report and sources used in the model, respectively.

Table 113. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		55		
Running searches		15		
Identified citations	776			
Duplicate removal	0	0		
Citations to sift at title/abstract level	776	620		
Citations excluded	729			
Full papers to retrieve + review	47	330		
Sources cited in the report	6		0.77%	129
Sources used in the model	5		0.64%	155
TOTAL	5	**1020**		

Abbreviations: NNR: number needed to read

7.3.4 Case Study Utility Values in Tumour Profiling Risk Stratification – Summary

The key observations from the breast cancer risk stratification utility case study were:

- Utility values identified in the usual practice search came from a systematic review (recurrence, distant metastasis) as well as from previous health economic models (chemotherapy and local recurrence). One utility value was calculated using an existing AML health economic model (AML health state). The identification and selection process was transparently recorded only for those utility sources identified in the systematic review.
- All search methods found a similar evidence base, although iterative searching found the most studies for inclusion in the report. This method also took the least time (530 minutes versus 870 and 1,020 minutes in iterative, usual practice and rapid review searches) (Table 114).
- Iterative searching was more precise and the NNR was lower than for either usual practice or rapid review (56 for iterative search, compared to 197 and 155 for usual practice and rapid review, respectively).
- Usual practice and rapid review took longer than iterative searching - 870 and 1,020 minutes respectively. Rapid review took 17% more time than usual practice. Due to methodological limitations associated with this study (i.e., original search protocol development time and search running time are not known), time for usual practice may be underestimated. Further, rapid review included literature review that covered all the health states, not only some like in usual practice. Therefore, rapid review was associated with increased transparency over usual practice.
- Model results were identical for all three search approaches, as all search approaches found similar sets of evidence. The differences in identified sources did not result in differences in model results in this case study, but that might not hold true for other models.
- In conclusion, the iterative search method identified all the relevant utility values and it was also the most efficient search method in this risk stratification case study.

Table 114. Breast cancer risk stratification case study: Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search	7	4	0.89%	0.51%	112	197	870	-
Iterative search	10	6	2.99%	1.79%	34	56	530	-39%
Rapid review	6	5	0.77%	0.64%	129	155	1,020	17%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

7.4 Baseline Risk of Clinical Events

7.4.1 Description of the Usual Practice (Control)

The breast cancer tumour profiling risk stratification model includes an input for probability of death due to distant metastasis (DM) (Table 113). Other potential model inputs, such as risk classification probabilities or distant recurrence-free interval (DRFI) varied by risk stratification test and/or were associated with systematic searches beyond the scope of this project. For this reason, the probability of death due to DM was chosen as the focus for this part of the case study.

Table 115. Overview of baseline risk of clinical event parameters in the breast cancer model

Model Input	Search method	Search dates	Search strategy
6-month probability of death due to DM	No literature review is reported	Not applicable	

Footnote: DM: distant metastasis.

Identification – Probability of Death Due to Distant Metastasis

The EAG report describes how the previous health economic model developed for DG34 applied the probability of death due to DM from Thomas *et al*.[457] However, this source was not used in DG58, because the clinical advisors commented that the majority of women with ER+ breast cancer who develop distant metastases in England now receive a cycline dependent kinase 4/6 inhibitors (CDK4/6i), such as abemaciclib, palbociclib or ribociclib, as first-line treatment. Receiving these treatments has an impact on OS, and therefore a new source for probability of death due to DM was needed. A health economic evaluation by Suri *et al*. was identified for ribociclib plus letrozole versus palbociclib plus letrozole for the treatment of post-menopausal women with HR+, HER2- advanced breast cancer.[458] The process of how this study was identified, or became to be selected, was not described in the EAG report. The EAG modeller was one of my supervisors (PT). Therefore, it was possible to find out further information about the usual practice searching. PT explained that he knew that CDK4/6 inhibitors had become first-line therapy for women with ER+ disease, so he looked for NICE appraisals of ribociclib, palbociclib or abemaciclib. The NICE documents did not present the health economic models in detail, so PT then searched for existing models/reviews of models of these drugs. This was either done just via Google, or possibly using Medline (PT does not recall exactly). A systematic review of pharmacoeconomic evaluations was identified (Zhu *et al*.)[459] and Suri *et al*. was the only UK-based publication included in this review.[458] PT retrospectively estimated that the searching took around 1.5 days (i.e., 12 hours). Because this estimate is done retrospectively, and prone to bias. This is recognised as a limitation of the study.

The EAG replicated the published OS Kaplan-Meier function for the ribociclib plus letrozole group by using the reported parameters of the baseline Weibull model for OS and HRs obtained from a matching-adjusted indirect comparison (MAIC). From this process, the EAG derived a mean OS estimate of 4.63 years. Therefore the model applies a 6-month probability of death of 0.102, assuming a constant event

rate. For the purposes of implementing any additional publications identified during this case study, the Weibull survival curve was fitted to the Kaplan-Meier data to estimate mean OS. Other models may have better statistical fit and/or clinical plausibility than the Weibull, which is recognised as a limitation of this study. This simplistic approach was taken for practicality as survival models are not the focus of this thesis.

Baseline Risk – Health Economic Model Results

In the base case, Oncotype DX produces more QALYs at lower cost than current decision-making (Table 116). No deterministic sensitivity analyses reported in the EAG report for this model parameter.

Table 116. Health economic model results Oncotype Dx versus current decision-making – Usual practice

	ICER (£)
Base case: Suri *et al*.[458]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Search Efficiency

No literature review was reported for the identification of risk from DM in the EAG report, and therefore it was not possible to run an update search. Therefore, it was not possible to identify how long the identification and selection process took. The time was estimated retrospectively by the health economist working on the assessment. Table 117 displays the estimated effort to identify risk of death from DM.

Table 117. Search tracker - Usual practice

Item	Original Search							
	Number of studies identified			Time in minutes			Precision	Number needed to read
	Original	Update	Total	Original	Update	Total		
Search protocol development								
Running searches				480		480		
Identified citation	0		0					
Duplicate removal	0		0					
title/abstract level	0		0					
Citations excluded	0		0					
Full papers	0		0					
Sources cited in the report	1		1				100%	1
Sources used in model	1		1				100%	1
TOTAL			**1**			**480**		

Abbreviations: NNR: number needed to read

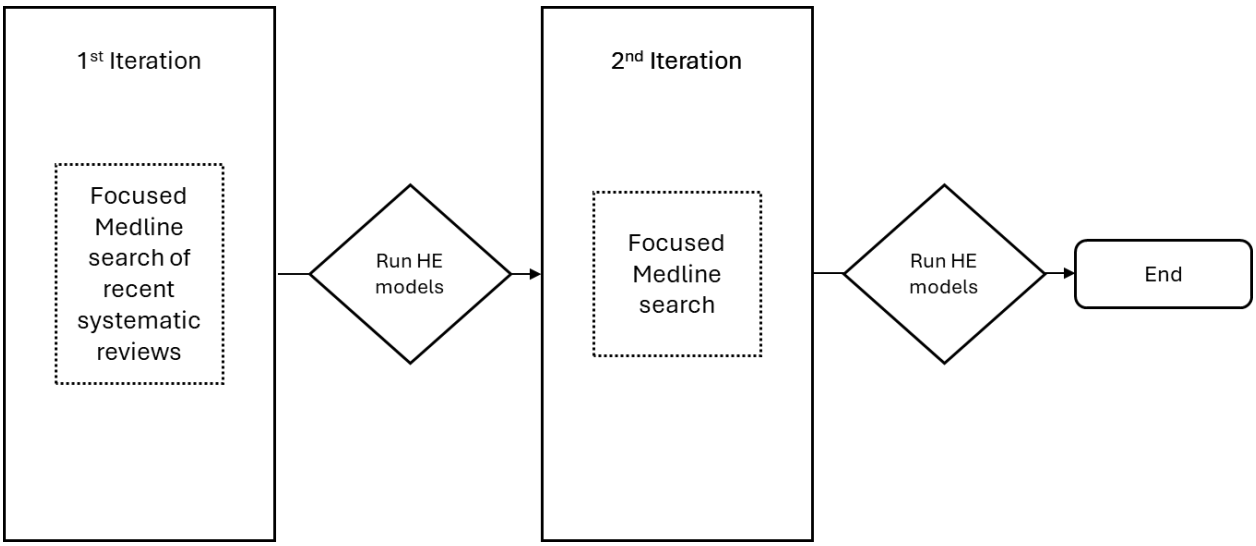
7.4.2 Alternative Search Method 1

Iterative Searching

The first iteration included an exploratory bibliographic search in titles of literature reviews only, with the aim of identifying a quick overview of evidence for baseline risk of death from DM. The second iteration included a search that was not limited by study type and designed to capture recent, primary studies reporting OS for CDK4/6 inhibitors in the target population.

Figure 32 shows the concept of the iterative search. In the previous case studies, a registry search was done as part of the baseline risk of clinical event search. The main reason for the registry search in case studies 1 and 2 was that the identified publications reported data from a registry but in a format that was not useable in the model. For example, the endpoint of interest and/or the population of interest were not included. Additionally, the data might not have been in the right format (e.g., % alive at month X instead of Kaplan-Meier OS plot). The publications were often using data from a registry and therefore, registries were searched for and listed. Accessing the registries would be the only way to access the most relevant data. Unlike in case study 2 (thyroid cancer), in this case study publications were identified that contained the survival data of interest for the relevant population. Several publications were identified with relevant data for the population of interest, and for this reason a registry search was not necessary.

Figure 32. Summary of alternative search method 1: Iterative search (breast cancer tumour profiling risk stratification case study)



Iteration 1

When starting to plan the search for this first iteration, it was not clear which CDK4/6 inhibitor efficacy should be used to model the probability of death due to DM. For this reason, in the first iteration, NICE guidelines were searched to better understand whether abemaciclib, palbociclib, ribociclib or all of them

should be searched for. In addition, a focused bibliographic database search was run to retrieve the latest systematic reviews of CDK4/6 inhibitor efficacy in first-line treatment of metastatic breast cancer. Details of these searches are summarised in Table 118.

Table 118. Iterative search: 1st iteration

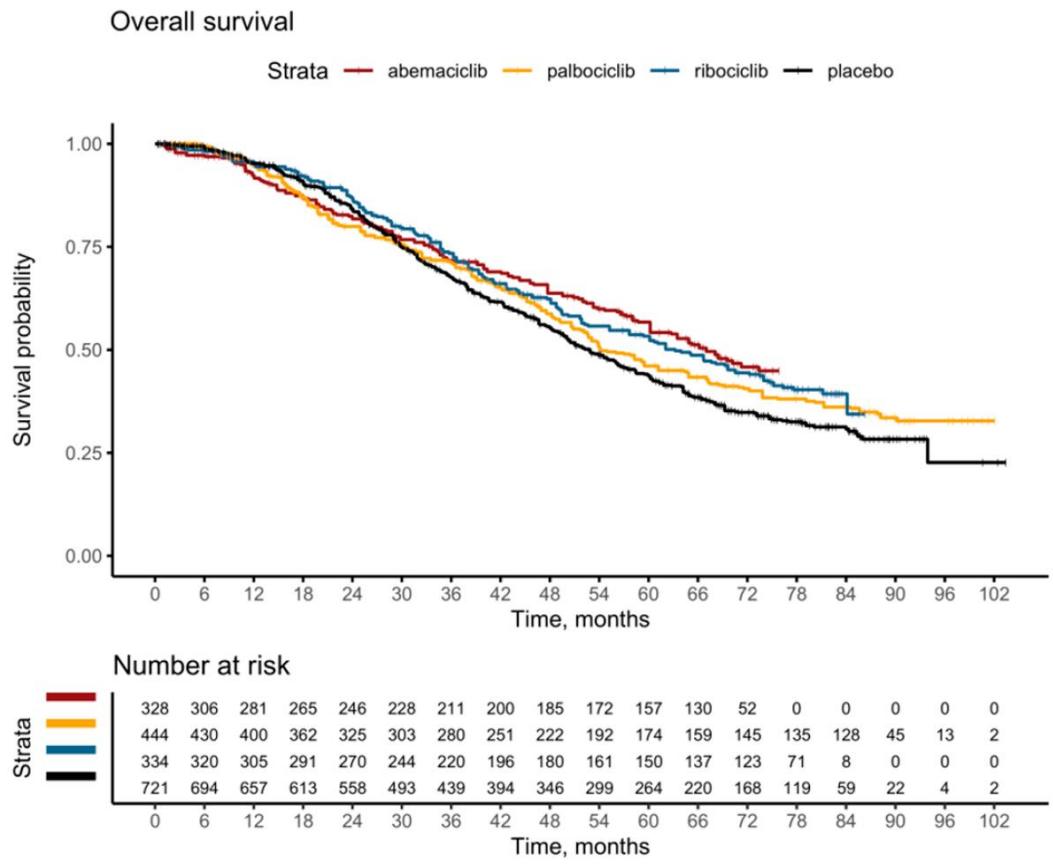
Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title/abstract only, time limit to applied and only systematic reviews/network meta-analyses searched
Type of studies	Systematic reviews only
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit: 1st July 2023 – 1st July 2024 (1 year)
Terms used	Metastatic breast cancer ((ab(breast cancer))) AND HR+ and HER2- ((ab(hormone receptor positive) OR ab(HR+))) AND ((ab(human epidermal growth receptor 2 negative) OR ab(HER2-))) AND Publication type ((ab(literature review)))
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	Identify whether any NICE guidelines on advanced breast cancer exist, to better understand which CDK 4/6 inhibitor is considered standard of care and should be searched for
Source names	NICE guidance: https://www.nice.org.uk/guidance/conditions-and-diseases/cancer/breast-cancer/products?GuidanceProgramme=guidelines
Search dates	June 30 2024
Limits	Not applicable

A review of the NICE website revealed that Clinical Guideline 81 (CG81) was published in 2009, and it was last updated on 16th August 2017.[460] CG81 does not include guidance on CDK4/6 inhibitors, most likely because abemaciclib, palbociclib, ribociclib have been approved in Europe after the last update. No detailed NICE submissions were identified, and therefore a bibliographic database search of systematic reviews in the target population was designed.

The focused bibliographic search was carried out on 1st July 2024. References were collected in a bibliographic management database (EndNote). A total of 29 publications were identified and of those 5 were selected for full-text review. One full-text study was selected for inclusion. This was a systematic review and indirect treatment comparison by Zhao *et al*. 2023.[461] The publication reported an indirect treatment comparison based on PALOMA-2, MONALEESA-2 and MONARCH-3 studies, including a total of 1,827 patients.[462-464] It included the same treatment and patient population that was used in the usual practice search: first-line treatment with ribociclib plus letrozole of post-menopausal patients with HR+/HER2- metastatic breast cancer.

Figure 33 displays the OS resulting from the indirect treatment comparison.

Figure 33. Overall survival of first-line CDK4/6 inhibitors in post-menopausal patients with HR+/HER2- metastatic breast cancer (Zhao *et al*. 2023[461])



Usual practice identified Suri *et al*. which was not identified in this iteration 1 search. Suri *et al*. is a health economic model, and iteration 1 did not search for health economic models, but rather reviews of OS in the target population. Suri *et al*. reported shape and scale parameters for Weibull distribution as well as HRs from an indirect treatment comparison, allowing the external assessors to re-build the survival model

and estimate mean OS.[465] Mean survival can be estimated from models, and it can also be estimated directly from Kaplan-Meier plots frequently reported in publications of clinical studies (although this would be a restricted mean if there is any administrative censoring so it would downwardly biased), as well as network meta-analyses. The reason why iteration 1 focused on recent systematic reviews and network meta-analyses is that there is a wealth of information available for CDK4/6 inhibitors, and pragmatic solution was required to search for the key evidence efficiently. Iteration 1 of the iterative search identified Zhao *et al*. that reports Kaplan-Meier estimates from the indirect treatment comparison for abemaciclib, palbociclib, ribociclib and placebo. These data can be used to model OS for any of these treatments, including ribociclib + aromatase inhibitor, the group that was chosen in usual practice to represent standard of care. This data could be used to derive mean OS for use in this risk stratification case study model to inform the 6-month probability of death from DM input. An overview of studies reporting OS is shown in Table 119.

Table 119. Overview of sources considered for OS

Study	Usual Practice (n=1)	Iteration 1 (n=1)
Suri *et al*.[458]	✓	
Zhao *et al*.[461]		✓

Baseline Risk Values Used in the Model from Iteration 1

In iteration 1, only one publication was included to be used in the model. This was a publication reporting an indirect treatment comparison results from three pivotal clinical studies. This publication was not identified by the usual practice search, although it is unclear if it might have been if the search could have been updated. The usual practice search did not contain any recorded steps for searching and therefore it could not be updated to current date. Similarly, iteration 1 of the iterative search did not identify the health economic model publication used from the usual practice search. The mean OS was estimated from the Kaplan-Meier data by using WebPlotDigitizer and R and fitting a Weibull model (Table 120).[466,467]

Table 120. Health state utility values used in the model - Iterative search, iteration 1

	Mean overall survival
Suri *et al*.[458]	4.63
Zhao *et al*.[461]	6.31

Health Economic Model Results

The model was run with estimated mean OS from Zhao *et al* (Table 121). Using Zhao *et al*. as a source resulted in very similar model results as using Suri *et al*. as a source.

Table 121. Health economic model results Oncotype Dx versus current decision-making – Iteration 1

	ICER (£)
Suri *et al*.[458]	Dominating
Zhao *et al*.[461]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Iteration 2

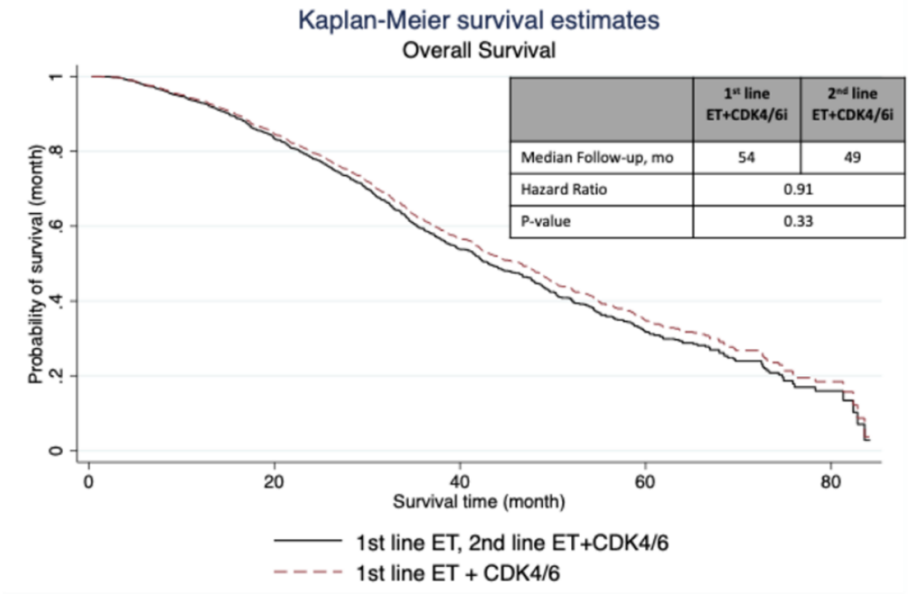
Iteration 1 gave an overview of the evidence base available for OS for CDK4/6 inhibitors. Iteration 2 expands the search for OS data beyond literature reviews to other publication types. A focused bibliographic database search for the last year was carried out to retrieve the very latest evidence. The intention was that if no relevant evidence is identified, the search can be expanded to cover further years as well as topics in Iteration 3. The search details are provided in Table 122.

Table 122. Iterative search: 2nd iteration

Bibliographic database search elements	Considerations
Sampling	Focused sampling in the words appearing in the title only.
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	Time limit 1 year (July 2023 – July 2024)
Terms used	Metastatic breast cancer ((ab(breast cancer))) AND (ab(metastatic)) HR+ and HER2- AND ((ab(hormone receptor positive) OR ab(HR+))) AND ((ab(human epidermal growth-factor receptor-2 negative) OR ab(HER2-))) CDK4/6i AND (ab(CDK4/6) OR ab(CDK 4/6)) Overall survival AND ab(overall survival)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	None
Search dates	None
Limits	None

Forty-six citations were sifted at title/abstract level. Five full-texts were reviewed. Of these, one publication by Kimmick *et al*. 2024 was included.[468] The aim of this identified study was to compare CDK4/6 inhibitors with endocrine therapy (ET) in the first- versus second-line setting for treatment of HR+, HER2-metastatic breast cancer using real-world evidence from Flatiron database. The analysis included 2,170 1st-line CDK4/6i patients, and reported a median OS of 54 months. Kaplan-Meier plots are also reported (Figure 34). These data could be used to derive mean OS for use in this risk stratification case study model to inform the 6-month probability of death from DM input.

Figure 34. Overall survival of first-line CDK4/6 inhibitors in post-menopausal patients with HR+/HER2- metastatic breast cancer (Kimmick *et al*. 2024[468])



Similar to iteration 1, Suri *et al*. from usual practice search was not identified in this iteration 2. In iterative search (1 + 2), two publications were identified, Zhao *et al*. and Kimmick *et al*. Both had Kaplan-Meier estimates available to perform further modelling. [461,468] Table 123 provides an overview of the OS sources considered.

Table 123. Overview of sources considered for OS

Study	Usual Practice (n=1)	Iteration 1 + 2 (n=2)
Suri *et al*.[458]	✓	
Zhao *et al*.[461]		✓
Kimmick *et al*.[468]		✓

Baseline Risk Values Used in the Model from Iteration 2

In iteration 2, a further publication by Kimmick *et al* was included.[468] The estimated mean OS was is shown in Table 124.

Table 124. Health state utility values used in the model - Iterative search, iteration 1

	Mean overall survival
Suri *et al*.[458]	4.63
Zhao *et al*.[461]	6.31
Kimmick *et al*.[468]	4.15

Health Economic Model Results

The model was run with estimated mean OS from Kimmick *et al*. Using Kimmick *et al* as a source resulted in very similar results as using Suri *et al*. and Zhao *et al*. as a source (Table 125).

Table 125. Health economic model results Oncotype Dx versus current decision-making – Iteration 1

	ICER (£)
Suri *et al*.[458]	Dominating
Zhao *et al*.[461]	Dominating
Kimmick *et al*.[468]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Decision to Stop Searching

Iteration 1 identified an indirect treatment comparison based on clinical trial data, that reported OS Kaplan Meier estimates for CDK4/6 inhibitors. Iteration 2 identified a further publication that reported real-world OS Kaplan Meier estimates for CDK4/6 inhibitors. Both publications also reported median OS. However, for quantification of economic benefits it is necessary to estimate the mean OS, instead of median time. The Kaplan-Meier function allows the mean to be estimated, although the step is out of scope of this this, as it was with OS and PFS in the thyroid cancer case study. Zhao *et al*. could be used in the base case analysis, and the real-world evidence from Kimmick *et al*. could be used as a sensitivity analysis, providing a range of feasible results from both types of data.

Given that the iterative search had identified two, recent publications that both reported relevant data in the right format (Kaplan-Meier) for the target population (post-menopausal patients with HR+/HER2- metastatic breast cancer), it was decided that searching would be stopped. Having one OS estimate from indirect treatment comparison based on clinical trials, plus one OS estimate from a large real-world study, allows an estimation of range of results that could be expected from the health economic model.

Summary of Iterative Search Results and Efficiency

In total, 75 studies were reviewed. In the end, two studies were included in the report and two in the model from this iterative searching. Summary of the iterative searching is presented in Figure 35. The iterative search output had a precision of 2.67% both in terms of items cited in the report and used in the model. The NNR was therefore 38. Table 126 displays the estimated search effort, separately for iteration 1 and 2, as well as the total. The development of the search protocol, running of the searches, citation identification, and publication screening, as well as the web searches for registries in the two iterations took 190 minutes (3 hours and 10 minutes).

Figure 35. Summary of results – Iterative search

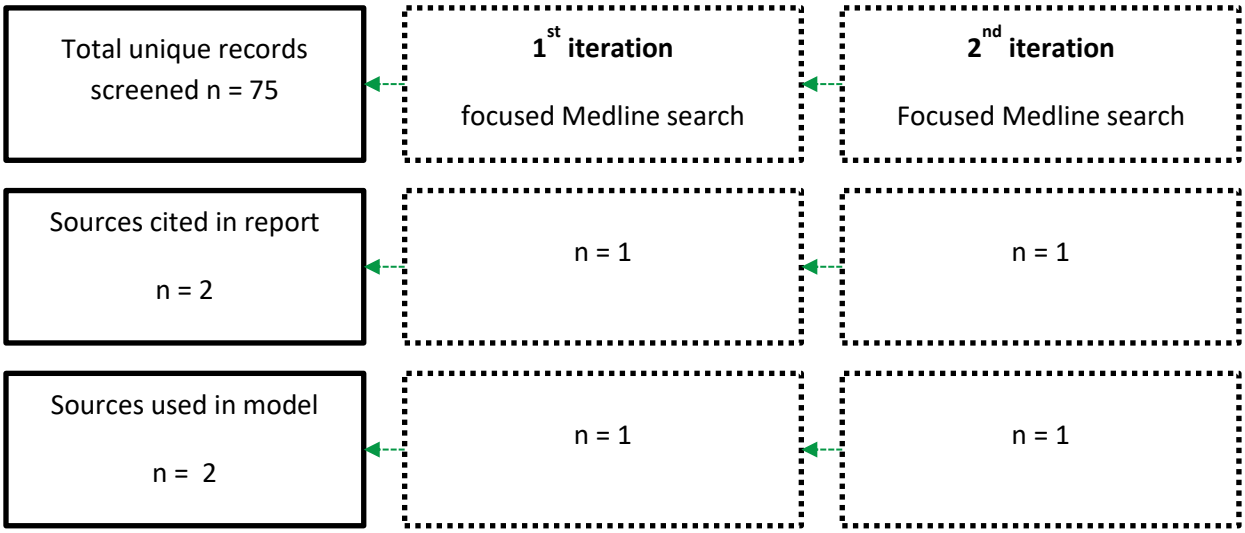


Table 126. Search tracker – iterative search

Item	Original and Update Search							
	Number of studies identified			Time in minutes			Precision	NNR
	Iteration 1	Iteration 2	Total	Iteration 1	Iteration 2	Total		
Search protocol development				15	15	30		
Running searches				10	10	20		
Identified citations	29	46	75			0		
Duplicate removal	0	0	0	0	0	0		
Title/abstract level	29	46	75	20	40	60		
Citations excluded	24	41	65			0		
Full papers to retrieve + review	5	5	10	40	40	80		
Sources cited in the report	1	1	2				2.67%	38
Sources used in model	1	1	2				2.67%	38
TOTAL				**85**	**105**	**190**		

Abbreviations: NNR: number needed to read.

7.4.3 Alternative Search Method 2

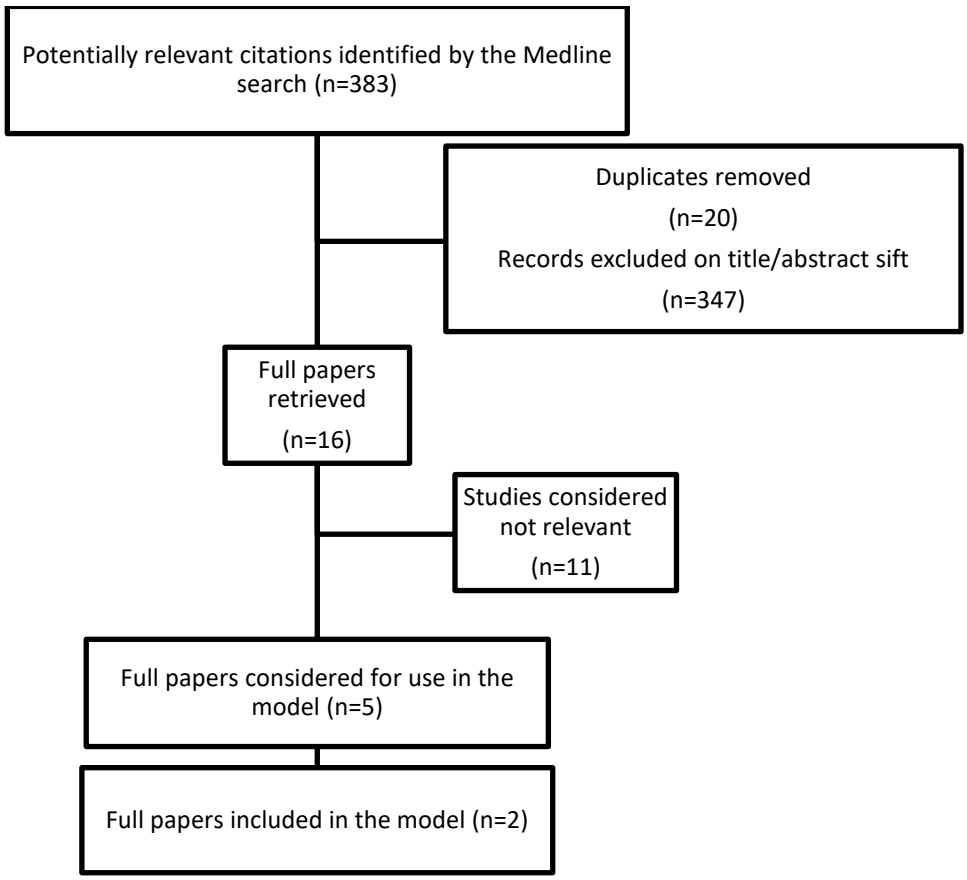
Rapid Review

For this rapid review, Medline (via ProQuest Dialog) was searched for the last 20 years. The search strategy is given below in Table 127. The rapid review limitations included using free text search and searching in titles/abstract. In the previous case studies only titles were searched, and not abstracts, but in this case study this resulted in too few findings that were not relevant enough. This literature search was conducted in July 2024. References were collected in a bibliographic management database (EndNote). The results are summarised in Figure 36.

Table 127. Breast cancer tumour profiling risk stratification case study: Alternative search – Rapid Review

Bibliographic database search elements	Considerations
Sampling	Focused sampling, with only one database, restricted synonyms, limiting the search to titles only
Type of studies	No limitation
Sources	MEDLINE (via ProQuest Dialog)
Limits	20 years i.e., Jan 2004 – May 2024 Not a case study
Terms used	Metastatic breast cancer ((ab(breast cancer))) AND (ab(metastatic)) CDK4/6i AND (ab(CDK4/6) OR ab(CDK 4/6)) Overall survival AND ab(overall survival)
Conceptual limitations	No conceptual limit
Non-database search elements	**Considerations**
Approaches	None
Source names	Not applicable
Search dates	Not applicable
Limits	Not applicable

Figure 36. Study selection results – Rapid review



The search identified five potential sources to be used in the model for probability of death due to DM.[461,468-471] Two of them were identified in the iterative search.[461,468] Three were unique to the rapid review.[469-471] None were identified in the usual practice and the usual practice source was not identified in either the iterative search nor the rapid review.

Three of the studies were analyses of registry data; the Flatiron Health database, the Survey Epidemiology and End Results (SEER)-Medicare database, SONABRE Registry and data collected from Instituto Portugues de Onkologia – Porto (Portugal).[468-471] OS Kaplan-Meier estimates were reported (Figure 34, Figure 37, Figure 38, Figure 39). The Kimmick *et al*. analysis of Flatiron data is based on a large sample size (n=2,170) and very recent (index date up to November 2021). The SEER-Medicare database analysis was two years older (index data up to 2019) and included less patients who received CDK4/6 inhibitors (n=169). The Dutch SONABRE registry provides OS by period of diagnosis, including a recent time period of 2017 – 2019. OS from this time period can be considered to be reflective of OS resulting from current treatment practice i.e., use of CDK4/6 inhibitors. There were n=493 patients at risk in the beginning of the time period for the 2017 – 2019 group. Coutinho-Almeida *et al*. is a recent analysis of Portuguese data on efficacy of CDK4/6 inhibitors (palbociclib n=246, ribociclib n=106). Both of these studies reported data in the right format for the relevant target population.

The remaining study was a systematic review with indirect treatment comparison.[461] Zhao *et al*. had also been identified in iterative search, and included OS Kaplan-Meier plots (Figure 33).

The results of the searches are summarised in Table 128.

Figure 37. Overall survival reported in Goyal *et al*. 2022 (Figure 2A)

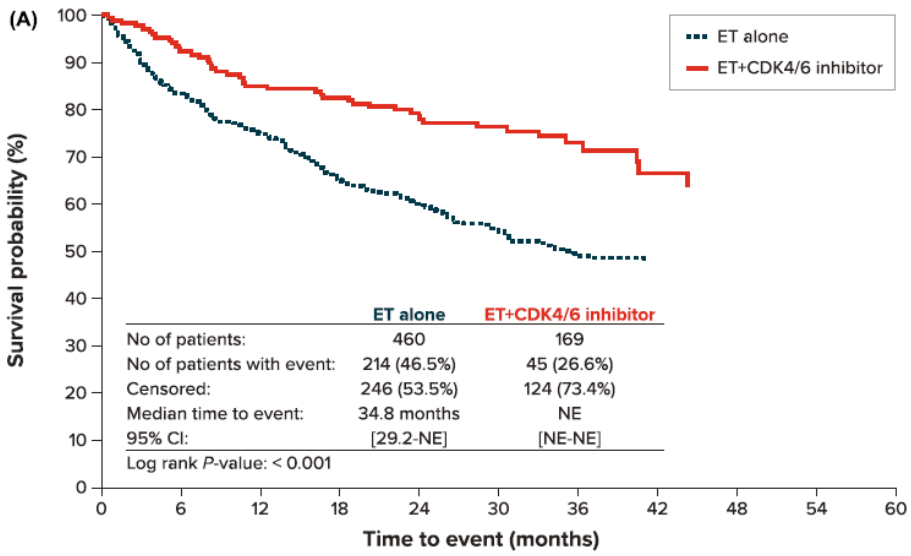


Figure 38. Overall survival reported in Meegdes *et al*. 2023 (Figure 1)

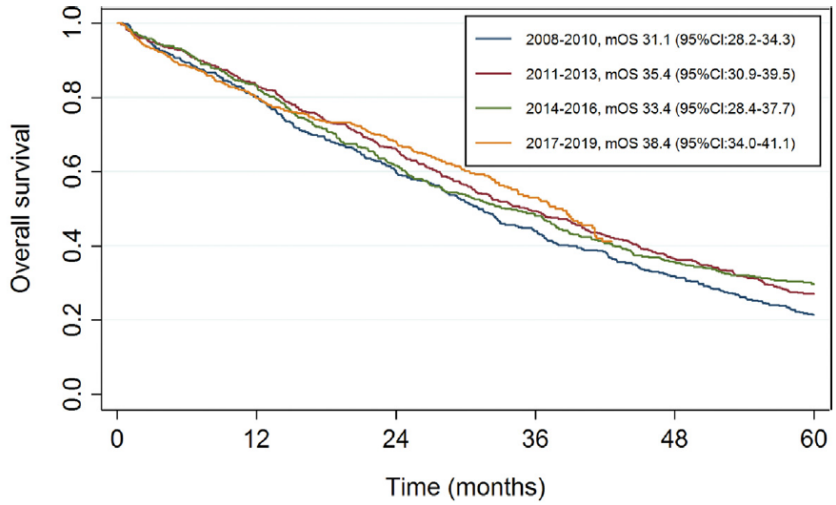


Figure 39. Overall survival reported in Coutinho-Almeida *et al*. 2024 (Figure 1)

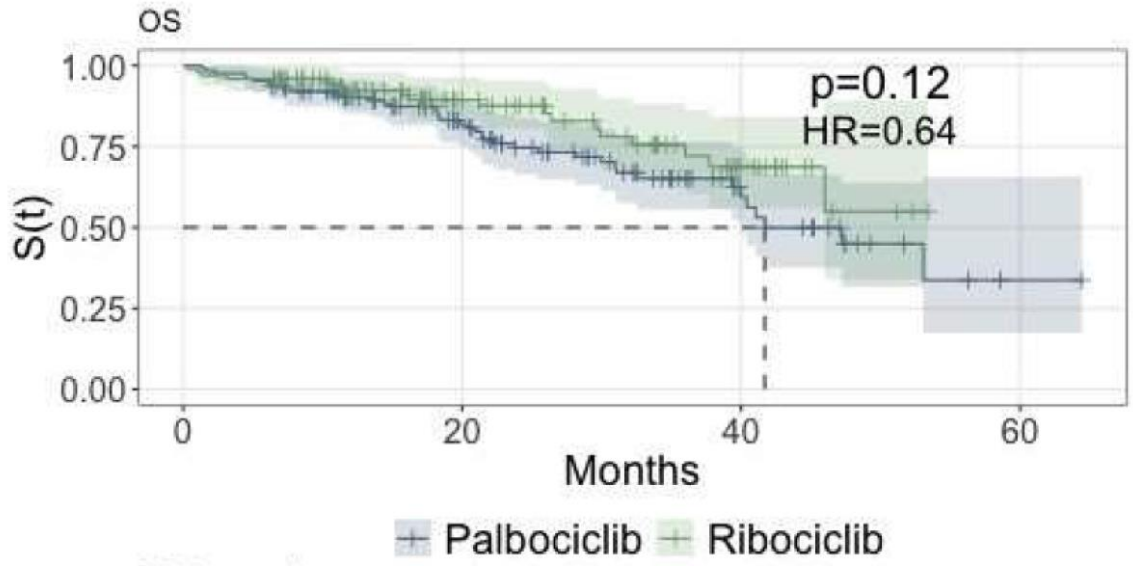


Table 128. Overview of sources considered for OS

Study	Usual Practice (n=1)	Iterative search (n=2)	Rapid Review (n=5)
Suri *et al*.[458]	✓		
Zhao *et al*.[461]		✓	✓
Kimmick *et al*.[468]		✓	✓
Coutinho-Almeida *et al*.[469]			✓
Goyal *et al*.[471]			✓
Meegdes *et al*.[470]			✓

Baseline Risk Values Used in the Model from Rapid Review

In rapid review three further sources were identified and the mean OS was estimated. The estimated mean OS values are shown in Table 129.

Table 129. Health state utility values used in the model – Rapid review

	Mean overall survival
Suri *et al*.[458]	4.63
Zhao *et al*.[461]	6.31
Kimmick *et al*.[468]	4.15
Coutinho-Almeida *et al*.[469]	7.04
Goyal *et al*.[471]	9.00
Meegdes *et al*.[470]	4.12

Health Economic Model Results

The model was run with estimated mean OS from the three additional publications (Table 130). Using any of the identified publications, the model results were almost unchanged.

Table 130. Health economic model results Oncotype Dx versus current decision-making – Iteration 1

	ICER (£)
Suri *et al*.[458]	Dominating
Zhao *et al*.[461]	Dominating
Kimmick *et al*.[468]	Dominating
Coutinho-Almeida *et al*.[469]	Dominating
Goyal *et al*.[471]	Dominating
Meegdes *et al*.[470]	Dominating

Abbreviations: ICER: incremental cost-effectiveness ratio.

Summary of Rapid Review Results and Efficiency

Table 130 displays the estimated effort to identify model inputs. In the focused Medline search, 383 studies were identified, of which two were used in the model. This effort was estimated to take 460 minutes. The search output had a precision of 1.31% in terms of items cited in the report and 0.52% for items used in the model. The NNR was therefore 77 and 192 for sources cited in the report and sources used in the model, respectively.

Table 131. Search tracker – Rapid review

Item	Number	Time (mins)	Precision	NNR
Search protocol development		20		
Running searches		15		
Identified citations	383			
Duplicate removal	20	20		
Citations to sift at title/abstract level	363	290		
Citations excluded	347			
Full papers to retrieve + review	16	110		
Sources cited in the report	5		1.31%	77
Sources used in the model	2		0.52%	192
TOTAL		**460**		

Abbreviations: NNR: number needed to read

7.4.4 Case Study Baseline Risk of Clinical Events in Tumour Profiling – Summary

The key observations from the baseline risk of clinical events case study were:

- The included model input was the risk of death from DM.
- From usual practice, one cost-effectiveness publication was included. A published model was rebuilt using published information that allowed the modellers to derive the 6-month death probability from DM. The detailed steps of how it was identified or selected were not described in the EAG report.[458]
- It was possible to obtain further information about the usual practice search from the health economist working on the external assessment, who was also a supervisor for this thesis (PT). From this information, it is estimated that the usual practice search took approximately 7 hours and 40 minutes. It was not possible to repeat the steps to identify the source, as PT could not recall them in detail. Therefore, unlike with other case studies, no update to the usual practice was done. It is possible that some of the publications identified by the alternative search methods could have been identified by usual practice search, if it had been possible to update it.
- Iterative searching identified two sources, neither of which was the same as the source identified in usual practice.[461,468] This is likely to be because the usual practice source is a health economic model whereas iterative search used search terms related to breast cancer and literature review (iteration 1) and OS (iteration 2). The two sources identified as part of iterative searching were an indirect treatment comparison and a real-world study, reporting Kaplan Meier OS estimates in the target population.[461,468]

- Rapid review identified the most included sources (five), with some overlap to iterative search findings (two of the five). Three sources unique to rapid review were identified.
- Mean OS was estimated using the newly identified Kaplan-Meier plots. The model results remained consistent with all the identified sources.
- In terms of effort, the usual practice search took longer than the iterative search and about the same as rapid review (Table 132).
- The precision and NNR were 38 and 192 for iterative search and rapid review, respectively, for sources used in the model. Precision and NNR are not known for usual practice as only one publication was reported, with no information of those that were reviewed but discarded.
- Iterative searching had a higher precision and lower NNR than rapid review (Table 132).

Table 132. Breast cancer tumour profiling risk stratification case study: Summary of usual practice and experimental search methods

	Sources: report	Sources: model	Precision Report	Precision Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
Usual Practice search	1	1	100.00%	100.00%	1	1	480	-
Iterative search	2	2	2.67%	2.67%	38	38	190	-60%
Rapid review	5	2	1.31%	0.52%	77	192	460	-4%

Abbreviations: NNR: Number needed to read. **Footnote:** [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

7.5 Case Study in Breast Cancer Tumour Profiling Risk Stratification – Summary

The key observations from the breast cancer tumour profiling risk stratification case study were:

- Iterative searching has emerged as a promising approach for both utility and baseline risk of clinical event inputs. In this case study, it offered an advantage in terms of speed compared to usual practice and rapid review. Moreover, the most relevant sources were identified, and the model outputs obtained through iterative searching remain comparable to those achieved using usual practice. Rapid review took longer than either iterative searching or usual practice, and identified fewer references. However, the impact on the model results was minor.
- Rapid review took longer than usual practice or iterative searching for utility input identification, and about the same as usual practice for baseline risk of clinical event input identification. While rapid review took longer than usual practice for the utility input identification, it also covered all five health states whereas usual practice only covered two health states. Therefore, rapid review was associated with increased transparency over usual practice. For the baseline risk of clinical events input, rapid review identified the most studies; however, the value of identifying more than one or two studies is likely to be limited, given that the model results were consistent across the values tested.

8 Discussion of Case Study Results

8.1 Chapter Overview

This chapter provides a discussion of the case study results. Section 8.2 summarises iterative search findings and Section 8.3 summarises rapid review findings. Section 8.4 gives details of stopping searching. Section 8.5 provides my personal reflection on the factors impacting the results, sources of bias and broader issues. Section 8.6 Is a reflection on the reporting tools adopted/developed for this PhD study.

8.2 Summary of Results: Iterative Searching

8.2.1 Methods Guidance Available in Literature

In contrast to systematic reviews, there is no common definition or method with practical steps describing how to carry out iterative searching. A publication by Zwakman *et al*. was helpful when formulating the methods for this PhD study (Section 4.2.2.2).[472] This publication described PALETTE, an iterative method for conducting a literature search for a review in palliative care. This publication provides a practical guidance for searching one patch at the time, and the ability to abandon the patch, as judged relevant by the researcher. Bates *et al*. and Schlosser *et al*. have described a general approach for searching literature that is consistent with expansive searching.[20,473] Searching literature in an expansive manner allows search strategies to emerge as the research investigation takes shape. Adjusting and readjusting search methods aims to ensure that data collection efforts produce meaningful results, rather than repetitive data. Information behaviours that are related to iterative searching have been described in earlier parts of this thesis (Sections 2.4 and 4.2.2.2).

8.2.2 Methods as Employed During This Study

Iterative searching was carried out six times in this PhD study: twice for each of the three case studies. Three of the searches were designed to identify health state utility inputs, and the other three were designed to identify baseline risk of clinical event inputs. For each of the iterative searches, an initial information retrieval need was defined based on the health economic model information need. An information retrieval plan was constructed and recorded using the draft search framework for each of the searches. One to two search iterations were run, depending on the case study. The rationale for ending searching, as defined in the methods (Section 4.2.2.2), was recorded. In all case studies, iterative searches resulted in similar health economic model results to usual practice.

8.2.2.1 Health State Utility

In all three case studies (UC, thyroid cancer and breast cancer risk stratification), the first iteration of the health state utility search included an initial HTA document review, as well as an associated, focused bibliographic database search. In the first iteration, different proportions of the studies identified by the usual practice search were found: 29%, 100% and 100% in the UC, thyroid cancer and breast cancer risk

stratification case studies, respectively. Iterative searching identified new studies (compared to usual practice search) in two of the case studies: UC (n=5) and breast cancer risk stratification (n=1). No new studies were identified in the thyroid cancer case study. This is likely due to the fact that the type of thyroid cancer that was the focus of the assessment (i.e., MTC) is rare and the evidence base is limited. This also served as a rationale to stop searching for further health state utility values in thyroid cancer. Therefore, only one search iteration was performed in the thyroid cancer case study.

Searching was continued to the second iteration in the UC case study, as there was uncertainty about whether all relevant input data had been identified. This may, at least in part, be due to iterative searching not being an established search method for health economic input identification. If evidence existed on the usefulness of iterative searching in the context health economic model input identification, I may have had more confidence to stop after this first iteration, as many inputs had already been identified. The second iteration identified some additional studies; the percentage of usual search practice inputs identified increased from 29% to 43% and two new unique studies were also identified. However, these new inputs did not have a marked impact on the model conclusions. The additional sources did not replace the base case source. Rather, they were used for additional scenarios that returned similar results to an already known scenario source, where Swinburn provided alternative utility values. Since there was no change to the base case and the scenario result results remained similar,searching was stopped after the second iteration.

A second iteration was also run for the breast cancer risk stratification case study. The results from the first iteration did not find inputs for all the health states, and therefore a second iteration was run that was aimed at finding model inputs specifically for those missing health states. Two further inputs were identified in the second iteration. In both UC and breast cancer case studies, two additional relevant studies were identified in iteration 2, compared to iteration 1. These did not make any difference to the model conclusion that could be made, so a decision was made to stop searching after iteration 2.

Table 133 summarises the details of iterative health state utility searches.

Table 133. Health state utility Iterative search

Case study	Ulcerative colitis	Thyroid cancer	Breast cancer risk stratification
Iteration 1			
Bibliographic search	Yes	Yes	Yes
HTA document review	Yes	Yes	Yes
UP inputs identified	8/28 (29%)	4/4 (100%)	7/7 (100%)
Number of new inputs*	5	0	1
HE model results similar to UP?	Yes	Yes	Yes
Modelled uncertainty similar to UP?	Yes	Yes	Yes
Decision to stop	No – Uncertainty associated with iterative searching as method for model input identification	Yes – Rare disease, unlikely to find further relevant evidence	Partially – Less evidenced uncovered for two of the health states
Iteration 2			
Bibliographic search	Yes	N/A	Yes
HTA document review	No	N/A	No
UP inputs identified	12/28 (43%)	N/A	7/7 (100%)
Number of new inputs*	7	N/A	3
Model results similar to UP/iteration 1?	Yes	N/A	Yes
Model uncertainty similar to UP/iteration 1?	Yes	N/A	Yes
Decision to stop	Yes – No further relevant evidence expected (marginal relevance)	N/A	Yes – No further relevant evidence expected (marginal relevance)

Footnote: *Number of inputs not identified in the usual practice search. [§]Cumulative value from 1st plus 2nd iterations. **Abbreviations**: HE: health economic, N/A: not applicable, UP: usual practice

In terms of search efficiency, iterative searching was more efficient than the usual practice search in all three case studies for health state utility input identification (see Table 134). The reduction in time needed to perform the searching relative to usual practice ranged from 89% to 39%.

Table 134. Search efficiency in iterative searching: Health state utility

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
UC	19	6	23.17%	7.32%	4	14	380	-74%
Thyroid cancer	5	4	29.41%	23.53%	3	4	130	-89%
Breast cancer risk stratification	10	6	2.99%	1.79%	34	56	530	-39%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

8.2.2.2 Baseline Risk of Clinical Events

In all three case studies (UC, thyroid cancer and breast cancer risk stratification), the first iteration of the baseline risk search only used focused bibliographic database searching (Medline only). In the first iteration, different proportions of the studies identified by the usual practice search were found: 57%, 17% and 0% in UC, thyroid cancer and breast cancer risk stratification case studies, respectively.

In the UC case study, several inputs had already been identified in the first iteration of the search. However, no UK-specific mortality data were identified. As mortality can differ between countries, it is important that this model input is using local data, when possible. Therefore, a second iteration was run in an attempt to identify UK-specific mortality sources. The second iteration identified two further studies: one had been identified in the first iteration and the other one was new. This iteration increased the percentage of overlap between the usual practice search inputs and iterative search inputs from 57% to 64%. The total number of unique baseline risk of clinical event inputs identified in the UC iterative search was 10. The conclusions from the health economic model did not differ from usual practice after running the model with the new data. No further search iterations were conducted, as likelihood of retrieving further relevant evidence was deemed to be low.

In the thyroid cancer case study, the title/abstract review was stopped early due to lack of relevant data in the first search iteration. There were several publications that had potential to inform the model, but the publication either summarised the data in a way that did not allow it to be included in the health economic model, or the population that it was reported for did not match the target population. As a validatory step, the remaining titles/abstracts were reviewed and it could be confirmed that the discarded patch did not include any relevant studies. The “close-but-not-quite” from the first patch showed that several registries exist that potentially could contain relevant data to inform the model. Therefore, the second iteration was a registry search to identify and list those registries that have potential to contain relevant OS and OFS data in thyroid cancer. Several registries were identified, but these could not be

accessed as they often require a subscription and may be associated with a fee. Therefore, the identified sources were listed but not accessed.

The first iteration of the breast cancer risk stratification case study was a review of systematic reviews. Usual practice data source was a health economic model, and therefore it was not identified in the first (or subsequent) iteration(s) of the search. The first iteration identified one alternative data source that could be used in the model. Searching continued in a second iteration, without restrictions by study type. Two further studies were identified. When the newly identified data were included in the health economic model, it was shown that the model conclusions did not change significantly. Therefore, searching was stopped after the second iteration.

Table 135 summarises the details of iterative baseline risk of clinical event searches.

Table 135. Baseline risk of clinical event Iterative search

Case study	Ulcerative colitis	Thyroid cancer	Breast cancer risk stratification
Iteration 1			
Bibliographic search	Yes	Yes – Selection stopped early	Yes
HTA document review	No	No	No
UP inputs identified§	8/14 (57%)	1/6 (17%)	0/1 (0%)
Number of new inputs*	9	1	1
Model results similar to UP?	Yes	Model not run	Yes
Modelled uncertainty similar to UP?	Yes	Model not run	Yes
Decision to stop	Partially – No UK specific mortality data identified	No – Change of tactic to registry search	No - This iteration focused on reviews only
Iteration 2			
Bibliographic search	Yes	No	Yes
HTA document review	No	No	No
Registry search	Yes	Yes	No
UP inputs identified§	9/14 (64%)	N/A	0/1 (0%)
Number of new inputs*§	10	N/A	3
Model results similar to UP/iteration 1?	Yes	Model not run	Yes
Model uncertainty similar to UP/iteration 1?	Yes	Model not run	Yes
Decision to stop	Yes – No further relevant evidence expected	Yes – Several potentially relevant registries identified	Yes – No further relevant evidence expected

Footnote: *Number of inputs not identified in the usual practice search. §Cumulative value from 1st plus 2nd iterations. **Abbreviations**: HE: health economic, N/A: not applicable, UP: usual practice

In terms of search efficiency, iterative searching was more efficient than usual practice search in two of the case studies for baseline risk of clinical event input identification (see Table 136). The reduction in time needed to perform the searching was 86% and 60% for the UC and breast cancer risk stratification case studies, respectively. Unlike the other health state utility or baseline risk of clinical event iterative searches, the search effort increased in the UC baseline risk of clinical events search. This is because in the usual practice search, only one of the baseline risk parameters was associated with a description of information retrieval steps. For the other four, no details were provided on how they came to be incorporated in the model. Iterative searching (as well as and rapid review reported in the Section 8.3) recorded information retrieval steps for all five model inputs, therefore increasing transparency and inevitably taking longer.

Table 136. Iterative searching: Baseline risk of clinical event

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
UC	17	6	3.28%	1.35%	31	74	460	+90%
Thyroid cancer	2	1	1.53%	0.76%	66	131	500	-86%
Breast cancer risk stratification	2	2	2.67%	2.67%	38	38	190	-60%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search tim4

8.3 Summary of Results: Rapid Review

8.3.1 Methods Guidance Available in Literature

Rapid reviews aim to expedite the process of evidence synthesis compared to full systematic reviews. These reviews employ various methodological shortcuts to achieve this goal. Multiple publications consistently highlight the absence of a universally agreed definition and methodology for rapid reviews.[474-478] Definitions proposed include both elements of a shorter time frame as well as a reduced scope.[475,477] Some common shortcuts, related to the searching part of rapid reviews (in line with the scope of this thesis), include narrowing the scope of the search, searching fewer databases, minimising reliance on grey literature, and including only specific types of studies (e.g., English-language only or recent studies). Additionally, rapid reviews often rely on existing systematic reviews, reduce manual hand searching of reference lists and relevant journals, and streamline the time frame for article retrieval. Despite these approaches, there are no firm guidelines for searching for rapid review methodology. Existing literature acknowledges the complexity of this area and highlights the need for empirical evidence to better understand the impact of these shortcuts.[474,475,477] Only two studies have been identified that compare systematic literature review and rapid review.[479,480] The studies concluded that the core conclusions of the rapid and full reviews did not differ extensively. However, Haby *et al*. report having quality assessed these comparisons.[481] As a result, the authors had concerns about the robustness of the systematic

reviews included in these studies, and encouraged the findings to be viewed with caution.[481] Limited studies comparing full and rapid reviews make it challenging to assess the impact of the rapid review methods. One study suggests that rapid reviews may enhance the clarity and accessibility of research evidence for decision makers.[482]

A literature review of rapid reviews by Haby *et al*. highlights the lack of definition, defined methods, and research evidence regarding the implications of methodological choices for both rapid reviews and systematic reviews.[481] The authors work shows that rapid reviews can be of high quality when well conducted. Published literature suggests that, rather than focusing solely on developing a formalised rapid review methodology, which may not always be appropriate, researchers and users should emphasise increasing the transparency of reporting the methods used in each review.[474,477]

8.3.2 Methods as Employed During This Study

Rapid review methods were employed six times in this PhD study: twice for each of the three case studies. Three of the rapid reviews were designed to identify health state utility inputs, and the other three were designed to identify baseline risk of clinical event inputs. In all of the rapid reviews, the method was similar to systematic literature review in that the search was only run once using a pre-defined search protocol. Each rapid review was associated with limits, or "shortcuts".

8.3.2.1 Health State Utility

The rapid reviews conducted across the three case studies that were aimed at retrieving utility inputs, were similar in all case studies but not identical (Table 137). In all case studies, the search was conducted in Medline only. The initial numbers of citations identified varied; 129, 1,190 and 776 in the UC, thyroid cancer and breast cancer risk stratification case studies, respectively. The UC search did not have a time limit, as findings were manageable without one. Searching in thyroid cancer was more challenging, as the search had to be widened to cover more types of thyroid cancer than the rare type of interest (MTC), and also searching beyond titles and abstract was required. Searching beyond titles and abstract and limiting to the most recent 20 years was also done for the breast cancer risk stratification case study in order to ensure relevance of the results.

The proportion of the usual practice search inputs that were identified in the rapid reviews varied: 25%, 100% and 57% in the UC, thyroid cancer and breast cancer risk stratification case studies, respectively. In the thyroid cancer rapid review no further new utility inputs were identified, most likely for the same reason as with iterative search (an extremely limited evidence base in a rare disease). In the UC and breast cancer case studies, 4 and 2 new inputs were identified, respectively. All model conclusions remained similar to the usual practice search.

Table 137. Health state utility rapid reviews

Case study	Ulcerative colitis	Thyroid cancer	Breast cancer risk stratification
Database	Medline	Medline	Medline
Time limit	None	Last 10 years	Last 20 years
Free text terms	Yes	Yes	Yes
MeSH terms	Yes	Yes	Yes
Restricted synonyms	Partially	Yes	Yes
Searching in titles and abstracts only	Yes	No	Partially
Citation searching	Yes	No	No
UP inputs identified[§]	7/28 (25%)	5/5 (100%)	4/7 (57%)
Number of new inputs*[§]	4	0	2
Model results similar to UP/iteration 1?	Yes	Yes	Yes
Model uncertainty similar to UP/iteration 1?	Yes	Yes	Yes

Footnote: *Number of inputs not identified in the usual practice search. [§]Cumulative value from 1st plus 2nd iterations. **Abbreviations**: HE: health economic, N/A: not applicable, UP: usual practice

In terms of search efficiency, rapid review was more quicker than usual practice search in the UC case study, resulting in 69% reduction in search time (see Table 138). In the thyroid cancer and breast cancer risk stratification case studies, the search time was increased by 17% relative to usual practice. Due to methodological limitations associated with this study (i.e., usual practice search protocol development time and search running time are not known), time for usual practice may be underestimated. Further, rapid review in the thyroid case study covered all the health states, rather than only some, unlike usual practice. Therefore, rapid review was associated with increased transparency over usual practice. Similarly, in the breast cancer risk stratification case study, the usual practice identification and selection process was transparently recorded only for some of the inputs, whereas rapid review covered all of the health states, increasing the transparency of information retrieval. This was similar to iterative search reported in the previous Section 8.2.

Table 138. Rapid review: Health state utility

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
UC	11	2	8.53%	1.55%	11.73	64.50	440	-69%
Thyroid cancer	5	4	0.42%	0.34%	238	298	1400	17%
Breast cancer risk stratification	6	5	0.77%	0.64%	129	155	1,020	17%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

8.3.2.2 Baseline Risk of Clinical Events

The rapid reviews conducted across the three case studies employed very similar methodologies (Table 139). In all case studies, the search was conducted in Medline only and for the last 20 years. The searches included limited free text terms, and no MeSH terms. All searched in titles/abstracts only with restricted synonyms. UC and thyroid cancer case studies both identified 50% of the usual practice search inputs, but breast cancer risk stratification case study did not identify any. UC case study found a high number of new inputs (n=14), and the other two found fewer. None of these differences in the search methods made a material difference to the model results: All model conclusions remained similar.

Table 139. Baseline risk of clinical events rapid reviews

Case study	Ulcerative colitis	Thyroid cancer	Breast cancer risk stratification
Database	Medline	Medline	Medline
Time limit	Last 20 years	Last 20 years	Last 20 years
Free text terms	Yes	Yes	Yes
MeSH terms	No	No	No
Restricted synonyms	Yes	Yes	Yes
Searching in titles and abstracts only	Yes	Yes	Yes
Citation searching	No	No	No
UP inputs identified[§]	7/14 (50%)	3/6 (50%)	0/1 (0%)
Number of new inputs*[§]	14	1	5
Model results similar to UP/iteration 1?	Yes	Yes	Yes
Model uncertainty similar to UP/iteration 1?	Yes	Yes	Yes

Footnote: *Number of inputs not identified in the usual practice search. [§]Cumulative value from 1st plus 2nd iterations. **Abbreviations**: HE: health economic, N/A: not applicable, UP: usual practice

Rapid review was significantly more time consuming than usual practice in the UC case study. This is due to two factors. Firstly, the usual search practice time could not be estimated for some parts of the search (development of search protocol and running of the searches). Secondly, like iterative searching (Section 8.2), rapid review recorded information retrieval steps for all five model inputs, therefore increasing transparency and also time taken. Usual practice only recorded that for one out of five model inputs. Therefore, usual practice search was less transparent than the alternative search methods. Rapid review returned the most sources but did not provide further increases in transparency over iterative searching and took considerably longer than both usual practice and iterative searching. The thyroid cancer and breast cancer case studies were associated with decreases in search time, 84% and 4%, respectively. Table 140 reports a comparison of efficiency of rapid review in searching baseline risk of clinical events.

Table 140. Rapid review: baseline risk of clinical events

	Sources: report	Sources: model	Precision: Report	Precision: Model	NNR: Report	NNR: Model	Search time (minutes)	% Change, minutes[a]
UC	19	8	1.63%	0.68%	62	146	1,534	+539%
Thyroid cancer	5	3	1.07%	0.64%	94	156	590	-84%
Breast cancer risk stratification	5	2	1.31%	0.52%	77	192	460	-4%

Footnote: [a] percentage change is calculated as a percentage change in search time compared to usual practice i.e. (usual practice search time – alternative search time) / alternative search time.

Iterative search identified more usual practice studies than rapid review for utility inputs (in 2 of 3 case studies). In the third utility case study, iterative search and rapid review both found all the usual practice studies. For baseline risk of clinical events, iterative search found more of the usual practice studies in one case study and rapid review found more of the usual practice studies in another case study. In the third case study both methods identified the same amount of usual practice studies i.e. 0%. The proportion of usual practice studies identified through alternative search methods is shown in Table 141.

Table 141. Proportion of usual practice studies identified through alternative search methods

	Health state utility		Baseline risk of clinical events	
	Iterative search	Rapid review	Iterative search	Rapid review
UC	43%	25%	64%	50%
Thyroid cancer	100%	100%	17%	50%
Breast cancer risk stratification	100%	57%	0%	0%

Abbreviations: UC: ulcerative colitis.

8.4 Decision to Stop Searching

There is no inherent value in continuing searching unless the new data identified enriches knowledge generated by the health economic model. For health economic modelling, more is better only when it helps to better understand cost-effectiveness of the health technology, or the uncertainty associated with it. Simply having more of the same is not useful in achieving this objective, although some similar values may increase the confidence in the identified point estimate, e.g., in a meta-analysis. Continuing searching when no further relevant evidence is emerging can also have a negative impact as it increases the cost of the modelling activity and may clutter the evidence base unnecessarily. Fintgeld-Connett *et al*. discuss that when deciding whether searching should be stopped, it is important to consider whether the unidentified data might considerably change the conclusions.[483] The authors also state that ending a literature review is always a judgement call as it is not possible to know what data have not been identified. Booth *et al.* state that the key is to ensure that decisions about stopping searching should be transparently detailed in the context of each review.[484]

Stopping searching was an important consideration in this PhD study, in both rapid review and iterative searching. In rapid review, decisions needed to be made about how many databases to search, how much supplementary searching to carry out and which limits to apply (time, language). Stopping iterative searching makes stopping searching more explicit, and required justification. It also acknowledges the concept of sufficient information, and focuses on value of looking for more information. My PhD study showed that marginal relevance was an important concept when searching for health economic model inputs. Searching was most often stopped because of redundancy of additional information. One case study (thyroid cancer) was also associated with another concept: 'proxy relevance'. The type of thyroid cancer in that case study was MTC that is a rare cancer subtype, and the evidence base is extremely limited. Therefore, the searching was expanded beyond MTC to cover thyroid cancer in general, before making a decision to stop searching. Multidimensional relevance was also a concept that was important when considering stopping iterative searching. When considering the identified evidence base, it was important to take many factors into consideration, including relevance of the study population, sample size of the study, geographical coverage of the study and the format of the data reported to ensure suitability for modelling. For health state utility the following were considered important: utility measure, elicitation technique and sources of the preference weights used. This multidimensional relevance was an important concept when evaluating which identified inputs should only be reported in the model report as candidate sources, and which sources should actually be used in the health economic model. The totality of the identified evidence base was important, rather than a binary yes/no decision whether a specific source met the specified inclusion criteria. In this PhD study, the health economic models were already final versions, and evolving relevance did not come out as an important concept. However, in reality, where economic models are developed prospectively, evolving relevance might be an extremely important concept. Information retrieval needs to be able to evolve, as the health economic model concept evolves. Although not tested in this PhD study, I expect that iterative searching will be a useful technique that allows for development of the information query together with the model concept. Each case study in this thesis included a description of what led to the stopping of searching (Sections 5.3.2,

5.4.2, 6.3.2, 6.4.2, 7.3.2 and 7.4.2). Table 133 and Table 135 provide a brief overview of the rationale to stop searching for health state utilities and baseline risk of clinical events, respectively.

8.5 Reflections on Iterative and Rapid Review Search Practice

As I started this PhD study, I had limited information retrieval experience. A key challenge for me was to learn how to search, and how to implement the learnings, not only in the context of the initial reviews of existing literature, but also as the research method for my case studies. All the training that I undertook at the beginning of the PhD study taught me how to conduct comprehensive Cochrane style searches. None of the University or online courses were aimed at learning alternative search methods. After conducting the reviews of existing literature, I found that there is also a paucity of literature relating to alternative search methods, especially methods that describe detailed, practical steps. It was unclear to me how to decide what should be the search approach in the first iteration, what information to use to when deciding stopping/continuing searching, how to set limits on the searches, and so on.

Iterative searching was particularly challenging as I was not able to find empirically tested detailed instructions on how to conduct iterative searching. Therefore, I needed to partially develop these step-by-step search methods, using what I could find in the literature as well as discussing the search implementation with my supervisors. In the end, in my PhD study, iteration 1 was implemented fairly consistently across the utility case studies. The first iteration was an HT.A document review and a focused Medline search. If the iterative search proceeded to second iteration, for utility inputs, a further Medline search was run. The main rationale for searching the HTA documents for utility inputs was that often the most relevant utility data comes from the pivotal study (e.g., EQ-5D collected during the trial). These *post hoc* analyses are needed to estimate the utility values required for health economic modelling, and these are usually not published in scientific journals i.e., they are not identifiable through bibliographic database searches. For this reason, it was deemed important to retrieve relevant HTA documents. A further rationale for retrieving the HTA documents was that they often contain reviews of existing literature. The first iteration was complemented with a focused Medline search, that was based on the literature review reported in the latest HTA documents retrieved. In all utility case studies, HTA document searched proved to be very effective and no further marginally relevant data was identified after one or two iterations.

HTA documents were not deemed as important for the baseline risk of clinical events as they were for utility inputs. The baseline risk inputs were heterogeneous across the case studies, including surgery-related inputs, PFS, OS, and death due to metastatic disease. These inputs were not treatment-specific, and should be estimated for standard of care, ideally over long period of time. Pivotal trials are unlikely to be the most, or the only, useful sources of data. Usually pivotal trials lack long-term follow-up at the time of the HTA submission, that is ideally needed for these inputs. Time-to-event data can be extrapolated beyond the clinical trial duration, allowing modelling of long term outcomes for the patients. However, survival data extrapolation is not without uncertainty, and therefore other observational data sources would also be of interest even in presence of data extrapolation from a clinical trial. For two of the baseline risk of clinical event iterative searches, a focused review of systematic reviews was conducted. This allowed for an efficient understanding of potential data sources. This tactic could not be

implemented in the MTC case study, as the disease is extremely rare and no useful systematic reviews were identified. A focused Medline search was run instead. The second iterations across the case studies involved running further focused Medline searching and/or conducting registry searching. Registry searches could potentially be useful for baseline risk of clinical event inputs, as often long-term data are not available in literature for the population of interest but could be derived from a registry.

During iterative searching there were two times when I felt uncertain whether I was implementing the method in an optimal manner, and would have felt more certain in my decisions if there had been some prior research published on the topic. One of these times was when I stopped the title/abstract selection in the middle of the first iteration (baseline risk input in the thyroid cancer case study). I made that decision because it seemed so unlikely that I would find the data that I was looking for analysed for the correct population in a format that would allow me to include it in the health economic model. As a validatory step, I completed the screening and found no further relevant citations. The other time was when I decided not to implement the second iteration, and to stop after the first iteration. This was when searching for the utility input for the thyroid cancer case study, where the evidence base extremely limited. Further searching seemed futile.

Varied methods for rapid reviews are described in the literature. In this thesis, a selection of those methods were applied, e.g., limiting by date and by database searched. The methods were similar but not identical across the case studies, depending on the case study context. The rapid review methods that were applied in this thesis were using one database only, restricting synonyms and searching with free-text only. Often a time limit was applied. Published literature suggests that rather than developing one formalised rapid review method, it might be more appropriate to ensure transparency of the rapid review methods used.[411,414]

Rapid review search methods were easier to comprehend, because I was able to think of it as a Cochrane style search where additional limits have been applied to make the review manageable. For both health state utility and baseline risk of clinical event inputs, the searches included limits that were based on the context of the search e.g., rarity of the condition and number of products on the market both which translate to the extent of the evidence base available. All rapid reviews were run only on Medline and included restricted synonyms. All but one were associated with time limitations, most commonly the search was for the last 20 years only. With iterative searching my uncertainty came mostly from my inexperience in constructing search protocols and from the choices that I needed to make to determine the most appropriate scope for the search. With iterative searching, I felt more freedom to go for an initial narrow scope as I knew I would be able to widen the scope as needed. However, with rapid review, I felt I needed to be inclusive enough to be sure to capture the most relevant evidence with one search strategy as I was only running that one search. This was reflected in the higher numbers of records to screen, which in some case studies resulted in a higher burden of the search.

Carrying out information retrieval research is challenging, and there are no established methods. Most research in this field adopts a retrospective design where existing literature reviews are examined and compared.[484-486] In these retrospective studies researchers have looked to establish which proportion of citations were identified through which database/method, and examined the differences. Very few

studies have attempted prospective searching. One such example is the study by Ertaylan *et al.* 2017.[487] The authors found that the search methodology used, and articles identified, differed between two review teams, despite the use of a standardised search framework. The differences were attributed to differences in expertise and research background of team members, search terms used for the database searches and the inclusion and exclusion criteria applied. Therefore, my PhD study represents a rare instance of prospective information retrieval research, contributing to research methodology. Reflecting back I feel this was a useful approach. It would have been possible to retrospectively look at which proportion of usual practice search results were derived from each bibliographic database when usual practice was a Cochrane style search. However, this would not be possible if only one database was searched in usual practice, and even more so if no search steps were recorded. The way search methods are implemented in my PhD study allowed comparison regardless of usual practice.

8.6 Reporting Tools

To be able to report the search processes, results and performance of this PhD study, several reporting structures (tables, figures) were used in this thesis. These are described in the thesis Sections below.

8.6.1 Search Framework (recording and reporting tool)

Part of this project is to review and develop a recording and reporting framework or tool for searches for health economic model inputs. Booth (2006) developed the STARLITE framework as a proposed standard for reporting literature searches.[159] Many of the elements such as sampling strategy, type of studies, approaches (e.g., other than bibliographic databases), range of years, limits, inclusions and exclusions, terms used and electronic sources included are similar to the elements that need to be reported for the searches for model input parameters. Booth's STARLITE framework was adopted before initiating searching by including an additional section for non-biographic search recording, and used to record all the alternative searches of this PhD study. Table 142 displays the search reporting framework employed. For iterative searching the framework was repeated for each iteration.

The framework was adopted prior to commencing searching for the case studies. It was applied and tested when searching for the case studies, and was generally found to be fit for purpose. Both the bibliographic database and non-database search sections of the framework proved important, especially in iterative searching. There is scope for further development of the framework. In the bibliographic section of the framework search terms are recorded. However, the space to enter a full search strategy is too small. In case it was not possible to fit the whole search strategy in this table, I used a second table with the details of the search protocol recorded there. However, this caused some confusion during the supervisor review and needed to be clarified. Therefore, it would be worth considering whether there is another way to connect the overview of the searching shown in this framework, and the full search protocol. However, overall this search framework was easy to use and transparently reported the key features of each search.

Table 142. Final search framework (based on STARLITE[159])

Bibliographic database search elements	Considerations
Sampling	Should the sampling be comprehensive, selective or purposive? Why was specific type of sampling chosen?
Type of studies	What type of studies will be included?
Sources	Which databases and platforms will be sampled?
Limits	What limits can be applied and if so how can they be justified?
Terms used	What search strategy will be used for the main databases?
Conceptual limitations	Can further conceptual limitations be applied such as by geographical location, setting, specific focus of study, etc.? How can these additional limits be justified?
Iterative search specifications	How will the initial search terms be informed? If no suitable evidence is identified, is further database searching is likely to be fruitful? If so, which parts of the inclusion/exclusion criteria should be relaxed? Has progress in the model development process changed the information needs? When to stop searching?
Non-database search elements	**Considerations**
Approaches	What approaches other than bibliographic database searching will be implemented?
Source names	What will be the exact sources to be accessed and what is the link (e.g. web address), if applicable?
Search dates	When will the source be searched?

8.6.2 Reporting Search Concept, Results and Burden

Several reporting tools were used in this thesis. A PRISMA diagram was suitable for reporting the results of the usual practice search and study selection processes, and it was also used for rapid review reporting.[157] For iterative searching, two novel, alternative reporting diagrams were required. Firstly, Figure 40 allows researchers to present the overall iterative search concept. The diagram captures the first iteration of the search: which source(s) were searched. This is followed by the assessment of how relevant the findings were and a decision of whether to continue searching. If further searching was deemed useful, the diagram allows the recording of the next iteration. Further iterations can be added as needed. I found the use of this diagram easy and the resulting diagram was helpful also for me, when I needed to remind myself what was done in each iterative search. Secondly, Figure 41 was used to capture the studies identified in the iterative searches. The diagram was adopted from Paisley.[1] It allowed for transparent reporting of the search results for each iteration, similar to PRISMA for usual practice and

rapid review searches. Further development could include standardisation of the items reported in this reporting tool.

Search tracker is a table template that was developed to capture the burden of the search methods. It was used across all search methods. It allowed recording the number of studies and time spend on each task transparently. I had also set this table up in Excel that calculated the precision and NNR, so that I only needed to copy the results over in order to report them in a word document. The format of the table changed somewhat over the course of the PhD study. Initially, precision and NNR were not included in this table but reported in the text. I decided to include them during the second case study so that all burden related measurements are on the same table. I then went back and changed the format of the tables in the first case study as well, to be consistent with the other two case studies. The naming of the items in each row was also refined during the project. The table format provides a good starting point to record the burden of searches for other researchers. In the future it would be worth considering whether there is too much overlap between the iterative reporting tool (Figure 41), and the search tracker (Table 143as both capture the number of citations at each stage of the review. An example search tacker for an iterative search is shown on Table 109.

Figure 40. Reporting the iterative search concept

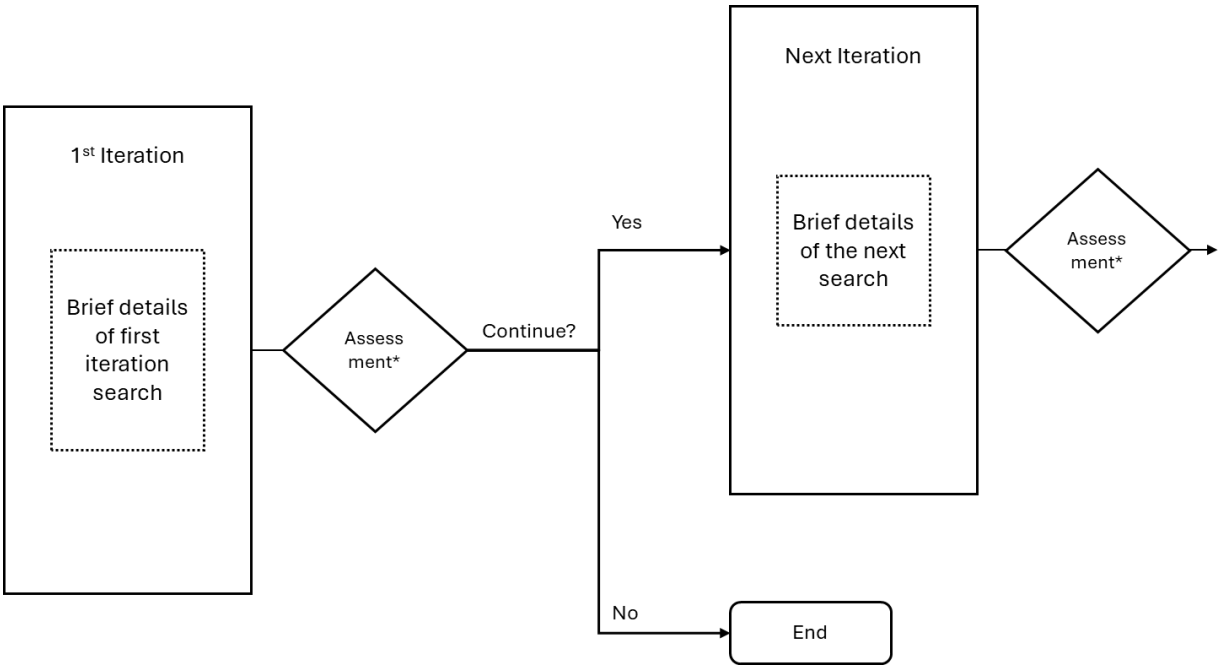


Figure 41. Reporting numbers of findings from the iterative searches

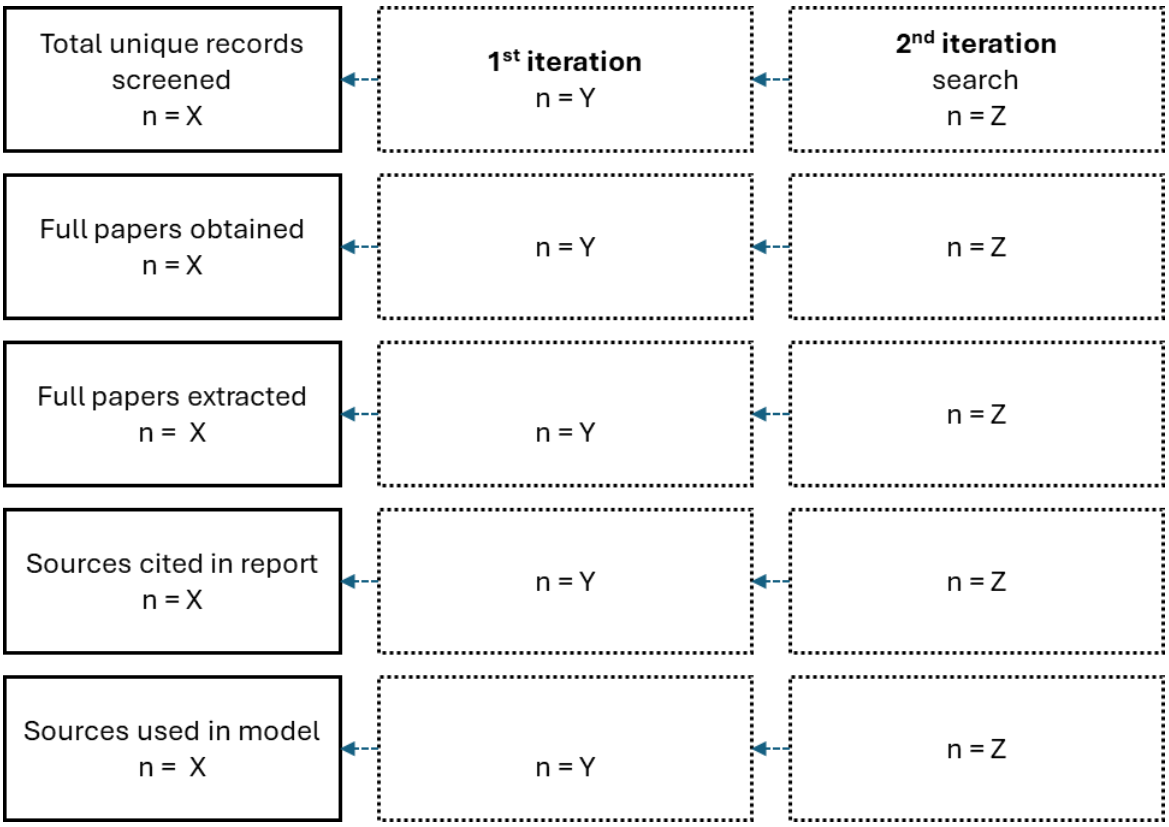


Table 143. Search tracker – Iterative search

Item	Iterative search							
	Number of studies identified			Time in minutes			Precision	NNR
	Iteration 1	Iteration 2	Total	Iteration 1	Iteration 2	Total		
Search protocol development								
Running searches								
Identified citations								
Duplicate removal								
Title/abstract level								
Citations excluded								
Full papers to retrieve + review								
Sources cited in the report								
Sources used in model								
TOTAL								

Abbreviations: NNR: number needed to read

9 Discussion

9.1 Overview

This PhD study explored the identification of evidence for key input parameters in health economic models used in HTAs to evaluate the cost-effectiveness of healthcare technologies. It aimed to empirically test search methods for identifying evidence, addressing the gap noted by Paisley in 2012, where no established methods for retrieving the full range of model input parameters had been reported.[1] Initially Cochrane reviews were designed to identify RCTs, and over the following decades methods have been developed to address a diverse range of questions across health science for a broad spectrum of evidence types.[103,488,489] Traditional systematic literature review methods often struggle to meet the diverse data needs of health economic models. This study compared usual search practices with two alternative methods—iterative searching and rapid review—across case studies in ulcerative colitis, thyroid cancer, and breast cancer tumour profiling. The objectives included developing and testing these alternative approaches, evaluating their efficiency, burden, relevance, and impact on model outputs, and creating a search framework for reporting methods. The motivation stemmed from the critical role of health economic models in HTA decision-making and the need for transparent, efficient search methods to ensure robust model inputs.

This chapter summarises the results from the thesis (Section 9.2) and provides the key findings and contributions to the field (Section 9.3). In Sections 9.4 and 9.5 strengths and limitations have been discussed, and areas of future research outlined. Finally, Section 9.6 provides a conclusion.

9.2 Main Findings

9.2.1 Narrative Reviews of Existing Literature

This thesis presented three narrative literature reviews supporting the PhD project (**Chapter 3**). The first review examined information retrieval methods in evidence-based medicine. The second focused on recommendations for searching for data for health economic models. The third assessed how search methods were reflected in health economic modelling guidance documents from HTA and professional agencies. The reviews aimed to summarise key methods, themes, and publications in order to place the PhD study in the context of existing research. These reviews were narrative reviews but used multiple search methods, such as reviewing academic reports, hand searching journals, browsing guideline-producing organisations' websites, searching Medline, and reference checking relevant publications.

The first literature review found eleven guidance documents from ten recognised organisations that produce or commission reviews (Table 6). These documents provided methodological guidance on searching for various types of reviews, including systematic, rapid, qualitative, and mixed method reviews.[4,5,98-106] Additional methodological advice on searching was identified for various search-related topics, including citation searching, search filters, web searching, iterative searching and automated retrieval methods (Section 3.2.4.2.2). Some publications offered methodological advice for specific review types, either complementing existing guidance or providing recommendations in the absence of guidance.

These covered approaches to searching, developing search strategies, search filters, and reporting standards. Methodological advice also existed for specific aspects of HTA, such as clinical effectiveness, safety, costs, and economic evaluations. The included documents all covered a similar range of search methods, from the standard search strategies used for Cochrane and other well known organisations, to less-used (or reported) techniques, such as rapid reviews, citation searching, iterative searching and berry-picking.

The second literature review found that modelling literature provided little information on formal search methods. NICE DSU TSD 13 detailed how to search for and review evidence for models submitted to NICE.[26] Golder *et al*. identified issues with search methods for models, including data quality and efficiency of search approaches.[8] Glanville & Paisley reported that health economic model searches could be less exhaustive and more targeted than review searches.[47] Paisley (2016) suggested minimum searching levels for each model parameter type, noting these had not been tested empirically.[10] Paisley (2012) concluded that developing search methods for models was an area for further research, with few additional publications identified.[1]

The third literature review examined how HTA and professional organisations' modelling guidelines reflected evidence for health economic model input identification. It aimed to determine the extent and basis of recommended information retrieval techniques. A scoping review methodology was used, recognising that country-specific HTA guidelines were not commonly indexed in bibliographic databases. Forty-two guidelines were identified, with varying levels of detail on information retrieval for model inputs. The most common recommendations concerned relative treatment effect estimates, often based on comprehensive search methods like Cochrane or CRD guidance.[5,86] Methods for AE searching varied (Section 3.3.4), while iterative searching for health state utility values was emerging as a method but lacked empirical research.[228] For costs, resource use and baseline risks, guidance was limited to local sources. The review highlighted limited health economic model-specific search methods, especially outside relative treatment effect identification.

The three reviews of existing literature on information retrieval found established methods for systematic reviews and some guidance for specific review types or data, but practical guidance for non-Cochrane style reviews was limited. Iterative searching, emphasised in guidelines like TSD 13[26] and the ISPOR Task Force Report,[162] aligned with the iterative nature of health economic model development and emerged as a search method to test in this PhD study. Review 3 found that many HTA/health economic model guidelines required systematic and transparent searches for model inputs, though specifics were often lacking. Clinical efficacy, typically informed by Cochrane style systematic searches, was an exception.

Iterative search method was the first technique chosen to implement in this case study. Rapid review was chosen as the second method to test. This is because rapid reviews, are faster yet methodologically similar to full systematic reviews, and had therefore potential to address resource-intensive challenges in identifying health economic input parameters. The literature review also identified a number of search reporting frameworks, that provided a basis for adopting one for this PhD study. One reporting tool was adopted for reporting the searches (sources searched, any limits applied etc.). A novel set of reporting tools was develop specifically for iterative searching, so that the iterative concept and the findings by iteration can be transparently reported. A search tracker was also developed to report the number of publications identified as well as the time it took to carry out the searching.

9.2.2 Implications to Methodological Approach

Chapter 4 described the methods employed in this PhD study. The narrative reviews of existing literature had identified iterative searching and rapid review as suitable methods for testing. The usual practice search approach served as a control arm against which the alternative search methods could be compared. Three possible usual practice search methods were included in this project: full systematic literature search, minimum search recommendations by health economic model type, and no specified search methods/rationale. The methods were tested on two model input types: health state utilities and baseline risk of clinical events. These types of inputs were chosen because: 1) they are not associated with established and tested search methods, 2) they are often impactful model inputs, and 3) they are representative of the challenges that are also associated with other model inputs. This PhD study used a case study approach to explore alternative search methods, as no empirical research on the alternative search methods was found in the literature, and therefore it was judged unlikely that other methods such as interviews would be feasible. The case study involved testing the identified inputs in an executable health economic model. The three case studies differed from each other in terms of disease areas, health technology and modelling approaches employed, offering an opportunity to test the search methods for different types of case studies.

The decision to stop searching is an important consideration in conducting iterative searching. Two concepts, relevance and saturation, were helpful in determining when to cease data collection and when to stop searching for model inputs. Relevance is a dynamic, multidimensional concept that can accommodate the complex and emerging nature of information. Most health economic modelling information queries could be characterised as dynamic, rather than static, although static information needs also exist. In case study 2 (Thyroid Cancer) a decision was made to stop searching further health state utility values after iteration 1. The first iteration included a HTA document review and a narrow bibliographic database search. Based on the initial HTA document review, an extremely limited evidence-base was observed in this specific type of thyroid cancer (MTC). For this reason, the literature search in iteration 1 was carried out using search terms related to thyroid cancer rather than MTC. Therefore, a decision what/how to search was modified according to the availability of information in absence of ideal information. The concept of proxy relevance was therefore used in this case study, due to rarity of the disease and the resulting lack of evidence. After carrying out both the HTA document review and the bibliographic database search for iteration 1, the model was run using the five identified sources. The alternative values produced similar model results: ICER varied between £148,19 and £165,890 per QALY gained. Given the low likelihood of identifying new data and the low potential marginal relevance of any new data identified, a decision was made to stop searching utilising the principles of marginal relevance.

Marginal relevance is a concept that can be applicable to any model input or development stage. In this study, testing inputs in an executable health economic model was employed to help determine marginal relevance. Once a health economic model is programmed, it can be used to assess the potential impact of incorporating additional evidence, regardless of the type of model input. Similarly, marginal relevance is a valuable concept even before the model is programmed. Although this PhD study did not cover model conceptualisation, some reflections can be made. During model conceptualisation, determining the model health states requires a thorough understanding of the disease and the impact of the health technology

on the disease and patients, which can be developed by acquiring literature and other information on the topic, such as expert opinion. During this early stage of model conceptualisation, much of the information new and therefore has high potential to inform the model structure. Once the disease and the impact of the health technology are more understood (as much as needed for modelling purposes), further knowledge will not alter the model structure and leads to a redundancy of additional information. Proxy relevance can be an important concept in the context of searching data for rare diseases, like MTC, where evidence base is limited. Additionally, it can also be helpful for model inputs if no data has been identified (e.g., for a particular disease consequence or an adverse event). For example, if no utility decrement can be found for cataplexy, could epilepsy seizure utility decrement be used instead? Proxy relevance requires careful consideration and possibly input from experts.

Multi-dimensional relevance evaluates the attributes of different options, as defined by the information characteristics. In Case Study 1 (UC), several surgery-related inputs were considered: the probability of elective surgery, the probability of surgical complications, the probability of complications requiring further surgery, the probability of developing pouchitis, and the probability of death due to surgery. When searching for and selecting values for these inputs, multiple factors needed to be considered: study population, inclusion/exclusion criteria, loss to follow-up, sample size, missing data, type of surgery, and the geographical location of the study. Also, comparing how the endpoints were defined in the published sources versus the definitions used in the model played a role in determining the most appropriate source. In the UC case study, one iteration was sufficient for all other basely risk of clinical event inputs except one. A second iteration was conducted for the probability of death due to surgery because no UK data had been identified in the first iteration. For this input, local data was deemed relevant, prompting a second search iteration. Multi-dimensional relevance is particularly important for health state utility selection. Additional considerations include using health state utility values from a single source (versus combining several sources with potentially varying instruments/methods) and the instruments/methods used (versus those recommended by the HTA agency for which the model is built). These considerations were applied in all of the health state utility case studies in this PhD study.

Evolving relevance refers to decisions that depend on the development of understanding over time. Definitions of relevance are subject to iterative sense-making and information seeking. In this PhD study, evolving relevance was crucial concept for iterative searching. In Case Study 3 (breast cancer tumour profiling risk stratification), the first iteration provided an overview of the evidence base, while the second iteration expanded upon it. Initially it was not clear for which CDK4/6 inhibitor the data (OS) should be searched for. The first iteration provided an overview of CDK4/6 inhibitors, as well as some of the available evidence for overall survival (OS) for CDK4/6 inhibitors. The second iteration expanded the search to include other publication types and the most recent evidence. Evolving relevance can be an important concept also for model conceptualisation,[31] even if this PhD study did not provide an opportunity to test it in this context.

Saturation was part of the concept of relevance, because if further information added nothing to the task, it was not relevant (marginal relevance). In this research project, constant comparison was used to arrive at conclusions about relevance and saturation, and when to stop searching. In theoretical saturation the researchers sample until no further data is found for the different categories.[253,254] In this PhD study theoretical saturation was reached when searching for health state utility values in case study 2 (thyroid cancer). The disease (MTC) is ultra rare, and no new data was found for any of the health states (health

states being the categories). Inductive thematic saturation defined saturation as a point in analysis when no new theme occurs. This concept is most relevant during model concept development than for this PhD study. For example, initially the health states may not yet be finally decided, and information is gathered to finalise the definitions. Researchers may refine the health states until no new themes occur and there are an increasing instances of the same themes (i.e. health states). Due to the scope of this PhD study, theoretical saturation was more relevant concept than inductive thematic saturation, because it is more focused on the degree of refinement of existing themes rather than identifying new, emerging themes. As the above examples highlight, both saturation and relevance were open to interpretation, and the decision to stop searching varied between different searches.

This PhD study compared usual search practice and alternative search methods, and to do this comparison, performance outcomes were needed. Besides the standard measures (sensitivity, precision) available in the literature on conducting systematic reviews, further outcomes measuring the burden and relevance of findings were also needed. Therefore, the performance outcomes in this PhD study included time spent, sensitivity, burden (precision and NNR), and relevance. The usual practice searches were updated to ensure comparability with alternative search methods. The study estimated the time spent on the search protocol, running of searches, citation identification, duplicate removal, and publication screening.

This study reviewed and adopted reporting methods for alternative searches to maintain transparency. Existing standards were primarily identified for Cochrane style systematic reviews, but a slightly modified version of Booth's STARLITE standards were proposed for both quantitative and qualitative literature searches.[159] A draft reporting tool was therefore created based on STARLITE, but with the addition of a detailed non-database searching section to capture the heterogeneous nature of sources for health economic model inputs. This framework aimed to help in planning and recording searches, as well as providing an easy step to reporting them.

9.2.3 Findings from the Case Studies

This PhD thesis included three case studies in **Chapters 5- 7**: UC, thyroid cancer and breast cancer tumour risk stratification. A brief summary is given here for each of the case studies.

The first case study on **ulcerative colitis (UC)** examined the use of infliximab, adalimumab, and golimumab for treating moderately-to-severely active UC after conventional therapy failure. It used a health economic model to assess the cost-effectiveness of second-line treatments, conventional therapies, and immediate colectomy. The model, based on a Markov structure, included surgery-related risks like elective surgery, complications, pouchitis, and death. Iterative searching and rapid review methods were tested for identifying health state utilities and baseline risk inputs. For health state utilities iterative searching, which involved focused bibliographic searches and Medline searches, was found to be 74% less time-consuming than the usual practice and identified many of the key sources. Rapid review was more efficient but missed some key sources. For baseline risk of clinical events, the usual practice search for identifying baseline risk inputs was not as transparent as it could be, since the search steps for only one of the inputs were reported. The use of iterative searching and rapid review methods increased

transparency by recording information retrieval steps for all five model inputs. However, rapid review had lower precision and higher NNR compared to the other two methods. The time required for rapid review was more than for iterative search or usual practice search. Registry searches were also conducted but data access was limited. All identified data showed little difference in model results, with iterative searching being the most efficient and effective method.

The second case study focused on **medullary thyroid cancer (MTC)** and compared cabozantinib and vandetanib with best supportive care (BSC) for treating unresectable locally advanced or metastatic MTC. An independent Assessment Group developed a health economic model to estimate the incremental cost-effectiveness of these treatments using a partitioned survival approach based on three health states. Iterative searching and rapid review methods were tested for identifying health state utilities and baseline risk inputs. For utilities, iterative searching was 89% less time-consuming than the usual practice search, while rapid review was slightly more time-consuming. The model included progression-free survival (PFS) and overall survival (OS) as baseline risk inputs. Iterative search for these inputs was stopped early due to limited useful data, and a second iteration focused on identifying MTC registries, though data access was limited. Rapid review identified most sources from the usual practice search and two additional sources, but these were not used in the model due to comparability concerns. Overall, iterative searching was the most efficient method, and combining iterative and rapid review methods offered a systematic and transparent approach to identifying supplementary evidence.

The third case study focused on **breast cancer tumour risk stratification** using Oncotype DX, a test that potentially predicts breast cancer recurrence and chemotherapy benefits. The External Assessment Group (EAG) developed a health economic model to estimate the test's cost-effectiveness compared to conventional decision-making, using a hybrid decision tree and Markov structure. For health state utilities iterative searching was 39% faster than usual practice and identified the most publications. For baseline risk inputs, iterative searching included a focused Medline search of systematic reviews and expanded to other publication types, identifying relevant OS data from two recent publications. Rapid review identified the same two sources plus three unique ones, but these were not used due to comparability concerns. Iterative searching was the most efficient method, with all methods showing similar model results. The use of iterative searching and rapid review methods increased transparency by recording information retrieval steps across both methods. Iterative search was more efficient than usual practice or rapid review, a reduction of 60% compared to usual practice was observed. All identified data across the search methods showed little difference in model results.

9.2.4 Case Study Reflection

Chapter 8 discussed the results of the case studies conducted as part of this PhD study.

Iterative searching emerged as promising information retrieval approach, especially for health state utility input identification. In all of the health state utility case studies, it was consistently more efficient compared to usual practice. Additionally, iterative searching also identified the most relevant sources and/or new utility inputs that resulted in similar health economic model conclusions as the inputs from the usual practice search. The conclusions about search efficiency held true, despite a limitation

associated with this study: usual practice search time is likely to be underestimated as it was not available from the original published reports describing the usual practice searches. The retrospective time estimations could only be made for the title/abstract and full text selection parts of the search and not for search protocol development or running of the search. Therefore, iterative searching is likely to be even more efficient than this study demonstrated.

Searching for baseline risk of clinical events is challenging, as data requirements vary significantly between different disease areas and can be specific to health care jurisdictions. The case studies in this PhD study were different from each other in several ways, including covering different disease areas (two oncology and one chronic diseases), as well as different types of health technologies (two drugs and one predictive test). The potential data sources for baseline risks also vary greatly: from published studies identifiable from bibliographic databases to national statistics, clinical/economic professional organisations and charities. Using iterative searching to retrieve baseline risk of clinical events data was in general less efficient than using iterative searching to retrieve utility inputs. However, it still emerged as a potential alternative method to retrieve baseline risk of clinical event model inputs, as it was mostly more efficient than usual practice, with the exception of the UC case study. It also allows the piecemeal approach of identifying registries, which cannot easily be accommodated in the Cochrane-style approach.

Iterative searching for health economic model input identification is a dynamic process that requires oversight from the researcher, and continuous recalibration. During the iterative searches in this thesis, the knowledge of the context in which the search was being performed was developed, and helped to shape the next iteration (if relevant) or form the rationale for stopping searching. It proved possible to develop a fairly consistent approach to iterative searching across the case studies (especially for health state utility identification), and the process was easy to record, using the reporting tools employed in this PhD study.

The efficiency of conducting rapid review compared to using usual practice varied between the case studies. For utility input identification rapid review did not provide an advantage over usual practice in most of the case studies (two out of three). Searching for health state utilites using a single search query approach can be challenging. One reason is that there is a need for an initial search to determine the extent of the evidence base. It can be done by conducting a broad search, followed by refined search and/or inclusion criteria as understanding evolves. Another approach, employed in this PhD study, is to conduct a narrower search to get an initial, quick understanding whether any highly relevant data emerges, relaxing the criteria as needed in later iterations. An additional challenge with utility identification is that due to indexing issues: quality of life topics, like utilities, are not easy to retrieve from bibliographic databases. Usual practice searching and rapid review approaches are similar in the way that they approach searching by using one pre-defined search query. This is in contrast to iterative searching that allows for an iterative approach that addresses the specific challenges of searching for utility inputs. In the context of searching for utility inputs, both rapid review and usual practice searching were impacted by similar issues resulting from the single search query concept, as opposed to the iterative nature of iterative searching. Using rapid review method to retrieve baseline risk of clinical events did not show a clear trend to either support or reject this method. The time comparison to usual practice search ranged

from a 539% increase to 84% decrease in search time. This is likely a reflection of the varied nature of baseline risk of clinical events, from one disease area to another, as well as geographical differences. Another factor is that often usual practice searches were not (at all or partially) reported i.e. no or little time could be assigned to them. It is possible that the usefulness and efficiency of rapid review methods varies greatly from one health economic model to another. This PhD study with three case studies does not provide a sufficient evidence base to conclude under which specific circumstances rapid review might be most useful. Further research is needed to determine this.

One of the most challenging aspects of the project was the lack of literature on alternative search methods, especially those that describe detailed practical steps how to carry out the searching. It was unclear how to decide on the search approach, what information to use when determining stopping or continuing searching, and how to set limits on the searches. Stopping searching was a crucial consideration in both rapid review (e.g. which databases to search, and iterative searching, although in iterative searching it is more explicit decision to continue/discontinue to the next iteration and reporting the rationale for the decision. In rapid review stopping is considered more implicitly e.g. when deciding how many databases to search, how many synonyms to include, or how much supplementary searching to conduct. Marginal relevance was an important concept, as searching was often stopped due to the redundancy of additional information. Multidimensional relevance was also significant when considering stopping iterative searching, taking into account factors such as the study population, sample size, geographical coverage, and data format of the studies identified. Evolving relevance was another key concept, as information retrieval needs to evolve alongside the model concept.

Both saturation and relevance as dynamic concepts are open to interpretation, and therefore the decision to stop searching was different for different case studies. In this study an important part of saturation and relevance assessment was done by running the health economic model with the identified inputs, when possible. If it was not possible to run the health economic model, observations were made about the additional information that the new data are likely to bring to the assessment of the decision problem.

The key reasons for **continuing** searching in this PhD study included:

- Data were not yet identified for all the model inputs (e.g., a utility value was not found for all the health states);
- Little data, or data that were not optimal, were identified and there was uncertainty about the possibility of further, more relevant data existing;
- Lack of confidence in the iterative search method when applied to health economic model input identification (most prominent in the first case study).

The key reasons for **stopping** searching were:

- Rare disease with very limited evidence base;
- No further relevant evidence found or expected, a decision often derived from running the identified inputs in the health economic model and examining the impact on the model results (both base case and sensitivity analyses).

The approach to stopping searching in the case studies of this PhD study was consistent in terms of the theoretical framework adopted. Searching was mostly stopped based on decisions accounting for the redundancy of additional information i.e., marginal relevance. This allowed stopping searching in a way that best met the aims and objectives of each search.

Compared to usual practice search, the iterative searching was not only more efficient in terms of search time, but it more consistently included all appropriate inputs in the searches. For example, all health state utility inputs for the breast cancer risk stratification case study were included in the iterative search (recurrence free, distant metastasis, AML, chemotherapy and local recurrence), whereas usual practice only reported search steps for two of those health states (recurrence-free and distant metastasis). The search recording and reporting framework specifically developed for this PhD study allowed transparent reporting of the bibliographic as well as non-bibliographic searching.

This PhD study, using the three case studies, demonstrated that iterative searching has potential to be less time consuming, more comprehensive and more transparent search method than current usual practice across a number of inputs– certainly for health utilities and baseline risk of clinical events, and potentially for other inputs also.

Research in information retrieval is challenging, and there are no established methods how to compare the search approaches, especially in prospective comparisons. Most research in this field adopts a retrospective design, examining existing literature reviews and comparing the proportion of citations identified through each database or method. However, few studies have attempted prospective searching, making this study a rare instance of prospective information retrieval research. It also potentially offers some means of measuring and recording processes for future, prospective studies in this field. The implementation of search methods in this study allowed for comparison regardless of usual practice.

By necessity, this PhD study also contributed to the development of a recording and reporting frameworks for health economic model inputs identification searches. The STARLITE framework was adopted before commencing searching for case studies and was found to be fit for purpose. The framework was found to be easy to use and it allowed transparent reporting of the key features for each search. Several additional existing and new reporting tools were used in the thesis, including the PRISMA diagram for reporting results of usual practice search and the study selection processes, and two new reporting diagrams for iterative searching. These diagrams allow presentation of the overall iterative search concept, as well as reporting the results identified in iterative searches (by search iteration). A search tracker was developed to capture the burden of search methods, recording the number of studies and time spent on each task transparently. The table format changed somewhat over the course of the PhD study, with precision and NNR included in the table for consistency across the case studies. The table format provides a good starting point for recording the burden of searches for other researchers. Further development of these reporting tools could include standardization of the items reported.

9.3 Contribution of the Thesis to the Field

This PhD study has empirically demonstrated that the single query with strict inclusion/exclusion criteria approach is less efficient approach to searching for certain inputs in health economic models than using an alternative, more flexible search approach.

This PhD study makes the following key contributions to the field:

1. **Contribution to the existing empirical evidence base on comparative evaluation of different search techniques**:

    - There are no established methods for carrying out information retrieval research. Nearly all research in this field adopts a retrospective design that look at establishing which proportion of identified studies were identified through which database/method.[445-447] Very few studies have attempted prospective searching.[448] This PhD study represents a rare instance of prospective information retrieval research, contributing to research methodology.

2. **Some empirical evidence on the effectiveness of different approaches in the context of health economic models**:

    - An important contribution of this PhD study has been to combine the searching for data with testing of the identified values in the health economic models. No publications were identified where sources identified by different search methods were tested in the health economic model, to allow decisions to be made about relevance of (any additional) data as well as stopping searching.

    - During the course of the study, it was possible to see how the iterative searching concept lends itself well to the nature of health economic model input identification. Often there are several, related model input values to search for (e.g., utility values for several health states), and the iterative process is able to incorporate the developing nature of the search requirement. This same would apply to model development process, although this thesis was not able to test that due to the study design: the health economic models in the three case studies examined in this PhD study were already developed and programmed at the start of the study. Therefore, the searching was done to inform an existing model structure as opposed to informing a developing model structure. Rapid review was also tested and found potentially useful search alternative. However, in some cases the usefulness of rapid review may be hindered by the same qualities that make the Cochrane style searching challenging to use for model input identification (pre-defined, single search query, strict inclusion/exclusion criteria).

    - HTA agencies have in recent years become more open to accepting alternative search methods. TSD 13 provides advice on how to search for and review evidence for models submitted for NICE.[25] ISPOR best practice guidance recommends iterative searching for health state utility values (without providing detailed guidance on its conduct, recording or

reporting).[204] Rapid reviews (such as rapid HTA, mini HTA, and rapid response) have gained widespread use in various HTA agencies.[449,450] The CADTH established its rapid review program in 2005 and consistently publishes hundreds of rapid reviews each year.[449] This thesis demonstrates that rapid review can be either as effective as usual practice, or it is associated with increased transparency when information retrieval steps are available for all model inputs. Usual practice searches often did not report the input identification steps, leading to lack of transparency.

3. **A transparent step-by-step description on how alternative approaches might be conducted and implemented**:

    - The application of iterative searching and rapid reviews are described step-by-step in the thesis. Therefore, this PhD study contributes to the description and standardisation of alternative search approaches.

    - This PhD study also attempts to operationalise the concepts of saturation and dynamic relevance in the search process, particularly with regards to the stopping rules and when assessing the value of information in terms of impact on the model results.

4. **A reporting framework**:

    - During this PhD study, a reporting framework was adopted from STARLITE, and other diagrams/tables were developed to capture the search concepts, results and performance especially in relation to reporting of iterative searches.

    - This PhD study contributes to the development of additional performance outcomes that can be used to compare search methods, that take into account time, burden and sensitivity of the searches.

This PhD study has so far resulted in two conferences posters, and one full published manuscript.[232,451,452] One further manuscript is currently being revised as a result of a peer review.[271,490,491] Further publications have been planned to report the findings from the other case studies, as well as the performance measures and reporting tools adopted/developed and their implications for practice.

9.4 Strengths and Limitations

The key strength of this study is that it focused on empirically testing different search methods using case studies, and testing the resulting inputs in a health economic model, when feasible. This enhances the assessment of relevance of the additional data on the model outputs. Further, this study not only empirically tests the alternative search methods, but also contributes to the development of those said methods. No common definition of iterative searching or rapid review exist in the literature, and additionally the practical steps of carrying out iterative searching are not described in the context of health economic model input identification. This thesis has contibuted to the advancement of information retrieval research methods. A search reporting framework and tools have been developed, applied and

tested that provide a valuable contibution to search reporting in this field. Other areas of development include performance measures for searches and operationalisation of stopping rules in searching.

This study represents only a limited number of case studies, and other case studies may generate different findings. A case study apporach can be helpful in highlighting different, or unusual cases as well as making general conclcusions across them. The case studies included in this PhD study represented a varied selection. There were important differences in terms of diseases (UC, thyroid cancer, breast cancer), health techonologies (two drug and a prognostic test) and dates when the assessment took place (2016, 2019 and 2024). There were also differences in the usual practice searches, ranging from comprehensive systemtic review to no record of how searching was conducted. Despite these differences, consistent patterns emerged, especially in relation to searching for utility inputs using iterative searching.

Additionally, search methods have been applied in a manner judged to be the most appropriate by me (in consultation with my supervisors).There was consistency between those applications: the iterative search was usually HTA document review + focused Medline search, followed by a further, expanded Medline search and/or registry search. Rapid review was mostly a Medline search with limits related to time and resticted synonyms. Selection of the most relevant utility values, for the base case and for the sensitivity analyses, is also subjective. These choices were moderated by evaluating the studies in terms of study population, sample size, geographical location, and other features that were relevant for a specific model input. Further, several sensitivity analyses were run in the model for each case study to better understand the impact of the data on the model conclusions. Other decisions on how to apply the search methods or which utility values were selected, may have led to different results. To mitigate this variability, I carried out searching as described in detail in the methods (Section 4.2.2.3). Additionally, frameworks were used (Sections 4.2.3 and 8.6.2) to ensure the searches provide sufficient details to be replicable. Due to these limitations the results cannot be generalised beyond these case studies. Further, the searching was done by one reviewer only, and the order in which the search methods were implemented may bias the results as the reviewer's knowledge of the topic area grows. However, having one reviewer to implement all the search methods was also an advantage, as it is likely to result in more consistency e.g., both when searching (as searching can vary person by person) and when recording time keeping. Also, the search methods were implemented in different order in each case study to account for the bias that may be introduced by the learning curve.

Time for the search protocol development for usual practice search was not available from published reports, and therefore not included in the study. This is likely to result in underestimation of the effort associated with usual practice searching. Time for search protocol development may have been different if an experienced information specialist had been involved. My approach to searching (in general, not just with the alternative search methods) developed during the PhD study, which may result in some minor differences between the case studies. An experienced information specialist may have had a more consistent and knowledgable approach from the outset. One of the search performance measures was sensitivity. Usually this is calculated as the number of relevant citations identified out of the total number of relevant sitations in existence. If the total number of citations in existence is defined as the citations identified by the usual practice search, in this study sensitivity would often be above 100% as alternative

search methods frequently identified publications not identified by usual practice. The burden measures, NNR and precision, were more straightforward to measure and report. Relevance as a performance metric can be considered subjective. In the case of this PhD study, running in the model was a helpful test to more objectively understand whether new sources made any difference for modelling, and therefore to the decision-problem.

Future stakeholder engagement, particularly involving information specialists, will be crucial for taking the findings from this PhD study further. Information specialists can significantly contribute by refining and guiding the search approaches and terminology, ensuring that the search strategies are both relevant and precise. Their expertise will also be invaluable in interpreting the findings from the case studies, providing deeper insights and contextual understanding. Additionally, organizing focus group discussions with information speciliasts, as well as other specialists (e.g. systematic reviewers, health economics, HTA speciliasts) can help in developing a more robust alternative search process and transparent reporting framework. Such discussions will facilitate the alignment and prioritisation of future research recommendations, ensuring that they are well-informed and practically applicable. By actively involving information specialists and other stakeholders, potential gaps and limitations can be identified and addressed more effectively, leading to more reliable and impactful outcomes.

This study did not investigate the use of mega databases (such as Semantic Scholar, Lens.org and OpenAlex) and/or artificial intelligence (AI) assisted literature review tools. Mega databases are searcheable online schlarly literature databases that aggregate bibliographic data from other databases and combine them into one.[492] While they contain vasts amounts of data, they are not curated and contain predatory journals and might therefore be of lower quality than curated databases, such as Web of Science or Scopus.[492] There are also concerns over mega databases ability to identify sources other than journal articles, e.g. books.[493]

AI has made significant advances in automating literature searches since 2017 (the start of this PhD study), especially for the screening and data extraction phases, with potential to make conducting systematic reviews substantially quicker.[494,495] Fast(er) systematic reviews would reduce the impact of alternative, quicker search methods investigated in this study. The AI assisted literature review tools have demonstrated significant effectiveness when utilised properly.[496] Nonetheless, currently these tools are still associated with challenges, especially for use in the context of policy-making. AI-assisted literature review tools depend on a variety of methodologies. Some use basic classifiers or Bag of Words methods for text representation, which are no longer considered the state-of-the-art.[497-501] Newer tools have adopted word and sentence embedding techniques that are advanced natural language processing technologies, particularly large language models (LLMs).[496] One problem that has been observed is that LLMs are trained on general data, and therefore they may result in less effective performance in specialist fields.[496] Additionally, LLMs may also generate inaccurate information that is known as “hallucianations”.[496] Understanding the decision-making process of LLMs is complex and the outputs can be inconsistent.[502] Contemporary AI models can be seen to operate as “black boxes”, making their internal processes difficult to comprehend.[502] For example, little information is available on why certain paper was deemed relevant or irrelevant.[502] For AI assisted systematic review tools to be adopted in policy-making,

where transparency and rigour are of critical importance, these issues hamper wide adoption until the challenges have been addressed.[496] Therefore, at least for now, the contributions from this PhD study are relevant until such issues have been resolved.

9.5 Areas of Further Research

Further research is recommended in the following areas:

- This thesis described steps for iterative searching of health economic model inputs. Further research is recommended to apply these procedures in other case studies and test the generalisability of the findings from this PhD study. This will help to operationalise and standardise the alternative search approaches.
- Further, this thesis developed and described measures of effectiveness of the search process beyond sensitivity and precision (NNR, change in search time, testing relevance in the health economic model). These concepts can be tested and developed further to advance information retrieval research, not only in the context of health economic modelling but for conducting comparisons of search approaches more generally. Development of such measures will help to define how the usefulness of searches should be measured.
- This empirical research reports the rationale for continuing/stopping searching for the three iterative searches. These decisions took into account availability of evidence and impact on model outputs. However, this study design did not allow examination of stopping decisions in terms of value of information to the modelling process, as the models in these case studies were already developed. Future research can look into incorporating all aspects of stopping searching to develop guidance.
- This thesis also outlines a proposed search recording and reporting framework for searching for model inputs. Other reporting tools were also used and developed in this PhD study. These should be tested by other researchers and, as necessary, further developed and assessed through further case studies.
- While the results for iterative searching were consistent across the case studies (very consistent for health state utility inputs and fairly consistent for baseline risk of clinical event inputs), this was not the case for rapid review. The effectiveness of rapid review as a search methodology for health economic inputs may be associated with some efficiency gains. However, the results varied across the case studies, and further research is needed to establish under which circumstances rapid review might be an effective search method for health economic model identification.

9.6 Conclusion

Searching to identify health economic model inputs is complex. Usual practice can result in a high number of citations to screen, but the trade-off between sensitivity and efficiency has not been researched. This thesis has demonstrated that an alternative approach, iterative searching, was minimally time intensive but productive, yielding highly relevant references for inclusion in the health economic model. No differences between usual practice and iterative searching were observed in the conclusions that could be drawn from the model for decision-making in the three case studies assessed. The usefulness of rapid review as an alternative approach significantly varied from one case study to another. This may indicate

that the appropriateness of the method is highly impacted by the context in which it is applied, such as the type of model input or disease or geographical area. More research is needed to determine when rapid review may be a more useful search method than usual practice.

This PhD study shows that a single query with strict criteria is less efficient for certain health economic model inputs compared to a flexible approach. It contributes to the empirical evidence base by providing rare prospective information retrieval research and demonstrates the usefulness of combining data searching with testing in health economic models. The study offers a step-by-step description of alternative search methods. It also adopted a reporting framework from STARLITE as well as developed/adopted further reporting tools. Additionally, new performance outcomes were developed to measure the burden and time it takes to carry out searching. Despite the limited number of case studies, consistent patterns emerged, particularly in searching for utility inputs using iterative searching. The study's limitations include subjective selection of model input values and potential bias from a single reviewer.

Researchers should consider applying the iterative search approach when conducting information retrieval for health state utilities and baseline risk of clinical events. Researchers are also encouraged to record the search steps, using the search recording and reporting framework developed as part of this study. This thesis included a set of three case studies, but further research is needed to explore whether the findings of these case studies can be generalised to other decision problems.

References

1. Paisley S. Identifying evidence to inform decision-analytic models of cost-effectivenss. A qualitative study of information seeking processess and behaviour Sheffield, England Sheffield University 2012.
2. Karnon J, Vanni T. Calibrating Models in Economic Evaluation. *PharmacoEconomics* 2011; **29**(1): 51-62.
3. Vreman RA, Mantel-Teeuwisse AK, Hövels AM, Leufkens HGM, Goettsch WG. Differences in Health Technology Assessment Recommendations Among European Jurisdictions: The Role of Practice Variations. *Value in Health* 2020; **23**(1): 10-6.
4. Centre for Reviews and Dissemination UoY. Centre for Reviews and Dissemination. Systematic reviews: CDR's guidance for undertaking reviews in health care. York, 2009.
5. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.0 [Updated July 2019]. 07/2019 2019. https://training.cochrane.org/handbook/current (accessed 05/09/2023 2023).
6. Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. *Journal of Health Services Research & Policy* 2005; **10**(4): 245-50.
7. Zechmeister-Koss I, Schnell-Inderst P, Zauner G. Appropriate evidence sources for populating decision analytic models within health technology assessment (HTA): a systematic review of HTA manuals and health economic guidelines. *Med Decis Making* 2014; **34**(3): 288-99.
8. Golder S, Glanville J, Ginnelly L. Populating decision-analytic models: the feasibility and efficiency of database searching for individual parameters. *IntJTechnolAssessHealth Care* 2005; **21**(3): 305-11.
9. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S, Philips Z. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *PharmacoEconomics* 2006; **24**(4): 355-71.
10. Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. *Pharmacoeconomics* 2016; **34**(6): 597-608.
11. International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Pharmacoeconomic Guidelines Around the World https://tools.ispor.org/peguidelines/ (accessed August 2019).
12. Gough D, Thomas J, Oliver S. Clarifying differences between review designs and methods. *Systematic Reviews* 2012; **1**(1): 28.
13. Higgins JPT, Davey Smith G, Altman DG, Egger M. Principles of Systematic Reviewing. Systematic Reviews in Health Research; 2022: 17-35.
14. Glanville J, Lefebvre C. Identifying Randomized Controlled Trials. Systematic Reviews in Health Research; 2022: 36-54.
15. Deeks JJ, Riley RD, Higgins JPT. Combining Results Using Meta-Analysis. Systematic Reviews in Health Research; 2022: 159-84.
16. Delaney A, Tamás PA. Searching for evidence or approval? A commentary on database search in systematic reviews and alternative information retrieval methodologies. *Research Synthesis Methods* 2018; **9**(1): 124-31.
17. Magarey JM. Elements of a systematic review. *International Journal of Nursing Practice* 2001; **7**(6): 376-82.

18. Lane D, Dykeman J, Ferri M, Goldsmith CH, Stelfox HT. Capture-mark-recapture as a tool for estimating the number of articles available for systematic reviews in critical care medicine. *Journal of Critical Care* 2013; **28**(4): 469-75.
19. Paynter RA FR, Stoeger E, et al. A prospective comparison of evidence synthesis search strategies developed with and without text-mining tools. *J Clin Epidemiol* 2021.
20. Bates M.J. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 1989; **13**: 407-24.
21. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: An exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. *Systematic Reviews* 2015; **4**(82).
22. Pearson M, Moxham T, Ashton K. Effectiveness of Search Strategies for Qualitative Research About Barriers and Facilitators of Program Delivery. *Evaluation & the Health Professions* 2011; **34**(3): 297-308.
23. McManus RJ, Wilson S, Delaney BC, et al. Review of the usefulness of contacting other experts when conducting a literature search for systematic reviews. *BMJ* 1998; **317**(7172): 1562-3.
24. Adams J, Hillier-Brown FC, Moore HJ, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. *Systematic Reviews* 2016; **5**(1): 164.
25. Booth A. Searching for qualitative research for inclusion in systematic reviews: a structured methodological review. *Systematic Reviews* 2016; **5**(1): 74.
26. Kaltenthaler E, Tappenden P, Paisley S. NICE DSU Technical Support Document 13: identifying and reviewing evidence to inform the conceptualisation and population of cost-effectiveness models. Sheffield, 2011.
27. HAS. A methodological guide. Choices in Methods for Economic Evaluation. 2012.
28. Altman DG. The scandal of poor medical research. *Bmj* 1994; **308**(6924): 283-4.
29. Michael F. Drummond MJS, Karl Claxton, Greg L. Stoddart, and George W. Torrance. Methods for Economic Evaluation of Health Care programmes. 4th ed. ed: Oxford University Press; 2015.
30. Briggs A CK, Schulpher M. Decision Modelling for Health Economic Evaluation. : Oxford University Press; 2006.
31. Kaltenthaler E, Essat M, Tappenden P, Paisley S. Identification and review of cost-effectiveness model parameters: a qualitative study. *IntJTechnolAssessHealth Care* 2014; **30**(3): 333-40.
32. Kaltenthaler E, Tappenden P, Paisley S. Reviewing the evidence to inform the population of cost-effectiveness models within health technology assessments. *ValueHealth* 2013; **16**(5): 830-6.
33. Charrois TL. Systematic reviews: what do you need to know to get started? *Can J Hosp Pharm* 2015; **68**(2): 144-8.
34. Beller EM, Glasziou PP, Altman DG, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLOS Medicine* 2013; **10**(4): e1001419.
35. Glandville J, Lefebvre C, Wright K, Editors. ISSG Search Filter Resource [internet]. 2008 [updated 2020 February 4]. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed 24 Aug 2020).
36. Paisley S. Classification of evidence in decision-analytic models of cost-effectiveness: a content analysis of published reports. *IntJTechnolAssessHealth Care* 2010; **26**(4): 458-62.
37. Glanville J LC, Manson P, Robinson S, Brbre I and Woods L, editors. ISSG Search Filter Resource [Internet]. updated 22 July 2024 2006. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed September 13 2024).

38. Andrew B, Jane N, Kate F, Graham M, Özge T, Elham S. Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ Global Health* 2019; **4**(Suppl 1): e001107.
39. Booth A. Clear and present questions: formulating questions for evidence based practice. *Library Hi Tech* 2006; **24**(3): 355-68.
40. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qual Health Res* 2012; **22**(10): 1435-43.
41. Cochrane. PICO ontology. http://linkeddata.cochrane.org/pico-ontology (accessed 14 June 2018).
42. Clinical research explained. Standard of Care. . 2024. https://viares.com/blog/clinical-research-explained/standard-of-care/ (accessed February 2025).
43. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 1: introduction. *Med Decis Making* 2013; **33**(5): 597-606.
44. NICE. NICE Health Technology Evaluations: The Manual. Process and methods. 2022. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741 (accessed 05/09/2023 2023).
45. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. *Evidence Based Medicine* 2016.
46. Health Technology Assessment process: Fundamentals. 2015. https://toolbox.eupati.eu/resources/health-technology-assessment-process-fundamentals/ (accessed February 2025).
47. Glanville J, Paisley S, Edditors. Searching for evidence for cost-effectiveness decisions. In: Shemilt C DM, ed. Evidence-Based Decisions and Economics : Health Care, Social Welfare, Education and Criminal Justice 2nd edition ed. Hoboken, NJ, USA: Wiley-Blackwell; 2010.
48. European Medicines Agency E. Esmya: new measures to minimise risk of rare but serious liver injury 2018.
49. Golder S, Loke YK, Wright K, Norman G. Reporting of Adverse Events in Published and Unpublished Studies of Health Care Interventions: A Systematic Review. *PLoS Med* 2016; **13**(9): e1002127.
50. CADTH. Guidelines for the Economic Evaluation of Health Technologies: Canada. 2017.
51. INESSS. Guidance Document for Submitting a Request to INESSS. 2018.
52. British National Formulary.
53. Gesundheit Bf. Wie Arzneimittelpreise entstehen und wie man sie senken kann. https://www.bundesgesundheitsministerium.de/arzneimittelpreise.html.
54. Confused about drug prices in Germany? This post will help you gain clarity. https://germanmarketaccesssimplified.com/confused-about-drug-prices-in-germany-this-guide-will-help-you-gain-clarity/.
55. Germany FIfDaMD. List of Reference Prices for Medicinal Products. https://www.bfarm.de/EN/Medicinal-products/Information-on-medicinal-products/Reference-Pricing/_node.html (accessed February 2025).
56. Daniel Mullins C, Onwudiwe NC, Branco de Araújo GT, et al. Guidance Document: Global Pharmacoeconomic Model Adaption Strategies. *Value in Health Regional Issues* 2014; **5**: 7-13.
57. Beauchemin M, Cohn E, Shelton RC. Implementation of Clinical Practice Guidelines in the Health Care Setting: A Concept Analysis. *ANS Adv Nurs Sci* 2019; **42**(4): 307-24.
58. NCCN Guidelines. https://www.nccn.org/guidelines/nccn-guidelines.
59. Sackett D, Straus S, Richardson W. Evidence-based medicine: how to practice and teach EBM. 2nd Ed. ed. Edinburgh: Churchill Livingstone; 2000.
60. El Instituto de Evaluación Tecnológica en Salud (IETS). Manual para la elaboración de evaluaciones económicas en salud. 2014.
61. Pharmaceutical Management Agency (PHARMAC). Prescription for Pharmacoeconomic Analysis - Methods for cost-utility analysis. 2015; **version 2.2**.

62. Tandvårds- och läkemedelsförmånsverkets allmänna råd (TLV). Ändring i Tandvårds- och läkemedelsförmånsverkets allmänna råd (TLVAR 2003:2) om ekonomiska utvärderingar. 2017.
63. Institute NHC. Guideline for economic evaluations in healthcare. 2016.
64. PBAC. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee. 2016.
65. PHARMAC. Pharmaceutical Management Agency of New Zealand (PHARMAC): Prescription for pharmacoeconomic analysis, 2015.
66. SMC. Guidance to manufacturers for completion of New Product Assessment Form (NPAF)2022. https://www.scottishmedicines.org.uk/media/2770/guidance-on-npaf.doc (accessed.
67. Health Intervention and Technology Assessment Program (Thailand): Ministry of Public Health guidelines for health technology assessment in Thailand (2nd ed). . 2014.
68. EuroQol Group. EQ-5D. https://euroqol.org/eq-5d-instruments/ (accessed 26 April 2018).
69. Szende A, Janssen B, Cabases J. Self-reported population health: an international perspective based on EQ-5D. 2014.
70. British Health Foundation. Data Science Centre https://bhfdatasciencecentre.org/ (accessed February 2025).
71. European Cystic Fibrosis Society Patient Registry. https://www.ecfs.eu/projects/ecfs-patient-registry/project.
72. National Cancer Institute. Surveillance, Epidemiology, and End Results Program. https://seer.cancer.gov/statistics/ (accessed February 2025).
73. Papaioannou D, Brazier J, Paisley S. NICE DSU Technical Support Document 9: The identification, review and synthesis of health state utility values from the literature. Sheffield, 2010.
74. Hirt J, Nordhausen T, Appenzeller-Herzog C, Ewald H. Citation tracking for systematic literature searching: A scoping review. *Res Synth Methods* 2023; **14**(3): 563-79.
75. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet* 2000; **356**(9237): 1228-31.
76. Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. *Health TechnolAssess* 2003; **7**(34): iii, ix-51.
77. Allison JJ, Kiefe CI, Weissman NW, Carter J, Centor RM. THE ART AND SCIENCE OF SEARCHING MEDLINE TO ANSWER CLINICAL QUESTIONS Finding the Right Number of Articles. *Int J Technol Assess Health Care* 1999; **15**(2): 281-96.
78. Fraser C, Murray A, Burr J. Identifying observational studies of surgical interventions in MEDLINE and EMBASE. *BMC Med Res Methodol* 2006; **6**(1): 41-.
79. Golder S, McIntosh HM, Duffy S, Glanville J. Developing efficient search strategies to identify reports of adverse effects in medline and embase. *Health Info Libr J* 2006; **23**(1): 3-12.
80. Raftery J, Roderick P, Stevens A. Potential use of routine databases in health technology assessment. *Health Technol Assess* 2005; **9**(20): III-92.
81. Philips Z, Ginnelly L, Sculpher M, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health TechnolAssess* 2004; **8**(36): iii-xi, 1.
82. University Library. Systematic Reviews + Evidence Synthesis. What is a search filter or hedge? https://guides.library.ucsc.edu/c.php?g=1440266&p=10697345.
83. ISSG IISS-G. Search Filter Resource https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed 19.07. 2018).
84. Health Information Research Unit MU. Hedges Project. https://hiru.mcmaster.ca/hiru/HIRU_Hedges_home.aspx (accessed 19.07. 2018).
85. Cochrane. Cochrane Library. http://www.cochranelibrary.com/ (accessed 22.07.2018 2018).

86. Centre for Reviews and Dissemination UoY. CRD Database. https://www.crd.york.ac.uk/CRDWeb/ (accessed 22.07. 2018).
87. KSR Evidence.
88. Derwent Drug File.
89. International Pharmaceutical Abstracts.
90. US National Library of Medicine NIoH. TOXLINE. https://toxnet.nlm.nih.gov/newtoxnet/toxline.htm (accessed 22.07. 2018).
91. INAHTA. International HTA database. https://www.inahta.org/hta-database/ (accessed 6 November 2020).
92. Center for the Evaluation of Value and Risk in Health (CEVR) TU. CEA Registry : Cost-Effectiveness Analysis Registry. http://healtheconomics.tuftsmedicalcenter.org/cear4/SearchingtheCEARegistry/SearchtheCEARegistry.aspx (accessed 22.07. 2018).
93. ScHARRHUD. (ScHARR Health Utilities Database). https://www.scharrhud.org/ (accessed 22.07. 2018).
94. Cooper NJ, Sutton AJ, Ades AE, Paisley S, Jones DR. Use of evidence in economic decision models: practical issues and methodological challenges. *Health Econ* 2007; **16**(12): 1277-86.
95. Gasparyan AY, Ayvazyan L, Blackmore H, Kitas GD. Writing a narrative biomedical review: considerations for authors, peer reviewers, and editors. *Rheumatol Int* 2011; **31**(11): 1409-17.
96. Whittemore R, Knafl K. The integrative review: updated methodology. *J Adv Nurs* 2005; **52**(5): 546-53.
97. Sutton A, Clowes M, Preston L, Booth A. Meeting the review family: exploring review types and associated information retrieval requirements. *Health Info Libr J* 2019; **36**(3): 202-22.
98. Relevo R, Balshem H. Chapter5: Finding Evidence for Comparing Medical Interventions. Methods guide for effectiveness and comparative effectiveness reviews Agency for Health Care Research and Quality (AHRQ); 2014.
99. Kugley S, Wade A, Thomas J, et al. Searching for studies: a guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews* 2017; **13**(1): 1-73.
100. Harris JL, Booth A, Cargo M, et al. Cochrane Qualitative and Implementation Methods Group guidance series-paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. *J Clin Epidemiol* 2018; **97**: 39-48.
101. EUnetHTA. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness: European network for Health Technology Assessment, 2019.
102. Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health* 2011; **14**(4): 429-37.
103. Aromataris E, Munn ZE. JBI Manual for Evidence Synthesis. 2020. https://synthesismanual.jbi.global.
104. Dobbins M. Rapid Review Guidebook. Hamilton, ON, Canada: ational Collaborating Centre for Methods and Tools, 2017.
105. Greenhalgh T, Wong G, al. e. The RAMESES Project. 2013-2020. http://www.ramesesproject.org/Standards_and_Training_materials.php.
106. Vortal; H. 2019. http://vortal.htai.org/?q=sure-info.
107. Kaunelis D, Glanville J. Cost and economic evaluation. Summarized Research in Information Retrieval for HTA (SuRE Info). 2020. Retrieved from http://vortal.htai.org/?q=node/336.

108. Dequen P, Sutton AJ, Scott DA, Abrams KR. Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery. *Value Health* 2014; **17**(4): 416-23.
109. Simmonds M, Elliott JH, Synnot A, Turner T. Living Systematic Reviews. *Methods Mol Biol* 2022; **2345**: 121-34.
110. Elliott JH, Turner T, Clavisi O, et al. Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. *PLOS Medicine* 2014; **11**(2): e1001603.
111. Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. *J Clin Epidemiol* 2017; **91**: 47-53.
112. Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. Introduction-the why, what, when, and how. *J Clin Epidemiol* 2017; **91**: 23-30.
113. Simmonds M, Salanti G, McKenzie J, Elliott J. Living systematic reviews: 3. Statistical methods for updating meta-analyses. *J Clin Epidemiol* 2017; **91**: 38-46.
114. Thomas J, Noel-Storr A, Marshall I, et al. Living systematic reviews: 2. Combining human and machine effort. *J Clin Epidemiol* 2017; **91**: 31-7.
115. Millard T, Synnot A, Elliott J, Green S, McDonald S, Turner T. Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Systematic Reviews* 2019; **8**(1): 325.
116. Wright J, Walwyn R. Literature search methods for an overview of reviews ('umbrella' reviews or 'review of reviews') CILIP Health Libraries Group Conference; 2016 15th - 16th Sept; Scarborough, UK; 2016.
117. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implementation science : IS* 2010; **5**: 56-.
118. Booth A, Noyes J, Flemming K, et al. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions: Integrate-HTA; 2016.
119. Saul JE, Willis CD, Bitz J, Best A. A time-responsive tool for informing policy making: rapid realist review. *Implementation Science* 2013; **8**(1): 103.
120. Duncan V, Waldron T, Groot G. Working with a librarian on a realist review: The RAMESES Project, 2017.
121. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy* 2005; **10 Suppl 1**: 21-34.
122. Booth A, Briscoe S, Wright JM. The "realist search": A systematic scoping review of current practice and reporting. *Research Synthesis Methods* 2020; **11**(1): 14-35.
123. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. *BMJ Global Health* 2019; **4**(Suppl 1): e000893.
124. Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC Medical Research Methodology* 2013; **13**(1): 118.
125. Smythe E, Spence D. Re-Viewing Literature in Hermeneutic Research. *International Journal of Qualitative Methods* 2012; **11**(1): 12-25.
126. Greenhalgh T, A'Court C, Shaw S. Understanding heart failure; explaining telehealth - a hermeneutic systematic review. *BMC Cardiovasc Disord* 2017; **17**(1): 156.
127. Li L, Smith HE, Atun R, Tudor Car L. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database Syst Rev* 2019; **3**(3): Mr000041.
128. Li L, Smith HE, Atun R, Car LT. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database of Systematic Reviews* 2019; (3).

129. Haynes RB, McKibbon KA, Wilczynski NL, Walter SD, Werre SR. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *Bmj* 2005; **330**(7501): 1179.
130. Lefebvre C, Manheim E, Glanville J. Chapter 6: Searching for studies. In Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 6.0. 2019.
131. Briscoe S, Bethel A, Rogers M. Conduct and reporting of citation searching in Cochrane systematic reviews: A cross-sectional study. *Res Synth Methods* 2020; **11**(2): 169-80.
132. Robinson KA, Dunn AG, Tsafnat G, Glasziou P. Citation networks of related trials are often disconnected: implications for bidirectional citation searches. *Journal of Clinical Epidemiology* 2014; **67**(7): 793-9.
133. Bramer WM. Reference checking for systematic reviews using Endnote. *Journal of the Medical Library Association* 2018; **106**(4): 542-6.
134. Janssens ACJW, Gwinn M. Novel citation-based search method for scientific literature: application to meta-analyses. *BMC Medical Research Methodology* 2015; **15**(1): 84.
135. Small H. Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. *Journal of the American Society for Information Science (pre-1986)* 1973; **24**(4): 265.
136. Hirt J, Nordhausen T, Fuerst T, Ewald H, Appenzeller-Herzog C. Guidance on terminology, application, and reporting of citation searching: the TARCiS statement. *BMJ* 2024; **385**: e078384.
137. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* 2015; **4**(1): 5.
138. Abstrackr. http://abstrackr.cebm.brown.edu/account/login (accessed 24 Aug 2020).
139. Colandr. https://www.colandrapp.com/signin (accessed 24 Aug 2020).
140. DistillerAI. https://www.evidencepartners.com/distiller-ai/ (accessed 24 Aug 2020).
141. EPPI-Reviewer. https://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4& (accessed 24 Aug 2020).
142. Rayyan. https://rayyan.qcri.org/welcome (accessed 24 Aug 2020).
143. RobotAnalyst. http://www.nactem.ac.uk/robotanalyst/ (accessed 24 Aug 2020).
144. Stansfield C, apos, Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. *Res Synth Methods* 2017; **8**(3): 355-65.
145. Shemilt I, Simon A, Hollands GJ, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Res Synth Methods* 2014; **5**(1): 31-49.
146. Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. *Res Synth Methods* 2017; **8**(3): 355-65.
147. McGowan BS. Using Text Mining Tools to Inform Search Term Generation: An Introduction for Librarians. *Johns Hopkins University Press* 2021; **21**(3): 603 - 18.
148. Mozelius P, Humble N. On the Use of Generative AI for Literature Reviews : an Exploration of Tools and Techniques. Porto: 23rd European Conference on Research Methodology for Business and Management Studies Porto : Academic Conferences and Publishing International Limited

Proceedings of the 23rd European Conference on Research Methodology for Business and Management Studies Porto : Academic Conferences and Publishing International Limited; 2024.
149. Stansfield C, Dickson K, Bangpan M. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Syst Rev* 2016; **5**(1): 191.
150. Jenkins M. Evaluation of methodological search filters—a review. *Health Information & Libraries Journal* 2004; **21**(3): 148-63.

151. Waffenschmidt S, Navarro-Ruan T, Hobson N, Hausner E, Sauerland S, Haynes RB. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Res Synth Methods* 2020; **11**(5): 617-26.
152. Taljaard M, McDonald S, Nicholls SG, et al. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *J Clin Epidemiol* 2020; **124**: 75-84.
153. Tummers M, van Hoorn R, Levering C, Booth A, van der Wilt GJ, Kievit W. Optimal search strategies for identifying moderators and predictors of treatment effects in PubMed. 2019.
154. Rader T, Mann M, Stansfield C, Cooper C, Sampson M. Methods for documenting systematic review searches: a discussion of common issues. *Research Synthesis Methods* 2014; **5**(2): 98-115.
155. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology* 2012; **12**(1): 181.
156. Churchill R, Lasserson T, Chandler J, et al. Standards for the reporting of new Cochrane Intervention Reviews. Methodological Expectations of Cochrane Intervention Reviews: Cochrane; 2020.
157. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535.
158. Wong G, Greenhalgh T, Westhorp G, Buckingham J. RAMESES publication standards: meta-narrative reviews. *BMS Medicine* 2013; **11**(20).
159. Booth A. "Brimful of STARLITE": toward standards for reporting literature searches. *Journal of the Medical Library Association : JMLA* 2006; **94**(4): 421-e205.
160. Rethlefsen M, Kirtley S, Waffenschmidt S, et al. PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. March 20, 2019 2019. https://doi.org/10.31219/osf.io/sfc38 (accessed 24 Aug 2020).
161. Ara Roberta, Brazier John, Peasgood Tessa, Suzy P. The Identification, Review and Synthesis of Health State Utility Values from the Literature. *PharmacoEconomics* 2017; **35**(S1): 43-55.
162. Brazier J, Ara R, Azzabi I. Identification, review and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. . *Value in Health* 2019; **22**(3): 267-75.
163. Golder S, Peryer G, Loke YK. Overview: comprehensive and carefully constructed strategies are required when conducting searches for adverse effects data. *J Clin Epidemiol* 2019; **113**: 36-43.
164. Thielen FW, Van Mastrigt G, Burgers LT, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). *Expert Review of Pharmacoeconomics & Outcomes Research* 2016; **16**(6): 705-21.
165. Weinstein MC, O'Brien B, Hornberger J, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2003; **6**(1): 9-17.
166. EUnetHTA. HTA Core Model® EUnetHTA Domain 4 - Clinical effectiveness (EUnetHTA Joint Action 2, Work Package 8. HTA Core Model, version 3.0 (PDF) 2016: Methodology/Where to find information? p. 146
167. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.0 [Updated August 2023]. 08/2023 2019. https://training.cochrane.org/handbook/current (accessed 25/01/2024 2020).
168. Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. *Pharmacoeconomics* 2016; **34**(6): 597-608.
169. Sharma D, Aggarwal AK, Downey LE, Prinja S. National Healthcare Economic Evaluation Guidelines: A Cross-Country Comparison. *Pharmacoecon Open* 2021; **5**(3): 349-64.

170. Hjelmgren J, Berggren F, Andersson F. Health economic guidelines--similarities, differences and some implications. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2001; **4**(3): 225-50.
171. Bracco A, Krol M. Economic evaluations in European reimbursement submission guidelines: current status and comparisons. *Expert Rev Pharmacoecon Outcomes Res* 2013; **13**(5): 579-95.
172. Mathes T, Jacobs E, Morfeld J-C, Pieper D. Methods of international health technology assessment agencies for economic evaluations- a comparative analysis. *BMC Health Services Research* 2013; **13**(1): 371.
173. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018; **18**(1): 143.
174. Armstrong R, Hall BJ, Doyle J, Waters E. 'Scoping the scope' of a cochrane review. *Journal of Public Health* 2011; **33**(1): 147-50.
175. GEAR. Guidelines comparison: what can I learn from the existing health economic evaluation guidelines? 2019. https://gear4health.com/gear/health-economic-evaluation-guidelines (accessed December 5 2024).
176. ISPOR. Pharmacoeconomic Guidelines Around The World. 2021. https://tools.ispor.org/peguidelines/ (accessed 3rd May 2021).
177. Iran Ministry of Health IFaDA, National Committee for Selecting and Registering Medicine. Regulation of submitting applications for medicine with regard to development of economic evaluation, 2014.
178. Health; SFOoP. Operationalisierung der Kriterien "Wirksamkeit, Zweckmässigkeit und Wirtschaftlichkeit". Grundlagendokument. In: (BAG) BfG, editor.; 2022.
179. Cuba Ministry of Public Health. Guía metodológica para la evaluación económica en salud. In: Escuela Nacional de Salud Publica AdE, editor. La Habana, Cuba: Gálvez González, A.M.; 2003.
180. Brazil Ministério da Saúde. Diretrizes Metodologicas, Diretriz de Avaliação Econômica. In: Secretaria de Ciência TeIE, editor. Brasília 2014.
181. Czech Pharmacoeconomic Society. Doporučené postupy České farmakoekonomické společnosti (ČFES) pro zdravotně-ekonomická hodnocení v ČR. Duben, 2016.
182. Hungarian pharmaceutical authority. Professional healthcare guideline of the methodology of health technology assessment, English version *The journal of the Hungarian pharmaceutical authority for doctors and pharmacists* 2017; **67**(1).
183. Rodríguez J, Suárez R, G G. Guía para la Conducción de Estudios de Evaluación Económica para la Actualización del Cuadro Básico y Catálogo de Insumos del Sector Salud en México. In: General CdS, editor. México; 2017.
184. Russian Federation Ministry of Health. Guidelines for conducting a comparative clinical and economic evaluation of drugs. In: Center for Healthcare Quality Assessment and Control, editor.; 2016.
185. China Guidelines for Pharmacoeconomic Evaluations. In: G L, ed.; 2020.
186. Columbia Institute for Health Technology Assessment. Manual para la elaboración de evaluaciones económicas en salud. In: Moreno Viscaya M MMA, Castro Jaramillo HE., editor. Bogotá; 2014.
187. Ministry of Health (Israel)—Pharmaceutical Administration. Guidelines for the submission of a request to include a pharmaceutical product in the national list of health services—Appendix D: instructions for performing an economic assessment. 2010.
188. Korea Health Insurance Review and Assessment Service. Guidelines for Economic Evaluation of Pharmaceuticals in Korea. Seoul. 2005.

189. Ministry of Health of the Slovak Republic. The announcement concerning pharmaco-economic analysis of drugs (Vyhláška Ministerstva zdravotníctva Slovenskej republikyopodrobnostiachfarmako-ekonomického rozboru lieku č. 422/2011 Z. z.). . 2011.
190. Slovenia HIIo. Rules on Classification of Medicines on the List, accepted by the Health Insurance Institute of Slovenia. 2010.
191. Behmane D, Lambot K, Irs A, Steikunas N. Baltic guideline for economic evaluation of pharmaceuticals (Pharmacoeconomic Analysis). 2002.
192. Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU TECHNICAL SUPPORT DOCUMENT 5: EVIDENCE SYNTHESIS IN THE BASELINE NATURAL HISTORY MODEL. 2011; 2011.
193. NICE. NICE Health Technology Evaluations: The Manual. Process and methods. 2022. https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741.
194. PHARMAC. Guidelines for funding applications to PHARMAC. New Zealand: Pharmaceutical Management Agency, 2017.
195. Walter E, Zehetmayr S. Guidelines on Health Economic Evaluation Vienna, Austria: Institute for Pharmaeconomic Research, 2006.
196. Cleemput I, Neyt M, Van De Sande S, Thiry N. Belgian guidelines for economic evaluations and budget impact analyses: second edition. Report 183C. *Belgian Health Care Knowledge Centre (KCE)* 2012.
197. Agency for Quality and Accreditation in Health Care. The Croatian Guideline for Health Technology Assessment Process and Reporting. In: Department for Development RaHTA, editor. Zagreb, Croatia; 2011.
198. Alban A, Gyldmark M, Pedersen A, Søgaard J. The Danish Approach to Standards for Economic Evaluation Methodologies. *Pharmacoeconomics* 1997; **Dec;12**(6): 627.
199. Elsisi GH, Kaló Z, Eldessouki R, et al. Recommendations for Reporting Pharmacoeconomic Evaluations in Egypt. *Value Health Reg Issues* 2013; **2**(2): 319-27.
200. EUnetHTA. Methods for Health Economic Evaluations - A guideline based on current practices in Europe European network for Health Technology Assessment, 2015.
201. HILA. Preparing a health economic evaluation to be attached to the application for reimbursement status and wholesale price for a medicinal product 2017.
202. IQWiG. General methods for the assessment of the relation of benefits to costs. Version of 6.1 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2022.
203. Indonesian Health Technology Assessment Committee. (InaHTAC). Ministry of Health Republic of Indonesia. Health Technology Assessment guideline. . 2017; (Jakarta).
204. HIQA. Health Information and Quality Authority (HIQA): Guidelines for the Economic Evaluation of Health Technologies in Ireland, 2020.
205. AIFA. Linea guida per la compilazione del capitolo 9 del Dossier a supporto della domanda di rimborsabilità e prezzo (May 2020). 2019. https://www.aifa.gov.it/documents/20142/1307543/2021.01.22_estratto_linee_guida_sezione_E.pdf (accessed May 2021).
206. Japan National Institute of Public Health. Center for Outcomes Research and Economic Evaluation for Health, C2H: Guideline for preparing cost-effectiveness evaluation to the Central Social Insurance Medical Council (Version 2.0). 2019.
207. Malaysia MoH. Pharmacoeconomic Guideline for Malaysia In: Division PS, editor.; 2012.
208. Zorginstituut Nederland. Guideline for economic evaluations in healthcare. 2016.
209. Norwegian Medicines Agency. Guidelines for the submission of documentation for single technology assessment of pharmaceuticals: Statens legemiddelverk, 2018.

210. Polish Agency for Health Technology Assessment. Health Technology Assessment Guidelines Warsaw, 2016.
211. Silva AE, Pinto CG, Sampaio C, Pereira JA, Drummond M, Trindade R. Guidelines for economic drug evaluation studies, 1998.
212. Singapore Agency for Care Effectiveness. Drug evaluation methods and process guide. 2018.
213. Matsos M. Guidelines for Pharmacoeconomic Submissions. Health Republic of South Africa. 2012.
214. López-Bastida J, Oliva J, Antoñanzas F, et al. Spanish recommendations on economic evaluation of health technologies. *The European Journal of Health Economics* 2010; **11**(5): 513-20.
215. Taiwan Society for Pharmacoeconomic and Outcomes Research (TaSPOR). Guidelines of Methodological Standards for Pharmacoeconomic Evaluations. 2006.
216. AMCP. AMCP Format for Formulary Submissions. Guidance on Submission of Pre-approval and Post-approval Clinical and Economic Information and Evidence, 2019.
217. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes: Oxford university press; 2015.
218. Gold MR. Cost-effectiveness in health and medicine: Oxford university press; 1996.
219. Wilkinson T, Sculpher MJ, Claxton K, et al. The International Decision Support Initiative Reference Case for Economic Evaluation: An Aid to Thought. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2016; **19**(8): 921-8.
220. Kobelt G. Health economics: an introduction to economic evaluation. London: Office of Health Economics, 2013.
221. Finc J. Making choices in health: who guide to cost effectiveness analysis. : Geneva: World Health Organization, 2003.
222. Matsos M. Guidelines for Pharmacoeconomic Submissions In: Health RoSA, editor.; 2012.
223. Collège des Économistes de la Santé. French guidelines for the economic evaluation of health care technologies, 2004.
224. Sculpher M, Fenwick E, Claxton K. Assessing Quality in Decision Analytic Cost-Effectiveness Models. *PharmacoEconomics* 2000; **17**(5): 461-77.
225. Sonnenberg AF, Roberts SM, Tsevat BJ, Wong LJ, Barry LM, Kent LD. Toward a Peer Review Process for Medical Decision Analysis Models. *Medical Care* 1994; **32**(7 Suppl): JS52-JS.
226. Soto J. HEALTH ECONOMIC EVALUATIONS USING DECISION ANALYTIC MODELING Principles and PracticesUtilization of a Checklist to Their Development and Appraisal. *J of Inter Tech of Health Care* 2002; **18**(1): 94-111.
227. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling Good Research Practices—Overview: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value in Health* 2012; **15**(6): 796-803.
228. Briggs AH, Weinstein MC, Fenwick EAL, Karnon J, Sculpher MJ, Paltiel AD. Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health* 2012; **15**(6): 835-42.
229. University of Exeter LibGuides. Grey Literature: What is Grey Literature? . 2023. https://libguides.exeter.ac.uk/c.php?g=670055&p=4756572 (accessed 14 September 2023).
230. Cleemput I, Neyt M, Van De Sande S, Thiry N. Belgian guidelines for economic evaluations and budget impact analyses: second edition. Brussels: Belgian Health Care Knowledge Centre(KCE), 2012.
231. Institute of Medicine Committee on Standards for Systematic Reviews of Comparative Effectiveness R. In: Eden J, Levit L, Berg A, Morton S, eds. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC): National Academies Press (US) Copyright 2011 by the National Academy of Sciences. All rights reserved.; 2011.

232. McDonagh MS, Jonas DE, Gartlehner G, et al. Methods for the drug effectiveness review project. *BMC medical research methodology* 2012; **12**: 140-.
233. NICE. Document B Company Evidence Submission, 2013.
234. NICE. Single technology appraisal: User guide for company evidence submission template, 2015.
235. Nissen T, Wynn R. The clinical case report: a review of its merits and limitations. *BMC Research Notes* 2014; **7**(1): 264.
236. Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *The Lancet* 2014; **383**(9912): 156-65.
237. Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014; **383**(9912): 166-75.
238. Belgisch Centrum voor Farmacotherapeutische Informatie
239. (APR-DRG) APRDRG. MZG : Publicaties : All-Patient-Refined-Diagnosis-Related-Groups-APR DGR.
240. FSBI Center for Expertise and Quality Control of Medical Care. Guidelines for conducting a comparative clinical and economic evaluation of drugs. 2016.
241. Brazier J, Ara R, Azzabi I, et al. Identification, Review, and Use of Health State Utilities in Cost-Effectiveness Models: An ISPOR Good Practices for Outcomes Research Task Force Report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2019; **22**(3): 267-75.
242. Zechmeister-Koss I, Schnell-Inderst P, Zauner G. Appropriate evidence sources for populating decision analytic models within health technology assessment (HTA): a systematic review of HTA manuals and health economic guidelines. *MedDecisMaking* 2014; **34**(3): 288-99.
243. Pearsons A, Neubeck L, Hendriks JM, Hanson CL. Justification, rationale and methodological approaches to realist reviews. *European Journal of Cardiovascular Nursing* 2022; **22**(1): 107-12.
244. distillersr. The Difference Between a Rapid Review vs Systematic Review. https://www.distillersr.com/resources/systematic-literature-reviews/the-difference-between-a-rapid-review-vs-a-systematic-review.
245. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. *BMC Medical Research Methodology* 2011; **11**(1): 100.
246. Green J, Thorogood N. Qualitative methods for health research. 2018.
247. NICE. National Institute for Clinical Excellence (NICE): Guide to the Methods of Technology Appraisal 2013. 4/4/2013 2013. nice.org.uk/process/pmg9.
248. Papaioannou D, Sutton A, Carroll C, Booth A, Wong R. Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal* 2010; **27**(2): 114-22.
249. Libraries FAU. Nursing - Literature Reviews & Systematic Reviews: Backwards & Forward Reference Searching. https://libguides.fau.edu/c.php?g=966176&p=7013756 (accessed 10th June 2022).
250. Hirt J, Nordhausen T, Fuerst T, Ewald H, Appenzeller-Herzog C. The TARCiS statement: Guidance on terminology, application, and reporting of citation searching. *medRxiv* 2023: 2023.10.25.23297543.
251. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. *Qual Quant* 2018; **52**(4): 1893-907.
252. Glaser BG, Strauss AL, Press S. The Discovery of Grounded Theory: Strategies for Qualitative Research: Aldine Publishing Company; 1967.
253. Bryman A. How many qualitative interviews is enough? . In: SE; B, R; E, eds. How Many Qualitative Interviews is Enough? Expert Voices and Early Career Reflections on Sampling and Cases in Qualitative Research. Southampton: ESRC National Centre for Research Methods, University of Southampton; 2012: 18-20.

254. Dey I. Grounding Grounded Theory: Guidelines for Qualitative Inquiry. San Diego: Academic Press; 1999.
255. Birks M, Mills J. Grounded Theory: A Practical Guide. 2nd Ed. ed. London: Sage; 2015.
256. Starks H, Trinidad SB. Choose your method: a comparison of phenomenology, discourse analysis, and grounded theory. *Qual Health Res* 2007; **17**(10): 1372-80.
257. Guest G, Bunce A, Johnson L. How Many Interviews Are Enough?:An Experiment with Data Saturation and Variability. *Field Methods* 2006; **18**(1): 59-82.
258. Pirolli P, Card S. Information foraging. *Psychological Review* 1999; **106**((4)): 643-75.
259. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *Jama* 2016; **316**(10): 1093-103.
260. Hartling L, Guise J-M, Hempel S, et al. Fit for purpose: perspectives on rapid reviews from end-user interviews. *Syst Rev* 2017; **6**(1): 32-.
261. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009; **26**(2): 91-108.
262. Watt A, Cameron A, Sturm L, et al. Rapid versus full systematic reviews: validity in clinical practice? *ANZ J Surg* 2008; **78**(11): 1037-40.
263. Schünemann HJ, Moja L. Reviews: Rapid! Rapid! Rapid! …and systematic. *Syst Rev* 2015; **4**(1): 4.
264. Yin RK. Case Study Research Design and Methods. Thousand Oaks, CA: Sage; 2014.
265. Baxter P, Jack S. Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report* 2008; **13**(4): 544-59.
266. Eisenhardt KM, Graebner ME. Theory building from cases: Opportunities and challenges. *Academy of management journal* 2007; **50**(1): 25-32.
267. Gerring J. What Is a Case Study and What Is It Good for? *American Political Science Review* 2004; **98**(2): 341-54.
268. Cooper C, Varley-Campbell J, Booth A, Britten N, Garside R. Systematic review identifies six metrics and one method for assessing literature search effectiveness but no consensus on appropriate use. *Journal of clinical epidemiology* 2018; **99**: 53-63.
269. SNDS. Système National des Données de Santé. 2021. https://www.snds.gouv.fr/SNDS/Accueil (accessed October 02 2021).
270. Eraksoy MBZ, T; Zivadinov;. Time for Change – Evolution of Real-world Evidence Outcome Measures in Multiple Sclerosis Exemplified by Fingolimod. *Eur Neurol Rev* 2014; (9(2)): 136-42.
271. Lister J, Paisley S, Carroll C, Tappenden P. Empirical Testing of Alternative Search Methods to Retrieve Utility Values for Health Economic Modelling. *Pharmacoeconomics* 2024.
272. Archer R, Tappenden P, Ren S, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (including a review of TA140 and TA262): clinical effectiveness systematic review and economic model. *Health Technol Assess* 2016; **20**(39): 1-326.
273. Ford AC, Moayyedi P, Hanauer SB. Ulcerative colitis. *BMJ* 2013; **346**: f432.
274. MDS. Manufacturer Submission of Evidence: Infliximab (Remicade). *Submission to NICE* 2014.
275. MDS. Manufacturer Submission of Evidence: Golimumab (Simponi). *Submission to NICE* 2014.
276. Alrubaiy L, Cheung WY, Dodds P, et al. Development of a short questionnaire to assess the quality of life in Crohn's disease and ulcerative colitis. *J Crohns Colitis* 2015; **9**(1): 66-76.
277. Armuzzi A, Tarallo M, Lucas J, et al. The association between disease activity and patient-reported outcomes in patients with moderate-to-severe ulcerative colitis in the United States and Europe. *BMC gastroenterology* 2020; **20**(1): 18.

278. Beilman CL, Thanh NX, Ung V, et al. Real-Life Treatment Paradigms Show Adalimumab Is Cost-Effective for the Management of Ulcerative Colitis. *Can J Gastroenterol Hepatol* 2016; **2016**: 5315798.
279. Van Assche G, Peyrin-Biroulet L, Sturm A, et al. Burden of disease and patient-reported outcomes in patients with moderate to severe ulcerative colitis in the last 12 months - Multicenter European cohort study. *Dig Liver Dis* 2016; **48**(6): 592-600.
280. Biedermann L, Dubinsky MC, Vermeire S, et al. Health-Related Quality of Life Outcomes With Tofacitinib Treatment in Patients With Ulcerative Colitis in the Open-Label Extension Study, OCTAVE Open. *Inflammatory Bowel Diseases* 2022; **15**: 15.
281. Burisch J, Hart A, Oortwijn A, et al. P293 Insights from patients with Ulcerative Colitis on disease burden: Findings from a real-world survey in Europe. *Journal of Crohn's and Colitis* 2022; **16**(Supplement_1): i324-i5.
282. CADTH. Pharmacoeconomic review report - tofacitinib (Xeljanz). 2019.
283. Casellas F, López Vivancos J, Malagela JR. Previous experience and quiality of life in patients with inflammatory bowel disease during relapse. *Rev Esp Enferm Dig* 2003; **95**(7): 476-9, 1-5.
284. Dulai PS, Jairath V. A Microsimulation Model to Project the 5-Year Impact of Using Hyperbaric Oxygen Therapy for Ulcerative Colitis Patients Hospitalized for Acute Flares. *Dig Dis Sci* 2021; **66**(11): 3740-52.
285. Gherardi A, Roze S, Kuijvenhoven J, Ghatnekar O, Yip Sonderegger YL. Budesonide with multi-matrix technology as second-line treatment for ulcerative colitis: evaluation of long-term cost-effectiveness in the Netherlands. *Journal of Medical Economics* 2018; **21**(9): 869-77.
286. Gibson PR, Vaizey C, Black CM, et al. Relationship between disease severity and quality of life and assessment of health care utilization and cost for ulcerative colitis in Australia: a cross-sectional, observational study. *Journal of Crohn's & colitis* 2013; **8**(7): 598-606.
287. Hagelund LM, Elkjaer Stallknecht S, Jensen HH. Quality of life and patient preferences among Danish patients with ulcerative colitis - results from a survey study. *Current Medical Research & Opinion* 2020; **36**(5): 771-9.
288. Hernandez L, Kuwabara H, Shah A, et al. Cost-Effectiveness Analysis of Vedolizumab Compared with Other Biologics in Anti-TNF-Naïve Patients with Moderate-to-Severe Ulcerative Colitis in Japan. *Pharmacoeconomics* 2020; **38**(1): 69-84.
289. Kawalec P, Stawowczyk E. Relationship between physician-based assessment of disease activity, quality of life, and costs of ulcerative colitis in Poland. *Przeglad gastroenterologiczny* 2018; **13**(1): 61-8.
290. Kuruvilla K, Osler T, Hyman NH. A comparison of the quality of life of ulcerative colitis patients after IPAA vs ileostomy. *Diseases of the Colon & Rectum* 2012; **55**(11): 1131-7.
291. Leidl R, Reitmeir P, König H-H, Stark R. The performance of a value set for the EQ-5D based on experienced health states in patients with inflammatory bowel disease. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2012; **15**(1): 151-7.
292. Richards DM, Hughes SA, Irving MH, Scott NA. Patient quality of life after successful restorative proctocolectomy is normal. *Colorectal Dis* 2001; **3**(4): 223-6.
293. Sardesai A, Dignass A, Quon P, et al. Cost-effectiveness of tofacitinib compared with infliximab, adalimumab, golimumab, vedolizumab and ustekinumab for the treatment of moderate to severe ulcerative colitis in Germany. *J Med Econ* 2021; **24**(1): 279-90.
294. Scott FI, Luo M, Shah Y, et al. Identification of the Most Cost-effective Position of Vedolizumab Among the Available Biologic Drugs for the Treatment of Ulcerative Colitis. *J Crohns Colitis* 2020; **14**(5): 575-87.
295. Swinburn P, Elwick H, Bean K, et al. PTU-127 The impact of surgery on health related quality of life in ulcerative colitis. *Gut* 2012; **61**(Suppl 2): A237-A.
296. NICE. Ozanimod for treating moderate to severe ulcerative colitis committee papers TA828. 2022.

297. van der Valk ME, Mangen MJ, Severs M, et al. Comparison of Costs and Quality of Life in Ulcerative Colitis Patients with an Ileal Pouch-Anal Anastomosis, Ileostomy and Anti-TNFalpha Therapy. *Journal of Crohn's & colitis* 2015; **9**(11): 1016-23.
298. Woehl A, Hawthorne A, McEwan P. The relation between disease activity, quality of life and health utility in patients with ulcerative colitis. Gut; 2008: BMJ PUBLISHING GROUP BRITISH MED ASSOC HOUSE, TAVISTOCK SQUARE, LONDON WC1H …; 2008. p. A153-A.
299. Min Ho PY, Hu W, Lee YY, et al. Health-related quality of life of patients with inflammatory bowel disease in Singapore. *Intest Res* 2019; **17**(1): 107-18.
300. McLeod RS, Churchill DN, Lock AM, Vanderburgh S, Cohen Z. Quality of life of patients with ulcerative colitis preoperatively and postoperatively. *Gastroenterology* 1991; **101**(5): 1307-13.
301. Richardson G, Griffiths AM, Miller V, Thomas AG. Quality of life in inflammatory bowel disease: a cross-cultural comparison of English and Canadian children. *Journal of Pediatric Gastroenterology & Nutrition*; **32**(5): 573-8.
302. Tsai HH, Punekar YS, Morris J, Fortun P. A model of the long-term cost effectiveness of scheduled maintenance treatment with infliximab for moderate-to-severe ulcerative colitis. *Alimentary Pharmacology & Therapeutics* 2008; **28**(10): 1230-9.
303. Vaizey CJ, Gibson PR, Black CM, et al. Disease status, patient quality of life and healthcare resource use for ulcerative colitis in the UK: an observational study. *Frontline Gastroenterol* 2014; **5**(3): 183-9.
304. Alam MF, Longo M, Cohen D, et al. Infliximab versus ciclosporin in steroid resistant acute severe ulcerative colitis: a model-based cost-utility analysis of data from CONSTRUCT pragmatic trial. *BMC health services research* 2023; **23**(1): 226.
305. Kuruvilla K, Osler T, Hyman NH. A comparison of the quality of life of ulcerative colitis patients after IPAA vs ileostomy. 2012; **55**: 1131-7.
306. Richards DM, Hughes SA, Irving MH, Scott NA. Patient quality of life after successful restorative proctocolectomy is normal. 2001; **3**: 223-6.
307. van der Valk ME, Mangen MJ, Dijkstra G, van Bodegraven AA, Fidder H, de Jong DJ. Is There a Difference in Quality of Life and Costs Between Ulcerative Colitis Patients with a Pouch or an Ileostomy? Gastroenterology Conference: Digestive Disease Week 19–22 May 2012; San Diego, CA.
308. Affairs USDoV. Cost-effectiveness Analysis. https://www.herc.research.va.gov/INCLUDE/PAGE.ASP?id=cost-effectiveness-analysis (accessed 21/07 2024).
309. Bates MJ. The design of browsing and berrypicking techniques for the online search interface. *Online Review* 1989; **13**(5): 407-24.
310. Pirolli P. Information Foraging. In: Liu L, ÖZsu MT, eds. Encyclopedia of Database Systems. Boston, MA: Springer US; 2009: 1485-90.
311. NICE. Upadacitinib for treating moderately to severely active ulcerative colitis TA856 [ID3953] committee papers.; 2022.
312. NICE. Filgotinib for treating moderately to severely active ulcerative colitis TA792 [ID3736] Committee Papers. 2022.
313. NICE. Ustekinumab for treating moderately to severely active ulcerative colitis TA633 [ID1511] committee papers. . 2020.
314. NICE. Tofacitinib for previously treated active ulcerative colitis TA547 [ID1218] committee papers. . 2018.
315. NICE. Vedolizumab for treating moderately to severely active ulcerative colitis [TA342] committee papers. . 2015.
316. Archer R, Tappenden P, Ren S, et al. Infliximab, adalimumab and golimumab for treating moderately to severely active ulcerative colitis after the failure of conventional therapy (including a review

of TA140 and TA262): clinical effectiveness systematic review and economic model. *Health Technology Assessment (Winchester, England)*; **20**(39): 1-326.
317. Arseneau KO, Sultan S, Provenzale DT, et al. Do patient preferences influence decisions on treatment for patients with steroid-refractory ulcerative colitis? *Clin Gastroenterol Hepatol* 2006; **4**(9): 1135-42.
318. Chaudhary MA, Fan T. Cost-Effectiveness of Infliximab for the Treatment of Acute Exacerbations of Ulcerative Colitis in the Netherlands. *Biologics in therapy* 2013; **3**: 45-60.
319. Punekar YS, Hawkins N. Cost-effectiveness of infliximab for the treatment of acute exacerbations of ulcerative colitis. *European Journal of Health Economics* 2010; **11**(1): 67-76.
320. Waljee AK, Higgins PD, Waljee JF, et al. Perceived and actual quality of life with ulcerative colitis: a comparison of medically and surgically treated patients. *Am J Gastroenterol* 2011; **106**(4): 794-9.
321. Poole CD, Connolly MP, Nielsen SK, Currie CJ, Marteau P. A comparison of physician-rated disease severity and patient reported outcomes in mild to moderately active ulcerative colitis. *J Crohns Colitis* 2010; **4**(3): 275-82.
322. Bashir NS, Walters TD, Griffiths AM, Ungar WJ. An Assessment of the Validity and Reliability of the Pediatric Child Health Utility 9D in Children with Inflammatory Bowel Disease. *Children (Basel, Switzerland)* 2021; **8**(5).
323. Yamabe K, Liebert R, Flores N, Pashos CL. Health-related quality of life outcomes and economic burden of inflammatory bowel disease in Japan. *ClinicoEconomics and outcomes research : CEOR* 2019; **11**: 221-32.
324. Stawowczyk E, Kawalec P. A Systematic Review of the Cost-Effectiveness of Biologics for Ulcerative Colitis. *PharmacoEconomics* 2018; **36**(4): 419-34.
325. Stawowczyk E, Kawalec P, Pilc A. Cost-utility analysis of 1-year treatment with adalimumab/standard care and standard care alone for ulcerative colitis in Poland. *European journal of clinical pharmacology* 2016; **72**(11): 1319-25.
326. Stawowczyk E, Kawalec P, Pilc A. Cost-Effectiveness Analysis of 1-Year Treatment with Golimumab/Standard Care and Standard Care Alone for Ulcerative Colitis in Poland. *PloS one* 2016; **11**(8): e0160444.
327. Stawowczyk E, Kawalec P, Pilc A. Cost-Utility Analysis of Infliximab with Standard Care versus Standard Care Alone for Induction and Maintenance Treatment of Patients with Ulcerative Colitis in Poland. *Pharmacotherapy* 2016; **36**(5): 472-81.
328. Feagan BG, Patel H, Colombel JF, et al. Effects of vedolizumab on health-related quality of life in patients with ulcerative colitis: results from the randomised GEMINI 1 trial. *Alimentary Pharmacology & Therapeutics*; **45**(2): 264-75.
329. Schneider Y, Saumoy M, Cohen-Mekelburg S, Steinlauf AF, Scherl EJ. The Cost-Effectiveness of Vedolizumab for Inflammatory Bowel Disease: A Review of the Current Literature. *Gastroenterology & hepatology* 2016; **12**(10): 617-21.
330. Tappenden P, Ren S, Archer R, et al. A Model-Based Economic Evaluation of Biologic and Non-Biologic Options for the Treatment of Adults with Moderately-to-Severely Active Ulcerative Colitis after the Failure of Conventional Therapy. *Pharmacoeconomics*; **34**(10): 1023-38.
331. Lindsay JO, Bergman A, Patel AS, Alesso SM, Peyrin-Biroulet L. Systematic review: the financial burden of surgical complications in patients with ulcerative colitis. *Alimentary pharmacology & therapeutics* 2015; **41**(11): 1066-78.
332. Huoponen S, Blom M. A Systematic Review of the Cost-Effectiveness of Biologics for the Treatment of Inflammatory Bowel Diseases. *PloS one* 2015; **10**(12): e0145087.

333. Ung V, Thanh NX, Wong K, et al. Real-life treatment paradigms show infliximab is cost-effective for management of ulcerative colitis. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2014; **12**(11): 1871-8.e8.
334. Park KT, Tsai R, Perez F, Cipriano LE, Bass D, Garber AM. Cost-effectiveness of early colectomy with ileal pouch-anal anastamosis versus standard medical therapy in severe ulcerative colitis. *Annals of Surgery* 2012; **256**(1): 117-24.
335. Connolly MP, Nielsen SK, Currie CJ, Marteau P, Probert CSJ, Travis SPL. An economic evaluation comparing concomitant oral and topical mesalazine versus oral mesalazine alone in mild-to-moderately active ulcerative colitis based on results from randomised controlled trial. *Journal of Crohn's & colitis* 2009; **3**(3): 168-74.
336. Arai K, Koganei K, Kimura H, et al. Incidence and outcome of complications following restorative proctocolectomy. *Am J Surg* 2005; **190**(1): 39-42.
337. Actis GC, Fadda M, David E, Sapino A. Colectomy rate in steroid-refractory colitis initially responsive to cyclosporin: a long-term retrospective cohort study. *BMC Gastroenterol* 2007; **7**: 13.
338. Al-Darmaki A, Hubbard J, Seow CH, et al. Clinical Predictors of the Risk of Early Colectomy in Ulcerative Colitis: A Population-based Study. *Inflamm Bowel Dis* 2017; **23**(8): 1272-7.
339. Constant BD, Baldassano R, Kirsch J, Mitchel EB, Stein R, Albenberg L. Tofacitinib Salvage Therapy for Children Hospitalized for Corticosteroid- and Biologic-Refractory Ulcerative Colitis. *Journal of Pediatric Gastroenterology and Nutrition* 2022; **75**(6).
340. Gower-Rousseau C, Dauchet L, Vernier-Massouille G, et al. The natural history of pediatric ulcerative colitis: a population-based cohort study. *Am J Gastroenterol* 2009; **104**(8): 2080-8.
341. Gustavsson A, Järnerot G, Hertervig E, et al. Clinical trial: colectomy after rescue therapy in ulcerative colitis - 3-year follow-up of the Swedish-Danish controlled infliximab study. *Aliment Pharmacol Ther* 2010; **32**(8): 984-9.
342. Manetti N, Bagnoli S, Rogai F, et al. Disease Course and Colectomy Rate of Ulcerative Colitis: A Follow-up Cohort Study of a Referral Center in Tuscany. *Inflamm Bowel Dis* 2016; **22**(8): 1945-53.
343. Misra R, Askari A, Faiz O, Arebi N. Colectomy Rates for Ulcerative Colitis Differ between Ethnic Groups: Results from a 15-Year Nationwide Cohort Study. *Can J Gastroenterol Hepatol* 2016; **2016**: 8723949.
344. Molnár T, Farkas K, Nyári T, Szepes Z, Nagy F, Wittmann T. Response to first intravenous steroid therapy determines the subsequent risk of colectomy in ulcerative colitis patients. *J Gastrointestin Liver Dis* 2011; **20**(4): 359-63.
345. Mocciaro F, Renna S, Orlando A, et al. Cyclosporine or infliximab as rescue therapy in severe refractory ulcerative colitis: early and long-term data from a retrospective observational study. *J Crohns Colitis* 2012; **6**(6): 681-6.
346. Solberg IC, Lygren I, Jahnsen J, et al. Clinical course during the first 10 years of ulcerative colitis: results from a population-based inception cohort (IBSEN Study). *Scand J Gastroenterol* 2009; **44**(4): 431-40.
347. Lynch RW, Down C, Roughton M, Arnott ID, Group UIAS. OC-165 No increase in surgical complication in patients treated with rescue therapy for acute severe ulcerative colitis: data from the UK IBD audit. *Gut* 2012; **61**(Suppl 2): A71-A.
348. Burisch J, Lophaven S, Munkholm P, Langholz E. Surgery, cancer and mortality among patients with ulcerative colitis diagnosed 1962-1987 and followed until 2017 in a Danish population-based inception cohort. *Aliment Pharmacol Ther* 2022; **55**(3): 339-49.
349. Dai N, Haidar O, Askari A, Segal JP. Colectomy rates in ulcerative colitis: A systematic review and meta-analysis. *Dig Liver Dis* 2023; **55**(1): 13-20.

350. Eriksson C, Cao Y, Rundquist S, et al. Changes in medical management and colectomy rates: a population-based cohort study on the epidemiology and natural history of ulcerative colitis in Örebro, Sweden, 1963-2010. *Aliment Pharmacol Ther* 2017; **46**(8): 748-57.
351. Hoie O, Wolters FL, Riis L, et al. Low colectomy rates in ulcerative colitis in an unselected European cohort followed for 10 years. *Gastroenterology* 2007; **132**(2): 507-15.
352. Kurti Z, Gonczi L, Lakatos L, et al. Epidemiology, Treatment Strategy, Natural Disease Course and Surgical Outcomes of Patients with Ulcerative Colitis in Western Hungary - A Population-based Study Between 2007 and 2018: Data from the Veszprem County Cohort. *J Crohns Colitis* 2023; **17**(3): 352-60.
353. Feuerstein JD, Curran T, Alosilla M, Cataldo T, Falchuk KR, Poylin V. Mortality Is Rare Following Elective and Non-elective Surgery for Ulcerative Colitis, but Mild Postoperative Complications Are Common. *Dig Dis Sci* 2018; **63**(3): 713-22.
354. de Silva S, Ma C, Proulx MC, et al. Postoperative complications and mortality following colectomy for ulcerative colitis. *Clin Gastroenterol Hepatol* 2011; **9**(11): 972-80.
355. Ordás I, Domènech E, Mañosa M, et al. Post-operative morbidity and mortality of a cohort of steroid refractory acute severe ulcerative colitis: Nationwide multicenter study of the GETECCU ENEIDA Registry. *Am J Gastroenterol* 2018; **113**(7): 1009-16.
356. Shawihdi M, Dodd S, Kallis C, et al. Nationwide improvement in outcomes of emergency admission for ulcerative colitis in England, 2005-2013. *Aliment Pharmacol Ther* 2019; **50**(2): 176-92.
357. Blommestein HM, Franken MG, Uyl-de Groot CA. A practical guide for using registry data to inform decisions about the cost effectiveness of new cancer drugs: lessons learned from the PHAROS registry. *Pharmacoeconomics* 2015; **33**(6): 551-60.
358. AHRQ Methods for Effective Health Care. In: Gliklich RE, Dreyer NA, Leavy MB, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Rockville (MD): Agency for Healthcare Research and Quality (US); 2014.
359. Competence Network IBD. 2023. http://www.kompetenznetz-ced.de.
360. NCT03162549. CgI. CorEvitas Inflammatory Bowel Disease (IBD) Registry. . https://clinicaltrials.gov/study/NCT03162549?term=NCT03162549&rank=1 (accessed August 2023).
361. ENEIDA Registry. 2023. https://www.encepp.eu/encepp/viewResource.htm;jsessionid=X_Zb7okekO53xGQudyq1_HcWwvRdGWcfUHTxSuSr3E7_QaIN4Azj!-1783076594?id=50619.
362. EPIMAD Registry. 2019. https://clinicaltrials.gov/study/NCT02392221?a=2.
363. SWIBREG Registry. 2023. https://www.swibreg.se/.
364. Swiss IBD Registry. 2022.
365. UK IBD Registry. 2023.
366. Eronen H, Oksanen P, Jussila A, Huhtala H, Helavirta I, Ilus T. Long-term outcomes of patients with acute severe ulcerative colitis treated with cyclosporine rescue therapy. *Scand J Gastroenterol* 2023; **58**(5): 483-8.
367. Monstad IL, Solberg IC, Cvancarova M, et al. Outcome of Ulcerative Colitis 20 Years after Diagnosis in a Prospective Population-based Inception Cohort from South-Eastern Norway, the IBSEN Study. *J Crohns Colitis* 2021; **15**(6): 969-79.
368. Núñez L, Mesonero F, Rodríguez de Santiago E, Die J, Albillos A, López-Sanromán A. High incidence of surgery and initiation of medical therapies after colectomy for ulcerative colitis or inflammatory bowel disease unclassified. *Gastroenterol Hepatol* 2023; **46**(5): 369-75.
369. Senanayake SM, Fernandopulle AN, Niriella MA, et al. The long-term outcomes of a cohort of Sri Lankan patients with ulcerative colitis: a retrospective study at two national referral centers and review of literature. *Clin Exp Gastroenterol* 2013; **6**: 195-200.

370. Targownik LE, Singh H, Nugent Z, Bernstein CN. The epidemiology of colectomy in ulcerative colitis: results from a population-based cohort. *Am J Gastroenterol* 2012; **107**(8): 1228-35.
371. Worley G, Nordenvall C, Askari A, et al. Restorative surgery after colectomy for ulcerative colitis in England and Sweden: observations from a comparison of nationwide cohorts. *Colorectal Dis* 2018; **20**(9): 804-12.
372. Tappenden P, Carroll C, Hamilton J, et al. Cabozantinib and vandetanib for unresectable locally advanced or metastatic medullary thyroid cancer: a systematic review and economic model. *Health Technol Assess* 2019; **23**(8): 1-144.
373. Pavlidis E, Sapalidis K, Chatzinikolaou F, Kesisoglou I. Medullary thyroid cancer: molecular factors, management and treatment. *Rom J Morphol Embryol* 2020; **61**(3): 681-6.
374. Roman S, Lin R, Sosa JA. Prognosis of medullary thyroid carcinoma. *Cancer* 2006; **107**(9): 2134-42.
375. Revised American Thyroid Association Guidelines for the Management of Medullary Thyroid Carcinoma. *Thyroid®* 2015; **25**(6): 567-610.
376. Elisei R, Schlumberger MJ, Müller SP, et al. Cabozantinib in progressive medullary thyroid cancer. *J Clin Oncol* 2013; **31**(29): 3639-46.
377. Sanofi Genzyme. Vandetanib for Treating Unresectable Locally Advanced or Metastatic Medullary Thyroid Cancer (ID56): Evidence Submission to NICE. Oxford: Genzyme Therapeutics, 2017.
378. Sanofi Genzyme. Vandetanib for Treating Unresectable Locally Advanced or Metastatic Medullary Thyroid Cancer (ID56): Evidence Submission to NICE - Appendices. Oxford: Genzyme Therapeutics, 2017.
379. Fordham BA, Kerr C, de Freitas HM, et al. Health state utility valuation in radioactive iodine-refractory differentiated thyroid cancer. *Patient Prefer Adherence* 2015; **9**: 1561-72.
380. All Wales Therapeutics and Toxicology Centre. AWMSG Secretariat Assessment Report. Vandetanib (Caprelsa®) 100 mg and 300 mg Film-coated Tablets. . Penarth: All Wales Medicines Strategy Group, 2014.
381. All Wales Therapeutics and Toxicology Centre. AWMSG Secretariat Assessment Report. Cabozantinib

(Cometriq®) 20 mg and 80 mg Hard Capsules. Penarth: All Wales Medicines Strategy Group, 2014.
382. Ipsen Ltd. Cabozantinib and Vandetanib for Treating Unresectable Locally Advanced or Metastatic Medullary Thyroid Cancer. Company Evidence Submission to the National Institute for Health and Care Excellence. Slough: Ipsen Ltd., 2017.
383. NICE. Vandetanib for treating medullary thyroid cancer (TA550). 2018.
384. NICE. Lenvatinib and sorafenib for treating differentiated thyroid cancer after radioactive iodine (TA535). 2018.
385. NICE. Cabozantinib for treating medullary thyroid cancer (TA516). 2018.
386. NICE. Selpercatinib for treating advanced thyroid cancer with RET alterations (TA742). 2021.
387. NICE. Cabozantinib for previously treated advanced differentiated thyroid cancer unsuitable for or refractory to radioactive iodine (TA928). 2023.
388. Scottish Medicines Consortium (SMC). Detailed Advice – Sorafenib 200 mg Film-coated Tablets (Nexavar®) SMC No. (1055/15). Glasgow, 2015.
389. Scottish Medicines Consortium (SMC). Cabozantinib 20 mg and 80 mg Hard Capsules (Cometriq®) SMC No. (1022/15). Glasgow, 2015.
390. Scottish Medicines Consortium (SMC). Detailed Advice – Lenvatinib 4 mg and 10 mg Hard Capsules (Lenvima®) SMC No. 1179/16. Glasgow, 2016.
391. Beusterien KM, Szabo SM, Kotapati S, et al. Societal preference values for advanced melanoma health states in the United Kingdom and Australia. *Br J Cancer* 2009; **101**(3): 387-9.

392. Kerr C, Fordham B, de Freitas HM, Pelletier CL, Lloyd A. Health State Utility Valuation in Radio-Iodine Refractory Differentiated Thyroid Cancer (RR-DTC). *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2014; **17**(7): A646.
393. Baek HS, Ha J, Kim K, et al. Cost-Utility Analysis of Early Detection with Ultrasonography of Differentiated Thyroid Cancer: A Retrospective Study on a Korean Population. *Endocrinol Metab (Seoul)* 2024.
394. Wells SA, Jr., Robinson BG, Gagel RF, et al. Vandetanib in patients with locally advanced or metastatic medullary thyroid cancer: a randomized, double-blind phase III trial. *J Clin Oncol* 2012; **30**(2): 134-41.
395. ClinicalTrials.gov. NCT00410761. An Efficacy Study Comparing ZD6474 to Placebo in Medullary Thyroid Cancer. 2006. https://clinicaltrials.gov/ct2/show/NCT00410761 (accessed 25 April 2024).
396. ClinicalTrials.gov. NCT00704730. Efficacy of XL184 (Cabozantinib) in Advanced Medullary Thyroid Cancer. 2008. https://clinicaltrials.gov/ct2/show/NCT00704730 (accessed 25 April 2024).
397. Schlumberger M, Elisei R, Müller S, et al. Overall survival analysis of EXAM, a phase III trial of cabozantinib in patients with radiographically progressive medullary thyroid carcinoma. *Ann Oncol* 2017; **28**(11): 2813-9.
398. Toraih E, Hussein M, Anker A, et al. Survival Outcomes of Medullary Thyroid Cancer With and Without Amyloid Deposition. *Endocr Pract* 2024; **30**(4): 311-8.
399. Liu CQ, Shen CK, Du YX, et al. Survival outcome and optimal candidates of primary tumor resection for patients with metastatic medullary thyroid cancer. *J Clin Endocrinol Metab* 2024.
400. Ramos HE, Hecht F, Berdelou A, et al. Long-term follow-up and safety of vandetanib for advanced medullary thyroid cancer. *Endocrine* 2021; **71**(2): 434-42.
401. Mathiesen JS, Kroustrup JP, Vestergaard P, et al. Survival and Long-Term Biochemical Cure in Medullary Thyroid Carcinoma in Denmark 1997-2014: A Nationwide Study. *Thyroid* 2019; **29**(3): 368-77.
402. Al-Qurayshi Z, Khadra H, Chang K, Pagedar N, Randolph GW, Kandil E. Risk and survival of patients with medullary thyroid cancer: National perspective. *Oral Oncol* 2018; **83**: 59-63.
403. Randle RW, Bates MF, Schneider DF, Sippel RS, Pitt SC. Survival in patients with medullary thyroid cancer after less than the recommended initial operation. *J Surg Oncol* 2018; **117**(6): 1211-6.
404. Papachristos AJ, Nicholls LE, Mechera R, et al. Management of Medullary Thyroid Cancer: Patterns of Recurrence and Outcomes of Reoperative Surgery. *The Oncologist* 2023; **28**(12): 1064-71.
405. Khadra H, Al-Qurayshi Z, Kandil E. Risk and Survival of Patients with Medullary Thyroid Cancer: National Perspective. *Journal of the American College of Surgeons* 2017; **225**(4): S73.
406. Baranauskas Z, Valuckas K, Smailyte G. Combined treatment and survival of medullary thyroid carcinoma patients. *Open Medicine* 2010; **5**(4): 426-30.
407. Buffet C, Leboulleux S, Kraeber-Bodéré F, et al. Cardiac Metastasis from Medullary Thyroid Cancers with Long-Term Survival under Vandetanib. *Eur Thyroid J* 2021; **10**(6): 517-22.
408. Tao Z, Deng X, Guo B, Ding Z, Fan Y. Subgroup analysis of steadily increased trends in medullary thyroid carcinoma incidence and mortality in the USA, 2000–2020: a population-based retrospective cohort study. *Endocrine-Related Cancer* 2024; **31**(5): e230319.
409. Zuhur SS, Ozturk BO, Keskin U, et al. Disease-free survival and the prognostic factors affecting disease-free survival in patients with medullary thyroid carcinoma: a multicenter cohort study. *Endocrine* 2024.
410. Censi S, Galuppini F, Clausi C, et al. Tumor Grade and Molecular Characteristics Associated with Survival in Sporadic Medullary Thyroid Carcinoma. *Thyroid* 2024; **34**(2): 177-85.
411. An Y, Lu J, Hu M, Cao Q. A prediction model for the 5-year, 10-year and 20-year mortality of medullary thyroid carcinoma patients based on lymph node ratio and other predictors. *Front Surg* 2022; **9**: 1044971.

412. Duval M, Ferreira CV, Marmitt L, et al. An Undetectable Postoperative Calcitonin Level Is Associated with Long-Term Disease-Free Survival in Medullary Thyroid Carcinoma: Results of a Retrospective Cohort Study. *Thyroid* 2023; **33**(1): 82-90.
413. Fernando A, Shanaz F, Nye D. HEREDITARY MEDULLARY THYROID CARCINOMA IN MULTIPLE FAMILY MEMBERS ASSOCIATED WITH RET MUTATION; A CASE SERIES HIGHLIGHTING THE IMPORTANCE OF EARLY GENETIC SCREENING IN ALTERING ASSOCIATED MORBIDITY AND MORTALITY. *Thyroid®* 2022; **32**(S1): P-1-A-135.
414. Raue F, Bruckner T, Frank-Raue K. Similar Stage-dependent Survival and Outcome in Sporadic and Hereditary Medullary Thyroid Carcinoma. *J Clin Endocrinol Metab* 2021; **106**(9): e3582-e91.
415. Lorusso L, Romei C, Piaggi P, et al. Ca19.9 Positivity and Doubling Time Are Prognostic Factors of Mortality in Patients with Advanced Medullary Thyroid Cancer with No Evidence of Structural Disease Progression According to Response Evaluation Criteria in Solid Tumors. *Thyroid* 2021; **31**(7): 1050-5.
416. Sahli ZT, Canner JK, Zeiger MA, Mathur A. Association between age and disease specific mortality in medullary thyroid cancer. *Am J Surg* 2021; **221**(2): 478-84.
417. Randle RW, Balentine CJ, Leverson GE, et al. Trends in the presentation, treatment, and survival of patients with medullary thyroid cancer over the past 30 years. *Surgery* 2017; **161**(1): 137-46.
418. Esfandiari NH, Hughes DT, Yin H, Banerjee M, Haymart MR. The effect of extent of surgery and number of lymph node metastases on overall survival in patients with medullary thyroid cancer. *J Clin Endocrinol Metab* 2014; **99**(2): 448-54.
419. Romei C, Cosci B, Bottici V, al. e. Significantly lower survival rate in medullary thyroid carcinoma patients with RET somatic mutations: A 10-year follow-up. *Thyroid®* 2007; **17**(s1): S-45-S-136.
420. Zheng-Pywell R, Chen H, Abraham D. Greater mortality seen in medullary thyroid cancer patients with post-surgical progression to metastases compared to stable metastases. Annals of Surgical Oncology; 2020 2020/02/01: Society of Surgical Oncology SSO 2020 – International Conference on Surgical Cancer Care; 2020. p. 1-230.
421. Schlumberger M, Elisei R, Müller S, et al. Final overall survival analysis of EXAM, an international, double-blind, randomized, placebo-controlled phase III trial of cabozantinib (Cabo) in medullary thyroid carcinoma (MTC) patients with documented RECIST progression at baseline. *Journal of Clinical Oncology* 2015; **33**(15_suppl): 6012-.
422. SEER. Surveillance, Epidemiology, and End Results Program. 2024. https://seer.cancer.gov/ (accessed 02 May 2024).
423. American College of Surgeons. National Cancer Database. 2024. https://www.facs.org/quality-programs/cancer-programs/national-cancer-database/ (accessed 02 May 2024).
424. American Thyroid Association. Medullary Thyroid Carcinoma Surveillance Study. 2020. https://www.thyroid.org/professionals/partner-relations/medullary-thyroid-carcinoma-registry-consortium/ (accessed 02 May 2024).
425. DATHYRCA. The Danish Thyroid Cancer Database. 2024. https://www.dahanca.dk/ (accessed 02 May 2024).
426. GPOH. GPOH-MET Register: Prospektives, multizentrisches Register zur Erfassung maligner endokriner Tumoren im Kindes- und Jugendalter. 2023. https://www.gpoh.de/studienportal/abgeschlossene_studien_register/gpoh_met_register/index_ger.html (accessed 02 May 2024).
427. Europe Rci. Surveillance of Rare Cancers in Europe (RARECARE). 2007. https://www.rarecarenet.eu/ (accessed 2024.
428. Locati L, Cavalieri S, Dal Maso L, et al. Rare thyroid malignancies in Europe: Data from the information network on rare cancers in Europe (RARECAREnet). *Oral Oncology* 2020; **108**: 104766.

429. Ahn HY, Chae JE, Moon H, Noh J, Park YJ, Kim SG. Trends in the Diagnosis and Treatment of Patients with Medullary Thyroid Carcinoma in Korea. *Endocrinol Metab (Seoul)* 2020; **35**(4): 811-9.
430. BNDMR. French National Registry for Rare Diseases. https://www.bndmr.fr/ (accessed 02 May 2024).
431. Würzburg. U. Registry for rare thyroid and parathyroid carcinoma. https://www.ukw.de/medizinische-klinik-i/endokrinologie/forschung/klinische-studien/ (accessed 02 May 2024).
432. RegistRare. RNMR: Italian National Rare Diseases Registry. https://www.registrare.org/schede-6-il_registro_nazionale_malattie_rare (accessed 02 May 2024).
433. RePER. Rare Diseases Patient Registry. https://registroraras.isciii.es/Comun/Inicio.aspx (accessed 02 May 2024).
434. SRSK. Swiss Rare Disease Registry. https://www.raredisease.ch/?cn-reloaded=1 (accessed 02 May 2024).
435. Stanford Research. CORDS Registry: Coordination of Rare Diseases at Sanford Registry. https://research.sanfordhealth.org/ (accessed 02 May 2024).
436. EuRRECa. European Registries for Rare Endocrine Conditions. https://endo-ern.eu/registries/eurreca/ (accessed 02 May 2024).
437. Grasic Kuhar C, Lozar T, Besic N, Music Marolt M. Outcome of Patients with Locally Advanced Metastatic Medullary Thyroid Cancer and Induction Therapy with Tyrosine Kinase Inhibitors in Slovenia. *Advances in Therapy* 2021; **38**(12): 5684-99.
438. NICE. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer. Diagnostics guidance [DG58]. 2024. https://www.nice.org.uk/guidance/dg58 (accessed 07 June 2024).
439. Harnan S, Tappenden P, Cooper K, et al. Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer: a systematic review and economic analysis. 2019; **23**: 30.
440. NICE. DG34 Tumour profiling tests to guide adjuvant chemotherapy decisions in early breast cancer: evidence review, 2024.
441. Cancer Research UK. Breast cancer incidence (invasive) statistics - breast cancer incidence by gender and UK country. https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive#heading-Zero (accessed 06 Jun 2024).
442. Kalinsky KM, Barlow WE, Gralow JR, et al. Abstract GS2-07: Updated results from a phase 3 randomized clinical trial in participants (pts) with 1-3 positive lymph nodes (LN), hormone receptor-positive (HR+) and HER2-negative (HER2-) breast cancer (BC) with recurrence score (RS) ≤ 25 randomized to endocrine therapy (ET) +/- chemotherapy (CT): SWOG S1007 (RxPONDER). *Cancer Research* 2022; **82**(4_Supplement): GS2-07-GS2-.
443. Sestak I, Buus R, Cuzick J, et al. Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor–Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncology* 2018; **4**(4): 545-53.
444. Albain KS, Barlow WE, Shak S, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010; **11**(1): 55-65.
445. Verrill M, Wardley AM, Retzler J, et al. Health-related quality of life and work productivity in UK patients with HER2-positive breast cancer: a cross-sectional study evaluating the relationships between disease and treatment stage. *Health Qual Life Outcomes* 2020; **18**(1): 353.
446. Lidgren M, Wilking N, Jönsson B, Rehnberg C. Health related quality of life in different states of breast cancer. *Qual Life Res* 2007; **16**(6): 1073-81.

447. Farkkila N RR, Jahkola T, Sintonen H, Hanninen J, Taari K, et al. Health state utilities in breast cancer. ISPOR 14TH ANNUAL EUROPEAN CONGRESS; 2011 November; 2011. p. PA459.
448. Yousefi M, Najafi S, Ghaffari S, Mahboub-Ahari A, Ghaderi H. Comparison of SF-6D and EQ-5D Scores in Patients With Breast Cancer. *Iran Red Crescent Med J* 2016; **18**(5): e23556.
449. Naik H, Howell D, Su S, et al. EQ-5D Health Utility Scores: Data from a Comprehensive Canadian Cancer Centre. *The Patient - Patient-Centered Outcomes Research* 2017; **10**(1): 105-15.
450. Bewersdorf JP, Patel KK, Goshua G, et al. Cost-effectiveness of liposomal cytarabine/daunorubicin in patients with newly diagnosed acute myeloid leukemia. *Blood* 2022; **139**(11): 1766-70.
451. Campbell HE, Epstein D, Bloomfield D, et al. The cost-effectiveness of adjuvant chemotherapy for early breast cancer: A comparison of no chemotherapy and first, second, and third generation regimens for patients with differing prognoses. *Eur J Cancer* 2011; **47**(17): 2517-30.
452. Exact Sciences. Tumour profiling tests to guide adjuvant chemotherapy decisions in lymph node positive early breast cancer. Company evidence submission. London: Exact Sciences, 2023.
453. Cytel CEA report (produced on behalf of Agendia). Cost-effectiveness analysis of the MammaPrint gene profiling test to guide adjuvant chemotherapy use in early breast cancer in the United Kingdom, 2023.
454. Younis T, Rayson D, Skedgel C. Adjuvant chemotherapy in breast cancer: is TC a cost-effective regimen compared to AC? *Cancer Research* 2009; **69**(2_Supplement): 6110-.
455. Saptaningsih AB, Nikmah USa, Aryandono T, Dwiprahasto I, At Thobari J. Cost-Effectiveness Analysis of Two Chemotherapy Regiments FAC vs. Taxane for Operable Breast Cancer Patients in Indonesia. 2022.
456. Vaidya A, Vaidya P, Both B, Brew-Graves C, Bulsara M, Vaidya JS. Health economics of targeted intraoperative radiotherapy (TARGIT-IORT) for early breast cancer: a cost-effectiveness analysis in the United Kingdom. 2017.
457. Thomas RJ, Williams M, Marshall C, Glen J, Callam M. The total hospital and community UK costs of managing patients with relapsed breast cancer. *Br J Cancer* 2009; **100**(4): 598-600.
458. Suri G, Mistry R, Young K, et al. Cost effectiveness of ribociclib plus letrozole versus palbociclib plus letrozole for the treatment of post-menopausal women with hormone receptor-positive (HR+), human epidermal growth factor receptor 2-negative (HER2-) advanced or metastatic breast cancer from a us private third-party payer perspective. *Value in Health* 2017; **20(9)**(A436).
459. Zhu L, Wang M, Luo X, et al. Pharmacoeconomic evaluations of CDK4/6 inhibitors plus endocrine therapy for advanced hormone receptor-positive (HR+) and human epidermal growth factor receptor-2 negative (HER2-) breast cancer: a systematic review. *Ann Transl Med* 2022; **10**(4): 233.
460. NICE;. Advanced breast cancer: diagnosis and treatment (CG81), 2009.
461. Zhao JJ, Fong KY, Chan YH, et al. Indirect Treatment Comparison of First-Line CDK4/6-Inhibitors in Post-Menopausal Patients with HR+/HER2- Metastatic Breast Cancer. *Cancers (Basel)* 2023; **15**(18).
462. Hortobagyi GN, Stemmer SM, Burris HA, et al. Overall Survival with Ribociclib plus Letrozole in Advanced Breast Cancer. *N Engl J Med* 2022; **386**(10): 942-50.
463. Goetz MP, Toi M, Huober J, et al. Abemaciclib plus a nonsteroidal aromatase inhibitor as initial therapy for HR+, HER2- advanced breast cancer: final overall survival results of MONARCH 3. *Ann Oncol* 2024.
464. Finn RS, Rugo HS, Dieras VC, et al. Overall survival (OS) with first-line palbociclib plus letrozole (PAL+LET) versus placebo plus letrozole (PBO+LET) in women with estrogen receptor–positive/human epidermal growth factor receptor 2–negative advanced breast cancer (ER+/HER2− ABC): Analyses from PALOMA-2. *Journal of Clinical Oncology* 2022; **40**(17_suppl): LBA1003-LBA.
465. ESMO;. Matching-adjusted indirect treatment comparison of ribociclib and palbociclib as first-line treatments for HR+, HER2– ABC. 2017.

466. LLC) ARA. WebPlotDigitizer 5.0. 2024. https://automeris.io.
467. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2024. https://www.R-project.org/ (accessed July 2024).
468. Kimmick G, Pilehvari A, You W, Bonilla G, Anderson R. First- vs second-line CDK 4/6 inhibitor use for patients with hormone receptor positive, human epidermal growth-factor receptor-2 negative, metastatic breast cancer in the real world setting. *Breast Cancer Res Treat* 2024.
469. Coutinho-Almeida J, Silva AS, Redondo P, Rodrigues PP, Ferreira A. CDK4/6 inhibitors and endocrine therapy in the treatment of metastatic breast cancer: A real-world and propensity score-adjusted comparison. *Cancer Treatment and Research Communications* 2024; **40**: 100818.
470. Meegdes M, Geurts SME, Erdkamp FLG, et al. Real-world time trends in overall survival, treatments and patient characteristics in HR+/HER2- metastatic breast cancer: an observational study of the SONABRE Registry. *Lancet Reg Health Eur* 2023; **26**: 100573.
471. Goyal RK, Chen H, Abughosh SM, Holmes HM, Candrilli SD, Johnson ML. Overall survival associated with CDK4/6 inhibitors in patients with HR+/HER2- metastatic breast cancer in the United States: A SEER-Medicare population-based study. *Cancer* 2023; **129**(7): 1051-63.
472. Zwakman M, Verberne LM, Kars MC, Hooft L, van Delden JJM, Spijker R. Introducing PALETTE: an iterative method for conducting a literature search for a review in palliative care. *BMC Palliat Care* 2018; **17**(1): 82-.
473. Schlosser RW, Wendt O, Bhavnani S, Nail-Chiwetalu B. Use of information-seeking strategies for developing systematic reviews and engaging in evidence-based practice: the application of traditional and comprehensive Pearl Growing. A review. *Int J Lang Commun Disord* 2006; **41**(5): 567-82.
474. Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implement Sci* 2010; **5**: 56.
475. Abrami PC, Borokhovski E, Bernard RM, et al. Issues in conducting and disseminating brief reviews of evidence. *Evidence & Policy* 2010; **6**(3): 371-89.
476. Harker J, Kleijnen J. What is a rapid review? A methodological exploration of rapid reviews in Health Technology Assessments. *Int J Evid Based Healthc* 2012; **10**(4): 397-410.
477. Watt A, Cameron A, Sturm L, et al. Rapid reviews versus full systematic reviews: an inventory of current methods and practice in health technology assessment. *Int J Technol Assess Health Care* 2008; **24**(2): 133-9.
478. Featherstone RM, Dryden DM, Foisy M, et al. Advancing knowledge of rapid reviews: an analysis of results, conclusions and recommendations from published review articles examining rapid reviews. *Systematic Reviews* 2015; **4**(1): 50.
479. Van de Velde S, De Buck E, Dieltjens T, Aertgeerts B. Medicinal use of potato-derived products: conclusions of a rapid versus full systematic review. *Phytother Res* 2011; **25**(5): 787-8.
480. Watt A, Cameron A, Sturm L, et al. Rapid versus full systematic reviews: validity in clinical practice? *ANZ journal of surgery* 2008; **78**(11): 1037-40.
481. Haby MM, Chapman E, Clark R, Barreto J, Reveiz L, Lavis JN. What are the best methodologies for rapid reviews of the research evidence for evidence-informed decision making in health policy and practice: a rapid review. *Health Research Policy and Systems* 2016; **14**(1): 83.
482. Opiyo N, Shepperd S, Musila N, et al. Comparison of alternative evidence summary and presentation formats in clinical guideline development: a mixed-method study. *PLoS One* 2013; **8**(1): e55067.
483. Finfgeld-Connett D, Johnson ED. Literature search strategies for conducting knowledge-building and theory-generating qualitative systematic reviews. *J Adv Nurs* 2013; **69**(1): 194-204.
484. Booth A. How much searching is enough? Comprehensive versus optimal retrieval for technology assessments. *Int J Technol Assess Health Care* 2010; **26**(4): 431-5.

485. Beyer FR, Wright K. Can we prioritise which databases to search? A case study using a systematic review of frozen shoulder management. *Health Info Libr J* 2013; **30**(1): 49-58.
486. Preston L, Carroll C, Gardois P, Paisley S, Kaltenthaler E. Improving search efficiency for systematic reviews of diagnostic test accuracy: an exploratory study to assess the viability of limiting to MEDLINE, EMBASE and reference checking. *Systematic Reviews* 2015; **4**(1): 82.
487. Ertaylan G, Le Cornet C, van Roekel EH, et al. A Comparative Study on the WCRF International/University of Bristol Methodology for Systematic Reviews of Mechanisms Underpinning Exposure-Cancer Associations. *Cancer Epidemiol Biomarkers Prev* 2017; **26**(11): 1583-94.
488. Chalmers I, Dickersin K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *Bmj* 1992; **305**(6857): 786-8.
489. Winkelstein W, Jr. The remarkable Archie: origins of the Cochrane Collaboration. *Epidemiology* 2009; **20**(5): 779.
490. Lister J, Paisley S, Tappenden P, Carroll C. Identification of Health State Utility Values for Health Economic Models: Empirical Testing of Alternative Search Methods. ISPOR Europe; 2023 12 - 15 November; Copenhagen; 2023.
491. Lister J, Paisley S, Tappenden P, Carroll C. A Scoping Review of Health Technology Assessment Agency (HTA) and Health Economic (HE) Modelling Guidelines for Identification of Health Economic Model Inputs ISPOR Europe; 2023 12 - 15 November; Copenhagen; 2023.
492. Aaron T. Mega databases, AI and semantic searches: magic buzz words or something real there? . 2024. https://www.iqwig.de/veranstaltungen/keynote-tay-mega-databases.pdf.
493. Aaron T. What can you actually find in a open citation index like OpenAlex? Implications for citation based tools based on it. 2022.
494. van den Bulk LM, Bouzembrak Y, Gavai A, Liu N, van den Heuvel LJ, Marvin HJ. Automatic classification of literature in systematic reviews on food safety using machine learning. *Current Research in Food Science* 2022; **5**: 84-95.
495. Kebede MM, Le Cornet C, Fortner RT. In‐depth evaluation of machine learning methods for semi‐automating article screening in a systematic review of mechanistic literature. *Research Synthesis Methods* 2023; **14**(2): 156-72.
496. Bolaños F, Salatino A, Osborne F, Motta E. Artificial intelligence for literature reviews: opportunities and challenges. *Artificial Intelligence Review* 2024; **57**(10): 259.
497. Walker VR, Schmitt CP, Wolfe MS, et al. Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr. *Environment international* 2022; **159**: 107025.
498. Marshall IJ, Noel‐Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. *Research synthesis methods* 2018; **9**(4): 602-14.
499. Van De Schoot R, De Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nature machine intelligence* 2021; **3**(2): 125-33.
500. Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys* 2023; **56**(2): 1-40.
501. Beltagy I, Lo K, Cohan A. SciBERT: A Pretrained Language Model for Scientific Text. Conference on Empirical Methods in Natural Language Processing; 2019; 2019.
502. Castelvecchi D. Can we open the black box of AI? *Nature News* 2016; **538**(7623): 20.

Appendix 1 Included publications – Review 3

Ref.	Publication	Review 1	Review 2
138	Abstrackr. http://abstrackr.cebm.brown.edu/account/login (accessed 24 Aug 2020).	✓	
111	Akl EA, Meerpohl JJ, Elliott J, Kahale LA, Schünemann HJ. Living systematic reviews: 4. Living guideline recommendations. Journal of clinical epidemiology 2017; 91: 47-53	✓	
161	Ara Roberta, Brazier John, Peasgood Tessa, Suzy P. The Identification, Review and Synthesis of Health State Utility Values from the Literature. *PharmacoEconomics* 2017; **35**(S1): 43-55.		✓
103	Aromataris E, Munn ZE. JBI Manual for Evidence Synthesis. 2020. https://synthesismanual.jbi.global	✓	
122	Booth A, Briscoe S, Wright JM. The "realist search": A systematic scoping review of current practice and reporting. *Research Synthesis Methods* 2020; **11**(1): 14-35.	✓	
124	Booth A, Harris J, Croot E, Springett J, Campbell F, Wilkins E. Towards a methodology for cluster searching to provide conceptual and contextual "richness" for systematic reviews of complex interventions: case study (CLUSTER). *BMC Medical Research Methodology* 2013; **13**(1): 118.	✓	
118	Booth A, Noyes J, Flemming K, et al. Guidance on choosing qualitative evidence synthesis methods for use in health technology assessments of complex interventions: Integrate-HTA; 2016.	✓	
133	Bramer WM. Reference checking for systematic reviews using Endnote. *Journal of the Medical Library Association* 2018; **106**(4): 542-6.	✓	
162	Brazier J, Ara R, Azzabi I. Identification, review and use of health state utilities in cost-effectiveness models: an ISPOR Good Practices for Outcomes Research Task Force Report. . *Value in Health* 2019; **22**(3): 267-75.		✓
131	Briscoe S, Bethel A, Rogers M. Conduct and reporting of citation searching in Cochrane systematic reviews: A cross-sectional study. *Res Synth Methods* 2020; **11**(2): 169-80.	✓	
4	Centre for Reviews and Dissemination UoY. Centre for Reviews and Dissemination. Systematic reviews: CDR's guidance for undertaking reviews in health care. York, 2009.	✓	✓
156	Churchill R, Lasserson T, Chandler J, et al. Standards for the reporting of new Cochrane Intervention Reviews. Methodological Expectations of Cochrane Intervention Reviews: Cochrane; 2020.	✓	
139	Colandr. https://www.colandrapp.com/signin (accessed 24 Aug 2020).	✓	
108	Dequen P, Sutton AJ, Scott DA, Abrams KR. Searching for indirect evidence and extending the network of studies for network meta-analysis: case study in venous thromboembolic events prevention following elective total knee replacement surgery. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2014; 17(4): 416-23.	✓	
140	DistillerAI. https://www.evidencepartners.com/distiller-ai/ (accessed 24 Aug 2020).	✓	
104	Dobbins M. Rapid Review Guidebook. Hamilton, ON, Canada: National Collaborating Centre for Methods and Tools, 2017	✓	
120	Duncan V, Waldron T, Groot G. Working with a librarian on a realist review: The RAMESES Project, 2017.	✓	

112	Elliott JH, Synnot A, Turner T, et al. Living systematic review: 1. Introduction-the why, what, when, and how. Journal of clinical epidemiology 2017; 91: 23-30.	✓	
110	Elliott JH, Turner T, Clavisi O, et al. Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap. PLOS Medicine 2014; 11(2): e1001603.	✓	
141	EPPI-Reviewer. https://eppi.ioe.ac.uk/CMS/Default.aspx?alias=eppi.ioe.ac.uk/cms/er4& (accessed 24 Aug 2020).	✓	
101	EUnetHTA. Process of information retrieval for systematic reviews and health technology assessments on clinical effectiveness: European network for Health Technology Assessment, 2019.	✓	
117	Ganann R, Ciliska D, Thomas H. Expediting systematic reviews: methods and implications of rapid reviews. *Implement Sci* 2010; **5**: 56-	✓	
35	Glandville J, Lefebvre C, Wright K, Editors. ISSG Search Filter Resource [internet]. 2008 [updated 2020 February 4]. https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed 24 Aug 2020).	✓	
163	Golder S, Peryer G, Loke YK. Overview: comprehensive and carefully constructed strategies are required when conducting searches for adverse effects data. *Journal of clinical epidemiology* 2019; **113**: 36-43.		✓
126	Greenhalgh T, A'Court C, Shaw S. Understanding heart failure; explaining telehealth - a hermeneutic systematic review. *BMC cardiovascular disorders* 2017; **17**(1): 156.	✓	
105	Greenhalgh T, Wong G, al. e. The RAMESES Project. 2013-2020. http://www.ramesesproject.org/Standards_and_Training_materials.php.	✓	
100	Harris JL, Booth A, Cargo M, *et al*. Cochrane Qualitative and Implementation Methods Group guidance series-paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. Journal of clinical epidemiology 2018; 97: 39-48	✓	
5	Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.0 [Updated July 2019]. 07/2019 2019. https://training.cochrane.org/handbook/current (accessed 25/08/2020 2020).	✓	✓
102	Hoaglin DC, Hawkins N, Jansen JP, *et al*. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2011; 14(4): 429-37.	✓	
83	ISSG IISS-G. Search Filter Resource https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home (accessed 19.07. 2020).	✓	✓
134	Janssens ACJW, Gwinn M. Novel citation-based search method for scientific literature: application to meta-analyses. *BMC Medical Research Methodology* 2015; **15**(1): 84.	✓	
107	Kaunelis D, Glanville J. Cost and economic evaluation. Summarized Research in Information Retrieval for HTA (SuRE Info). 2020. Retrieved from http://vortal.htai.org/?q=node/336.	✓	
99	Kugley S, Wade A, Thomas J, et al. Searching for studies: a guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews* 2017; **13**(1): 1-73.	✓	

128	Li L, Smith HE, Atun R, Car LT. Search strategies to identify observational studies in MEDLINE and Embase. *Cochrane Database of Systematic Reviews* 2019; (3).	✓	
115	Millard T, Synnot A, Elliott J, Green S, McDonald S, Turner T. Feasibility and acceptability of living systematic reviews: results from a mixed-methods evaluation. *Systematic Reviews* 2019; **8**(1): 325.	✓	
157	Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535.	✓	
123	Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. *BMJ Global Health* 2019; **4**(Suppl 1): e000893.	✓	
137	O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* 2015; **4**(1): 5.	✓	
10	Paisley S. Identification of Evidence for Key Parameters in Decision-Analytic Models of Cost Effectiveness: A Description of Sources and a Recommended Minimum Search Requirement. *Pharmacoeconomics* 2016; **34**(6): 597-608.	✓	
154	Rader T, Mann M, Stansfield C, Cooper C, Sampson M. Methods for documenting systematic review searches: a discussion of common issues. *Research Synthesis Methods* 2014; **5**(2): 98-115.	✓	
142	Rayyan. https://rayyan.qcri.org/welcome (accessed 24 Aug 2020).	✓	
98	Relevo R, Balshem H. Chapter 5: Finding Evidence for Comparing Medical Interventions. Methods guide for effectiveness and comparative effectiveness reviews Agency for Health Care Research and Quality (AHRQ); 2014.	✓	
160	Rethlefsen M, Kirtley S, Waffenschmidt S, et al. PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. March 20, 2019 2019. https://doi.org/10.31219/osf.io/sfc38 (accessed 24 Aug 2020).	✓	✓
132	Robinson KA, Dunn AG, Tsafnat G, Glasziou P. Citation networks of related trials are often disconnected: implications for bidirectional citation searches. *Journal of clinical epidemiology* 2014; **67**(7): 793-9.	✓	
143	RobotAnalyst. http://www.nactem.ac.uk/robotanalyst/ (accessed 24 Aug 2020).	✓	
119	Saul JE, Willis CD, Bitz J, Best A. A time-responsive tool for informing policy making: rapid realist review. *Implementation Science* 2013; **8**(1): 103.	✓	
145	Shemilt I, Simon A, Hollands GJ, et al. Pinpointing needles in giant haystacks: use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Res Synth Methods* 2014; **5**(1): 31-49.	✓	
113	Simmonds M, Salanti G, McKenzie J, Elliott J. Living systematic reviews: 3. Statistical methods for updating meta-analyses. *Journal of clinical epidemiology* 2017; **91**: 38-46.	✓	
125	Smythe E, Spence D. Re-Viewing Literature in Hermeneutic Research. *International Journal of Qualitative Methods* 2012; **11**(1): 12-25.	✓	
144	Stansfield C, apos, Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. *Res Synth Methods* 2017; **8**(3): 355-65.	✓	

[149]	Stansfield C, Dickson K, Bangpan M. Exploring issues in the conduct of website searching and other online sources for systematic reviews: how can we be systematic? *Syst Rev* 2016; **5**(1): 191.	✓	
[106]	SURE Vortal; H. 2019. http://vortal.htai.org/?q=sure-info	✓	✓
[97]	Sutton A, Clowes M, Preston L, Booth A. Meeting the review family: exploring review types and associated information retrieval requirements. *Health Info Libr J* 2019; **36**(3): 202-22.	✓	
[152]	Taljaard M, McDonald S, Nicholls SG, et al. A search filter to identify pragmatic trials in MEDLINE was highly specific but lacked sensitivity. *Journal of clinical epidemiology* 2020; **124**: 75-84.	✓	
[164]	Thielen FW, Van Mastrigt G, Burgers LT, et al. How to prepare a systematic review of economic evaluations for clinical practice guidelines: database selection and search strategy development (part 2/3). *Expert Review of Pharmacoeconomics & Outcomes Research* 2016; **16**(6): 705-21.		✓
[114]	Thomas J, Noel-Storr A, Marshall I, et al. Living systematic reviews: 2. Combining human and machine effort. *Journal of clinical epidemiology* 2017; **91**: 31-7.	✓	
[155]	Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology* 2012; **12**(1): 181.	✓	
[153]	Tummers M, van Hoorn R, Levering C, Booth A, van der Wilt GJ, Kievit W. Optimal search strategies for identifying moderators and predictors of treatment effects in PubMed. 2019.	✓	✓
[151]	Waffenschmidt S, Navarro-Ruan T, Hobson N, Hausner E, Sauerland S, Haynes RB. Development and validation of study filters for identifying controlled non-randomized studies in PubMed and Ovid MEDLINE. *Res Synth Methods* 2020; **11**(5): 617-26.	✓	
[158]	Wong G, Greenhalgh T, Westhorp G, Buckingham J. RAMESES publication standards: meta-narrative reviews. *BMS Medicine* 2013; **11**(20).	✓	✓
[116]	Wright J, Walwyn R. Literature search methods for an overview of reviews ('umbrella' reviews or 'review of reviews') CILIP Health Libraries Group Conference; 2016 15th - 16th Sept; Scarborough, UK; 2016.	✓	